UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Centro Singular de Investigación en Tecnoloxías da Información

Tese de doutoramento

PATTERN RECOGNITION BEYOND CLASSIFICATION: AN ABDUCTIVE FRAMEWORK FOR TIME SERIES INTERPRETATION

Presentada por:
Tomás Teijeiro Campo

Dirixida por:
Paulo Félix Lamas
Jesús María Rodríguez Presedo

Decembro de 2016


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Paulo Félix Lamas, Profesor Titular de Universidade da Área de Ciencias da Computación e Intelixencia Artificial da Universidade de Santiago de Compostela

Jesús María Rodríguez Presedo, Profesor Titular de Universidade da Área de Ciencias da Computación e Intelixencia Artificial da Universidade de Santiago de Compostela

FAN CONSTAR:

Que a memoria titulada **PATTERN RECOGNITION BEYOND CLASSIFICATION: AN ABDUCTIVE FRAMEWORK FOR TIME SERIES INTERPRETATION** foi realizada por **D. Tomás Teijeiro Campo** baixo a nosa dirección no Centro Singular de Investigación en Tecnoloxías da Información da Universidade de Santiago de Compostela (CiTIUS), e constitúe a Tese que presenta para optar ó título de Doutor.

Decembro de 2016

Paulo Félix Lamas
Director da tese

Jesús María Rodríguez Presedo
Codirector da tese

Tomás Teijeiro Campo
Autor da tese


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

In theory, there is no difference between theory and practice.
But, in practice, there is.

Jan L. A. van de Snepscheut


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecementos

En primeiro lugar, quero agradecer ós meus directores Paulo Félix e Jesús Presedo toda a súa implicación e esforzo no desenvolvemento desta tese. Pero non só iso, pois realmente o traballo que aquí se presenta non é máis ca un dos moitos resultados dun proceso de desenvolvemento intelectual, persoal e profesional para o que ningunhas palabras de agradecemento poden facer xustiza. Todas as discusións e reflexións que temos compartido, a esixencia por buscar sempre o mellor que poidamos dar, e a ambición científica son as pezas nas que se apoia non só este traballo, senón todo o que veña a partir de agora.

Por outra banda, quero dar as grazas ó Departamento de Electrónica e Computación e ó Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS) da Universidade de Santiago de Compostela a oportunidade que me brindaron para a realización desta tese nun contorno de traballo excelente e no que desfrutei cada día.

Pero sen dúbida os maiores responsables deste bo ambiente foron os meus compañeiros, e sobre todo amigos, cos que compartín estes anos cheos de frustracións e dificultades, pero tamén de diversión e alegrías: Pablo, Ángel, Jorge, Tino, Arturo, Adri, Tefa, Ana, Ismael, Jano, David, Borja, moitas grazas por facer de psicólogos, mestres, servizo técnico, ou do que fixera falta en cada situación.

Por suposto, tamén teño que agradecer a todas aquelas institucións e entidades que fixeron materialmente posible este traballo co seu financiamento. Principalmente, ó programa de Formación do Profesorado Universitario (FPU) do Ministerio de Educación do Goberno de España (ref. AP2010-1012), pero tamén ó Ministerio de Economía e Competitividade do Goberno de España a través do proxecto TIN2014-55183-R, e á Fundación Pedro Barrié de la Maza, que me deu a oportunidade de realizar unha estadía predoutoral no *Laboratory for Biomedical Informatics "Mario Stefanelli"* da Università di Pavia a través da súa convocatoria de bolsas do ano 2014. Neste sentido, quero agradecerlle a todos os membros do grupo, e en

particular ó Prof. Riccardo Bellazzi, a súa cálida acollida e o fantástico trato que me deron durante esta estadía.

Dende un punto de vista máis persoal, quero agradecer sinceramente á miña familia todo o seu apoio e motivación durante esta etapa. Seguramente supoña certa frustración non ser quen de explicar en que traballa o teu fillo, e por iso ese apoio e axuda incondicionais aínda son máis apreciados. Tamén gracias especiais a miñas irmás e a meus primos, por servirme de exemplo e inspiración durante toda a miña vida, e ó meu recentemente chegado afillado, que aínda que de xeito involuntario axudou a levar a cabo o final deste traballo con máis gañas. Non quero rematar este agradecemento persoal sen acordarme de todos os amigos que estiveron sempre dispoñibles para ir tomar algo, contar un conto e amañar o mundo. Moitas gracias a todos.

Por último quero agradecerlle moi especialmente a Nuria o seu apoio, a súa paciencia, o seu cariño e o seu esforzo durante todos estes anos por axudarme cada día a sentirme orgulloso e con confianza e gañas de seguir adiante. Esta tese tamén é túa.

Santiago de Compostela, Decembro de 2016

Contents


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Algorithms


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Figures

List of Tables


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumo da tese

A capacidade humana para o recoñecemento de patróns a partir da información sensorial é un dos aspectos chave que definen a nosa intelixencia. Xa dende moi nenos os humanos somos quen de recoñecer caras, obxectos, sons, etc., e esta habilidade é a base para posteriormente poder comunicarnos a través da fala ou da escritura, e chegar a desenvolver abstraccións de tan alto nivel coma as que se poden atopar por exemplo no estudo das matemáticas.

O propósito fundamental da Intelixencia Artificial é conseguir que as máquinas sexan quen de amosar un comportamento intelixente, no sentido de demostrar unhas habilidades cognitivas comparables ou que incluso superen ás dos seres humanos. Xa que logo, o recoñecemento de patróns é unha das áreas que máis interese suscita dentro da comunidade investigadora neste campo, e os seus resultados adoitan ter unha aplicación directa en multitude de problemas da ciencia e da enxeñaría.

Porén, existen aínda multitude de problemas que resultan moi doados de resolver para un ser humano e para os cales a Intelixencia Artificial aínda non proporcionou solucións comparables, coma a interpretación da fala, a lectura, ou a descrición de imaxes, entre moitos outros. Todos estes problemas teñen en común a necesidade de categorizar un conxunto de observacións, o que levou a abordalos de xeito natural como problemas de clasificación. Téñense desenvolvido multitude de paradigmas, estratexias e técnicas para a resolución de problemas de clasificación, xa sexa mediante métodos de aprendizaxe automática a partir de exemplos ou mediante unha definición formal dunha función de clasificación, pero todas as aproximacións teñen en común a aplicación dun mesmo esquema de razoamento, asumindo unha relación de dedución lóxica entre os valores das observacións de entrada e a clase identificada como resultado. Deste xeito, unha vez o clasificador toma unha decisión, esta decisión non poderá ser modificada a posteriori por parte do mesmo clasificador, e calquera erro cometido propagarase ás seguintes etapas de procesamento, dificultando a obtención de resultados acertados a medida que engadimos novas etapas de máis alto nivel.

Se analizamos o xeito no que os humanos levamos a cabo este tipo de tarefas, dificilmente atoparemos este tipo de razoamento monótono. Por exemplo, cando un humano le un texto non analiza letra a letra antes de concluír cada palabra, nin analiza palabra a palabra antes de concluír cada frase. De feito, somos *qn d cmprndr sn dficlutades frss ml fromadas*, nun proceso de razoamento que utiliza conxecturas de alto nivel (a primeira palabra da frase parece ser *"quen"*) para deducir información incompleta de baixo nivel (faltan as letras "u" e "e" nesa palabra).

O obxectivo principal desta tese de doutoramento é o de desenvolver un marco teórico e un conxunto de algoritmos que permitan a interpretación do comportamento dun sistema complexo ó longo do tempo a partir dun conxunto de observacións efectuadas sobre o propio sistema, imitando o xeito no que os humanos levamos a cabo esta tarefa. Esta interpretación implica a descrición e caracterización dos procesos e mecanismos subxacentes a esas observacións, que aparecerán na forma de series temporais.

A nosa hipótese de partida é que o modelo de razoamento que mellor dá conta do xeito no que os humanos levamos a cabo tarefas de recoñecemento de patróns é o razoamento abdutivo. A abdución foi definida polo filósofo Charles Sanders Peirce (1839-1914) como o proceso cognitivo polo cal os humanos respondemos ante unha observación que capta a nosa atención mediante unha busca da hipótese ou hipóteses que mellor expliquen esa observación. Comunmente, este paradigma de razoamento tamén se coñece como inferencia da mellor explicación, e no campo da Intelixencia Artificial tense empregado con éxito para a resolución de multitude de problemas, principalmente tarefas de diagnóstico.

A principal diferencia do razoamento abdutivo con respecto ó razoamento dedutivo que empregan os clasificadores comúns é o seu carácter non monótono, é dicir, que a verdade das premisas non garante a verdade da conclusión. Vexamos un exemplo típico de razoamento dedutivo en contraposición a un razoamento abdutivo:

> *Todos os corvos son negros.*
> *Aquel paxaro é un corvo.*
> *Polo tanto, aquel paxaro é negro. (Razoamento dedutivo)*

> *Todos os corvos son negros.*
> *Aquel paxaro é negro.*
> *Polo tanto, aquel paxaro é un corvo. (Razoamento abdutivo)*

Como se pode ver, dende un punto de vista lóxico o razoamento abdutivo é equivalente á falacia de verificación do consecuente, e polo tanto as conclusións poden ser incorrectas. Isto resulta evidente en tanto que unha hipótese explicativa é unha conxectura que pode resultar falsa, e polo tanto un esquema de interpretación abdutivo deberá ser quen de refutar ou modificar conclusións previas en base a nova información. Non obstante, mentres unha conclusión dedutiva é unha afirmación que está implicitamente contida nas súas premisas, unha conclusión abdutiva amplía a información das premisas, permitindo predicir nova evidencia.

O marco proposto pode verse como un método para a definición de sistemas expertos que empregan razoamento baseado en modelos para a interpretación de series temporais nun dominio determinado. Deste xeito, partindo dunha base de coñecemento existente e dado un conxunto de observacións iniciais en forma de series temporais, os algoritmos desenvolvidos proporcionan unha interpretación dos procesos subxacentes a esas observacións. Esta interpretación constrúese seguindo un proceso de hipótese-test, polo cal ante a aparición de nova evidencia conxectúrase aquela explicación que mellor dá conta do observado (hipótese), e compróbase que as consecuencias desa conxectura en forma de predicións realmente se cumpren (test). Cada nova hipótese pode á súa vez formar parte da evidencia dunha nova hipótese de maior nivel de abstracción, co que o ciclo se repite ata que toda a evidencia sexa explicada ou ata que non sexa posible conxecturar novas hipóteses consistentes.

O dominio escollido para a validación da nosa proposta é o da interpretación do electrocardiograma (ECG). O ECG é un sinal obtido mediante sensores colocados sobre a pel dun individuo, rexistrando a actividade eléctrica do corazón ó longo do tempo. Normalmente o ECG mídese simultaneamente en diversos puntos do corpo, proporcionando cada un deles unha perspectiva diferente dos ciclos de activación e recuperación eléctrica do miocardio, os cales guían a contracción e o relaxamento do músculo cardíaco. O ECG é unha proba non invasiva e de moi baixo custo, e dende hai décadas constitúe o principal método para o estudo e o diagnóstico de cardiopatías, incluíndo arritmias, bloqueos de condución eléctrica, hipertrofia das cavidades cardíacas, insuficiencia cardíaca, etc. Os avances tecnolóxicos dos últimos anos permiten que hoxe en día sexa posible levar a cabo unha monitorización remota e continua do sinal de ECG mediante dispositivos móbiles, pero aínda non se dispón de algoritmos o suficientemente robustos para que a análise deste sinal se realice de forma autónoma e cunhas prestacións comparables ás dun experto humano.

Dende un punto de vista cognitivo, a interpretación do ECG é unha tarefa semellante á da lectura (é frecuente denominar a interpretación do ECG como “lectura do ECG”). No sinal de ECG poden identificarse diversas ondas arquetípicas, das cales as máis importantes son: a onda P, que representa a activación eléctrica das aurículas; o complexo QRS, que representa a activación dos ventrículos; e a onda T, que representa a recuperación eléctrica dos ventrículos. A Figura 1.3, na páxina 21 deste documento amosa a forma típica destas ondas, que adoitan repetirse na mesma secuencia con cada ciclo cardíaco.

As diferentes ondas poden verse como “letras” que se combinan para formar “palabras” que describen as características de cada latexo. A análise da morfoloxía, a duración e os intervalos entre as diferentes ondas dun latexo permite identificar o funcionamento normal ou anómalo das diferentes partes do músculo cardíaco, e incluso caracterizar a orixe e causa de moitas anomalías. Á súa vez, a secuencia de latexos proporciona información moi relevante sobre o funcionamento dinámico do corazón, permitindo identificar e caracterizar distintos tipos de arritmia como taquicardias, latexos ectópicos, etc. Este último nivel podería asociarse coas “frases” segundo o símil da lectura. Deste xeito, podemos ver como a interpretación proporciona unha descrición e caracterización dos procesos subxacentes (fenómenos de activación e recuperación eléctrica do músculo) responsables das observacións que forman a evidencia inicial (secuencia de valores numéricos que miden unha diferencia de potencial eléctrico na pel).

Para a definición da base de coñecemento dun dominio específico proponse un formalismo baseado na noción de *patrón de abstracción*. Un patrón de abstracción define unha relación entre un observable hipótese e un conxunto de observacións que conforman a evidencia na que se sustenta esa hipótese, permitindo a definición de restricións temporais e de valor que deberán satisfacerse para que a hipótese sexa considerada consistente. Por exemplo, un patrón de abstracción que tivese como hipótese un latexo cardíaco definiría como evidencia unha onda P, un complexo QRS e unha onda T, e unhas restricións temporais e de valor entre esas observacións que recollen a descrición que pode atoparse en calquera manual de electrocardiografía. O exemplo 2.1.7, na páxina 38 deste documento recolle a definición formal deste patrón de abstracción.

En xeral, unha mesma hipótese pode abstraer un número indeterminado de observacións. Por exemplo, unha taquicardia defínese como un ritmo cardíaco que supera as 100 pulsacións por minuto, e non ten a priori unha duración máxima determinada. Polo tanto, unha mesma hipótese de taquicardia pode abstraer un número potencialmente infinito de latexos, sempre que entre eles haxa unha separación temporal inferior a 600 milisegundos. Para permitir a definición de patróns de abstracción cun número indeterminado de pezas de evidencia definiuse un formalismo baseado no uso de gramáticas formais atribuídas, que dan soporte á xeración dinámica de patróns de abstracción. A expresividade destas gramáticas é equivalente á das expresións regulares, se ben permiten a maiores a definición, na súa semántica, de restricións temporais e de valor entre cada un dos símbolos terminais.

Un problema de interpretación consta polo tanto dun conxunto de observacións iniciais e un modelo de abstracción do dominio expresado como un conxunto de gramáticas xeradoras de patróns de abstracción, e o obxectivo é a construción dunha interpretación que explique as observacións iniciais no maior nivel de abstracción posible, empregando un modelo de razoamento abdutivo. A resolución óptima deste problema ten unha complexidade NP-dura, polo que se fai preciso adoptar certas estratexias que fagan que a computación sexa viable. Deste xeito, proponse un método de resolución de problemas de interpretación abordándoos como problemas de busca heurística. Partindo dunha interpretación trivial, que non contén ningunha hipótese, realízase a exploración dun espazo de busca en forma de árbore de interpretacións, no que cada nodo engade unha nova hipótese ao nodo pai, ou explora unha nova predición da última hipótese incluída no nodo pai. Unha característica moi relevante desta estratexia é que permite reconstruír todos os pasos de razoamento que se levaron a cabo para acadar (ou descartar) unha determinada interpretación, o que resulta de moita utilidade á hora de definir, corrixir e revisar os patróns de abstracción que forman parte do coñecemento. A heurística que guía a exploración baséase en catro principios clásicos do razoamento explicativo:

- *Principio de cobertura*, que establece a preferencia por interpretacións que expliquen un maior número de observacións.

- *Principio de simplicidade*, tamén coñecido como *navalla de Ockham*, que establece a preferencia por interpretacións cun menor número de hipóteses.

- *Principio de abstracción*, que establece a preferencia por interpretacións en niveis de abstracción altos.

- *Principio de predición*, que establece a preferencia por interpretacións que sexan capaces de predicir observacións futuras.

O algoritmo CONSTRUE, que implementa o ciclo de hipótese-test, foi desenvolvido para a resolución de problemas de interpretación en dous escenarios diferentes: O primeiro, denominado *en diferido*, asume que no momento de comezar a interpretación xa se dispón de toda a evidencia inicial, e polo tanto as métricas de cobertura poden calcularse de xeito absoluto. O segundo escenario, denominado *continuo*, asume que a evidencia se vai adquirindo dinamicamente, como ocorre por exemplo en monitorizacións en tempo real. Este segundo escenario esixe que o algoritmo sexa capaz de ofrecer unha solución en calquera momento, e as restricións computacionais son maiores ca no escenario en diferido.

No ciclo de hipótese-test lévase a cabo unha construción dinámica de interpretacións mediante a combinación de diferentes modos de razoamento abaixo-arriba e arriba-abaixo guiados por un mecanismo inspirado na atención humana. Este mecanismo denomínase *foco de atención*, e contén a seguinte observación que deberá ser explicada, ou a seguinte predición derivada dunha hipótese que deberá ser comprobada. En base á natureza do contido do foco de atención, o esquema de razoamento aplicado variará do seguinte xeito:

1. Se o foco de atención contén unha observación non explicada, entón realízase un proceso de *abdución* para conxecturar unha hipótese que dea conta desa observación, e esa nova hipótese pasa a estar no foco de atención.

2. Se o foco de atención contén unha hipótese que xa explica unha ou varias observacións, entón lévase a cabo un proceso de *dedución*, polo cal se obtén unha predición a partir desa hipótese e do patrón de abstracción que a soporta. A nova predición xérase mediante un percorrido das regras da gramática asociada con esa hipótese, e pasa a estar no foco de atención.

3. Se o foco de atención contén unha predición feita a partir dalgunha hipótese, e existe unha observación consistente con esa predición, esa observación pasará a considerarse explicada pola hipótese mediante un procedemento de *subsunción*. A continuación, recupérase o foco de atención anterior, correspondente á hipótese na que se subsumiu a observación, permitindo a obtención dunha nova predición.

4. Por último, se o foco de atención contén unha predición para a cal non existe unha observación consistente, esa predición pasará a ser unha hipótese situada no foco de atención, polo que nos seguintes pasos buscarase a evidencia que soporte esta nova hipótese a un menor nivel de abstracción.

A Figura 3.1, na páxina 64 ilustra un exemplo completo de interpretación na que estes modos de razoamento se van sucedendo e complementando ata proporcionar unha interpretación completa das observacións iniciais.

O traballo recollido neste documento organízase do seguinte xeito:

- No Capítulo 1 faise unha revisión bibliográfica da abdución dende os puntos de vista filosófico, lóxico e da intelixencia artificial, e discútense as principais referencias e traballos nos que se basea esta proposta. Tamén se fai unha pequena introdución á análise do ECG, motivando a súa adopción como exemplo paradigmático de dominio no cal a interpretación abdutiva pode proporcionar resultados satisfactorios, en tanto que as maiores limitacións identificadas nas aproximacións existentes están relacionadas coa adopción dun esquema de razoamento dedutivo.

- O Capítulo 2 formaliza o marco de interpretación proposto. Isto inclúe a definición das nocións de observable e observación como entidades de representación básicas, e patróns de abstracción como a primitiva de descrición do coñecemento do dominio. Posteriormente, formalízanse nocións fundamentais coma a de relación de abstracción, hipótese, ou interpretación, e defínese o que é un problema de interpretación e a súa solución dende un punto de vista conxuntista. Por último, estúdase a complexidade computacional do problema, demostrando que a busca dunha solución óptima é un problema NP-duro.

- O Capítulo 3 describe CONSTRUE, o algoritmo proposto para a resolución de problemas de interpretación de forma efectiva. Este algoritmo aborda os problemas de interpretación como problemas de busca heurística, xerando dinamicamente unha interpretación a través dun ciclo de hipótese-test no que se combinan distintos modos de razoamento abaixo-arriba e arriba-abaixo, apoiándose nun mecanismo de atención que determina cal é o seguinte paso a realizar. En primeiro lugar descríbese a versión *en diferido* do algoritmo, na que se asume que toda a evidencia a explicar está dispoñible antes de comezar a interpretación. A descrición céntrase nas distintas etapas do ciclo de hipótese-test e no mecanismo de atención, amosando como esta estratexia é consistente coa intuición sobre como os humanos interpretamos información temporal en múltiples niveis de abstracción. A continuación, descríbese o algoritmo CONSTRUE-ONLINE, que xeneraliza a CONSTRUE permitindo a adquisición continua de nova evidencia durante o proceso de interpretación. Ademais, introdúcense unha serie de estratexias de optimización para mellorar a eficiencia da interpretación, entre as que se atopan un mecanismo de poda para limitar a exploración de hipóteses, unha nova estratexia de exploración da árbore de interpretacións que explota as nocións de necesidade e suficiencia da evidencia que soporta unha hipótese, e un formalismo de representación de coñecemento temporal baseado no STP (Simple Temporal Problem) que permite mellorar a eficiencia no razoamento temporal.

- No Capítulo 4 preséntanse as principais contribucións e resultados desta tese a través de dous experimentos que amosan a capacidade práctica da proposta con problemas ben coñecidos: O primeiro é o problema da detección de complexos QRS en sinais de ECG, no que se salienta a capacidade dun esquema de razoamento non monótono para corrixir conclusións previas. O segundo problema é o da clasificación de latexos cardíacos pola súa orixe, no cal demostramos que unha interpretación en múltiples niveis de abstracción permite describir o mesmo conxunto de características abstractas que empregan os expertos humanos nesta tarefa, e que a partir delas e cun conxunto simple de regras cualitativas conseguimos superar os resultados dos mellores clasificadores automáticos do estado do arte, e incluso os da maioría de propostas que precisan a asistencia dun experto humano.

- No Capítulo 5 discútense os principais aspectos teóricos da nosa proposta, ilustrándoos con exemplos reais do dominio da análise do ECG e facendo énfase naquelas características que permiten superar as principais limitacións das aproximacións tradicionais baseadas en clasificadores. Entre estas vantaxes destacan: 1) a capacidade de incluír de xeito natural a concorrencia de múltiples procesos nunha mesma interpretación, 2) a posibilidade de expresar ignorancia dun xeito explícito como resultado da interpretación, 3) a capacidade de buscar activamente aquela evidencia non dispoñible a priori, ou 4) a interpretabilidade dos resultados da interpretación e do proceso de razoamento. O capítulo remata cunha serie de reflexións semánticas sobre os principais conceptos definidos no marco de traballo.

- O Capítulo de conclusións presenta unha síntese dos resultados e das conclusións obtidas nesta investigación, ademais de debullar novas vías para a investigación futura.

- No Apéndice A descríbese o modelo de abstracción completo para o problema de interpretación de sinais de ECG en múltiples niveis de abstracción, incluíndo desviacións do sinal, activación/recuperación eléctrica das aurículas e ventrículos, e patróns de ritmos normais e diferentes tipos de arritmias. Ademais, e co obxectivo de facer reproducible esta investigación, o código fonte completo da implementación deste modelo de abstracción e dos algoritmos de interpretación foi publicado baixo unha licenza de software libre.

- O Apéndice B describe un novo algoritmo para a delineación de complexos QRS en sinais de ECG con múltiples derivacións. Este algoritmo baséase nunha técnica de simplificación de trazados, e está especialmente deseñado para traballar con sinais ruidosos. Emprégase no modelo de abstracción do ECG como procedemento de observación no patrón de abstracción de complexos QRS.


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 1

INTRODUCTION

The interpretation and understanding of the behavior of a complex system involves a cognitive activity aimed at guessing the processes and mechanisms underlying what is observed. The human ability to recognize patterns plays a paramount role as an instrument for highlighting evidence which should require an explanation, by matching information from observations with background knowledge retrieved from memory. Classification naturally arises as a pattern recognition task, defined as the assignment of observations to categories.

Let us first state precisely what is the problem under consideration: we wish to interpret the behavior of a complex system by measuring a set of physical quantities along time. These quantities are represented as time series.

The scientific community has devoted a great deal of effort on different paradigms, strategies, methodologies and techniques for time series classification, both in time and transformed domains: Artificial neural networks [Kehagias and Petridis, 1997], hidden Markov models [MacDonald and Zucchini, 1997], decision trees [Geurts, 2001], fuzzy rules [Bardossy et al., 1995], autoregressive models [Corduas and Piccolo, 2008], support vector machines [Kampouraki et al., 2009], nearest-neighbor classifiers [Chaovalitwongse et al., 2007] or Bayesian networks [Pavlovic et al., 1999], these are just some examples of the diversity of approaches used in the bibliography. Nonetheless, in spite of the wide range of proposals for building classifiers, either by eliciting domain knowledge or by induction from a set of observations, the resulting classifier behaves as a deductive system, assuming a logical implication between a time series fragment and the pattern identified as a result. The present work is premised on the thesis that some of the important weaknesses of time series classification lie in its deductive nature, and that an abductive approach can address these shortcomings.

Let us remember that a deduction contains in its conclusions information that is already implicitly contained in the premises, and for this reason is called *truth-preserving*. In this sense, a classifier ultimately assigns a label or a set of labels to each observation. This label can designate a process or a mechanism of the system being observed, but it is nothing more than a term that summarizes the premises satisfied by the observations. As a consequence, deductive reasoning is monotonic, and conclusions cannot be retracted after they have been inferred. Conversely, abduction goes from data to a hypothesis that best explains or accounts for the data. Abductive conclusions contain new information, not contained in the premises, capable of predicting new evidence, although they are fallible and new information may invalidate them. Abductions are thus *truth-widening*, and they can make the leap from the language of observations to the language of the underlying processes and mechanisms, responding to the problem under consideration in a natural way. For example, consider a simple rule stating that if a patient experiences a sudden tachycardia and a decrease in blood pressure, then we can conclude that is suffering from shock due to a loss of blood volume. From a deductive perspective, *loss of blood volume* is just a name provided by the rule for the satisfaction of the two premises. However, from an abductive perspective, *loss of blood volume* is an explanatory hypothesis, a conjecture, that expands the truth contained in the premises, enabling the observer to predict additional consequences such as, for example, pallid skin, faintness or dizziness.

Abduction was first described by the American philosopher Charles Sanders Peirce as the cognitive process that respond to surprising observations with a search for hypotheses that can explain them [Hartshorn et al., 1931]. The logical form of abduction formulated by Peirce is as follows:

> *The surprising fact,* ***C****, is observed;*
> *But if* ***A*** *were true,* ***C*** *would be a matter of course,*
> *Hence, there is reason to suspect that* ***A*** *is true.*

Peirce categorized human reasoning in three possible inference patterns: deduction, abduction, and induction. The difference between them is illustrated in the following syllogistic example [Flach, 1996]:

> *All the beans from this bag are white.*
> *These beans are from this bag.*
> *Therefore, these beans are white. (Deductive reasoning)*

All the beans from this bag are white.
These beans are white.
Therefore, these beans are from this bag. (Abductive reasoning)

These beans are from this bag.
These beans are white.
Therefore, all the beans from this bag are white. (Inductive reasoning)

In this form, abduction can be viewed as a *reverse modus ponens* reasoning [Kodratoff, 1994], so it is formally equivalent to the fallacy of affirming the consequent, thus assuming as *sufficient* a condition that is only *necessary*. On the other hand, induction is a generalization that goes from the characteristics of a sample of individuals to a conclusion about the whole population. This example highlights the main difference between monotonic (deduction) and non-monotonic logics (abduction and induction): any conclusion in a non-monotonic logic may be incorrect. In the case of the abductive conclusion, white beans could come from another bag, while for the inductive conclusion the bag could contain an unseen non-white bean.

There is still an open discussion about taxonomies for human reasoning, both in the philosophy [Magnani et al., 2010] and logic [Flach and Kakas, 2000a] fields, with significant efforts to integrate abduction and induction under the same non-monotonic reasoning scheme. Authors like Magnani believe that all creative reasoning is abductive, including aesthetic creations like art, literature or music [Magnani, 2015]. This is supported by the idea that abduction is a natural consequence of the neural topography of the brain [Thagard, 2010]. On the other hand, authors like Lipton discuss the integration of abduction and induction as necessary pieces for explanatory reasoning under the Inference to the Best Explanation (IBE) paradigm [Lipton, 2004]. Even Peirce modified his vision of human reasoning in his later *inferential theory* [Flach and Kakas, 2000b], where abduction, deduction and induction are considered three necessary stages of scientific inquiry, corresponding to hypothesis generation, prediction, and evaluation. It is not the subject of this work to deepen this discussion, so we will adopt a classical perspective from the domain of Artificial Intelligence aligned with Peirce's inferential theory, where abduction is identified with the inference to the best explanation [Peng and Reggia, 1990; Console and Torasso, 1991; Josephson and Josephson, 1994].

Of course, the result of a classifier can also be considered as a conjecture, but always from an external agent, since a classifier is monotonic as a logical system and its conclusions cannot be refuted from within. Multi-classifiers, or classifier ensembles [Hastie et al., 2009], aim to overcome the errors of individual classifiers by combining different classification paradigms to obtain a better result; thus, a classifier can be amended by others in the final result of the multi-classifier. However, even a multi-classifier represents a bottom-up mapping, and classification invariably fails above a certain level of distortion within the data.

The interpretation and understanding of a complex system usually unfolds along a set of abstraction layers, where at each layer the temporal granularity of the representation is reduced from below. A classification strategy provides an interpretation as the result of connecting a set of classifiers along the abstraction structure, and the monotonicity of deduction entails a propagation of errors from the first abstraction layers upwards, narrowing the capability of making a proper interpretation as new abstraction layers are successively added. To overcome this weakness, the usual strategy is to embrace ad-hoc error-checking procedures at the expense of greater complexity of the model and reducing its adaptability to other domains. Following an abductive process instead, an observation is conjectured at each abstraction layer as the best explanatory hypothesis for the data from the layer or layers below, within the context of information from above, and the non-monotonicity of abduction supports the retraction of any observation at any abstraction layer in the search for the best global explanation. Thus, bottom-up and top-down processing complement one another and provide a joint result [Josephson and Josephson, 1994]. As a consequence, abduction can guess the underlying processes from corrupted data or even in the temporary absence of data.

If we analyze how humans face this sort of abstraction tasks, we can hardly find a one-way reasoning flow. Even the most primary reasoning, at the perception level, shows a a non-monotonic behavior. An illustrative example is the cognitive phenomenon called “repetition blindness” [Kanwisher, 1987], by which you may have not noticed the repetition of the word “a” in the previous sentence. This kind of phenomena demonstrate that human reading does not follow the classical abstraction sequence *shape perception* → *character recognition* → *lexical analysis* → *syntax analysis*, but that some integration of top-down and bottom-up strategies is used to make the process more efficient and robust.

On the other hand, a classifier is based on the assumption that the underlying processes or mechanisms are mutually exclusive. Superpositions of two or more processes are excluded; they must be represented by a new process, corresponding to a new category which is different and usually unrelated to previous ones. Therefore, an artificial casuistry-based heuristics is adopted, increasing the complexity of the interpretation and reducing its adaptability to the variability of observations. In contrast, abduction can reach a conclusion from the availability of partial evidence, refining the result by the incremental addition of new information. This makes it possible to discern different processes just from certain distinguishable features, and at the end to infer a set of explanations as far as the available evidence does not allow us to identify the best one, and they are not incompatible with each other [Peng and Reggia, 1990].

In a classifier, the truth of the conclusion follows from the truth of each and every premise, and missing data usually demand an imputation strategy that results in a conjecture: a sort of abducing to go on deducing. In contrast, an abductive interpretation is posed as a hypothesize-and-test cycle, in which missing data is naturally managed, since a hypothesis can be evoked by every single piece of evidence in isolation and these can be incrementally added to reasoning. This fundamental property of abduction is well suited to the time-varying requirements of the interpretation of time series, where future data can compel changes to previous conclusions, and the interpretation process can be used to provide the current result as the best explanation at any given time.

Abductive reasoning has been widely and successfully adopted to different tasks in Artificial Intelligence, primary diagnosis [Console and Torasso, 1991; Cox and Pietrzykowski, 1986; de Kleer et al., 1992; Peng and Reggia, 1990; Poole et al., 1987], but also temporal abstraction [Campos et al., 2010; Henson et al., 2012; Alirezaie and Loutfi, 2014], question answering [Ferrucci et al., 2012], language understanding [Hobbs et al., 1993; Josephson and Josephson, 1994], story comprehension [Charniak, 1989], image understanding [Poole, 1990] or plan recognition [Kautz and Allen, 1986; Litman and Allen, 1987], among others. Some studies have proposed that perception might rely on some form of abduction [Josephson and Josephson, 1994], even if it involves many basic neuropsychological processes that are not completely described as generation and evaluation of explanatory hypotheses [Thagard, 2010], like for example perceptual *priming* or the various principles of visual grouping [Smith and Kosslyn, 2013]. All these tasks have in common a body of data demanding a hypothesis or set of hypotheses that best explain or account for the data. At the end, the best explanation provides more information than was available before abduction, because new predictions can be made from these hypotheses, not contained in the initial data.

Although abduction is generally an NP-hard problem [Bylander et al., 1991; Josephson and Josephson, 1994], this has not prevented its application after establishing the necessary restrictions to ensure tractability. Thus, a number of different logic-based abductive frameworks have been proposed, motivated by a trade-off between expressiveness and computational efficiency [Console et al., 1991; Cox and Pietrzykowski, 1986; de Kleer, 1986; Poole, 1989;

Flach, 1994]. Other authors propose a syntactically restricted representation based on set theory, translating the problem of inferring explanations to an equivalent set covering problem [Peng and Reggia, 1990]. These frameworks can be extended with a probabilistic schema enabling a computationally feasible ranking of explanations [Pearl, 1988; Peng and Reggia, 1990]. The present work relies on a compiled form of abduction [Josephson and Josephson, 1994] based on a set of pre-stored hypotheses in the form of *temporal abstraction patterns*, which narrow the generation of hypotheses. A temporal abstraction pattern develops abductively the notion of archetype over time series data: a number, possibly indefinite, of available observations appear in a particular arrangement, satisfying certain constraints that commonly induce us to identify the occurrence of a certain process ψ. This leads us to hypothesize the observation of that process, estimating certain features of ψ from the initial observations. Section 1.1 informally introduces and motivates the notion of temporal abstraction pattern and discusses the semantics of observation from an abductive perspective.

But abduction is not the only mechanism proposed for inferring explanations. Some authors have studied different consistency-based approaches for diagnosis tasks [Davis, 1984; Genesereth, 1984; de Kleer and Williams, 1984; Reiter, 1987]. In such approaches, a model of the correct behavior of the system to be diagnosed is available, and a diagnosis is a minimal set of assumptions about the abnormal behavior of some components of the system. Thus, a diagnosis explains a set of observations if it does not contradict them. Instead, abduction carries a stronger notion of explanation, whereby a diagnosis explains a set of observations if it directly implies (covers) them, and it fits better to the availability of a fault model of the system to be diagnosed. Both approaches were integrated in a unified framework for solving an spectrum of abduction problems with consistency constraints, where the selection of the subset of observations to be covered by an explanation leads to different definitions of explanations between the two ends of the spectrum [Console and Torasso, 1991]. In this sense, in our proposal we pursue interpretations satisfying the abductive definition of explanation, covering all the initial evidence. However, we assume that the knowledge may be incomplete, and that this incompleteness is not modeled —Console and Torasso suggest the use of anonymous premises to explicitly represent unknown causes—, so consistent explanations not covering all the evidence are also considered valid interpretations as a way to express ignorance. This is discussed in detail in Section 5.2.

Time representation and reasoning plays a central role in characterizing the behavior of most systems, and it has been promptly included as a fundamental piece of temporal abduction frameworks [Gamper and Nejdl, 1997; Hamscher, 1991; Long, 1996; Palma et al., 2006; Wainer and de Melo, 1997; Shahar, 1997]. Diagnosis is once again the main application of most of the proposals in literature. According to a knowledge-level analysis made in [Brusoni et al., 1998], these proposals are classified according to four dimensions: (1) the type of temporal phenomena they take into account: how the system to be diagnosed changes, when diagnosis is performed and when the diagnosis results are needed; (2) the ontology of time they adopt; (3) the notion of logical explanation they use —consistency-based versus abductive—; and (4) the notion of temporal explanation they use —consistency with the temporal information of observations versus entailment of such information—. Authors provide a unified framework, extending the spectrum of definitions of diagnosis proposed by Console and Torasso [1991]. The present work concerns time as a fundamental dimension, since we will cope with continuously evolving systems that leave behind a set of time series data resulting from a monitoring process. Two different scenarios for the interpretation of such systems are assumed: a scenario of *off-line interpretation*, where all the data implicated in the interpretation are available at the time of performing this task; and a scenario of *on-line interpretation*, where new data is acquired continuously during the interpretation. This last scenario can be further constrained to operate in real time, demanding to provide an interpretation on time to be able to control the system being monitored.

This work takes as a starting point the thesis that abduction is the proper type of inference for the pattern recognition task, considered as a part of a dynamic interpretation process. Pattern recognition plays the role of an explanatory task, in charge of identifying and characterizing over sensory data the underlying physical processes from their particular signature on the data. These signatures are manifestations of the underlying processes and pattern recognition thus results in a set of conjectures about those processes, providing at the end an interpretation about the available data. Furthermore, pattern recognition could perform recursively over their own results, aggregating information in a set of abstraction levels. Hence this work aims to be an application of the aforementioned thesis about the abductive nature of perception, wherein *”the process of interpretation is broken down into discrete layers where at each layer a best-explanation hypothesis is formed of the data presented by the layer or layers below, with the help of information from above”* [Josephson and Josephson, 1994].

1.1 Interpretation as a process-guessing task

Let $\mathcal{O}_t = \{o_1, o_2, \ldots, o_i\}$ be a set of observations made on a system S until time t, such that $\mathcal{O}_t \subseteq \mathcal{O}_{t+1}$. These observations appear as a set of time series obtained from a data acquisition process. It is assumed that these observations are manifestations of a set of unknown underlying processes $\Psi_t = \{\psi_1, \psi_2, \ldots, \psi_n\}$ taking place in this system at some extent during the observation period. We will address the problem of identifying and characterizing the processes of the set Ψ_t from the observations in $\mathcal{O}_t$.

In order to solve this interpretation task, prior knowledge is available as a set of patterns, assigning a set of observations to a certain process in so far as they appear in a recognizable form, thus behaving as a sort of signature. Since a process can be identified from a set of observations, the process itself is usually said to be observable, even though its observation is actually a conjecture, as long as it is fallible. This perception is reinforced as the observations are obtained by acquisition procedures closer to the process at the physical level, and therefore an identification between the process and its signature at the observational level is quite common. We will assume that every process $\psi \in \Psi_t$ is assigned an observable entity as a hypothesis h_ψ and, for the sake of simplicity, that every process occurs simultaneously to its pattern of observations.

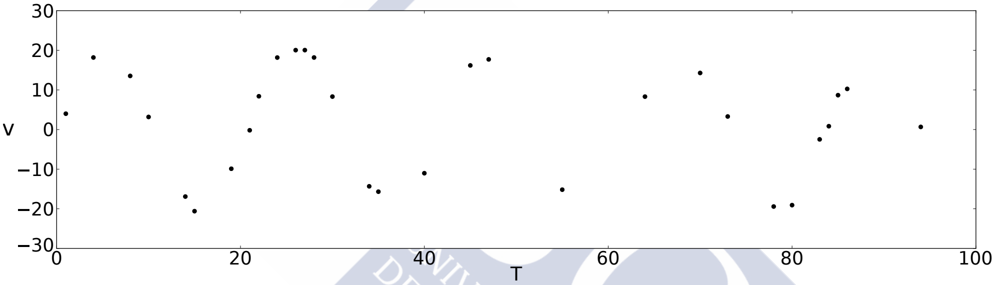
We propose a knowledge-based interpretation framework upon the principles of abductive reasoning, i.e., on the basis of a strategy of hypothesis formation and testing. Taking as a starting point a time series of physical measurements, a set of observations are guessed as conjectures of the underlying processes, through successive levels of abstraction. Each new observation will be generated from previous levels as the underlying processes aggregate, superimpose or concatenate to form more complex processes with greater duration and scope, and are organized into an abstraction hierarchy.

The knowledge of the domain is described as a set of abstraction patterns as follows:

$$h_\psi((\mathbf{A}_h, T_h^b, T_h^e) = \pi(\mathbf{A}_1, T_1, \ldots, \mathbf{A}_n, T_n)) \ \textit{abstracts}\ m_1(\mathbf{A}_1, T_1), \ldots, m_n(\mathbf{A}_n, T_n)$$
$$\{C(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, T_1, \ldots, \mathbf{A}_n, T_n)\}$$

where $h_\psi(\mathbf{A}_h, T_h^b, T_h^e)$ is an observable of the domain playing the role of a hypothesis on the observation of an underlying process ψ; $\mathbf{A}_h$ represents a set of attributes, and its temporal support is represented by two instants T_h^b and T_h^e, corresponding to the beginning and the end of the observable; $m_1(\mathbf{A}_1, T_1), \ldots, m_n(\mathbf{A}_n, T_n)$ is a set of observables of the domain which plays the role of the evidence suggesting the observation of h_ψ, where each of these has its own set of attributes $\mathbf{A}_i$ and temporal support T_i, represented here as a single instant for the sake of simplicity, although this may also be an interval; C is a set of constraints among the variables involved in the abstraction pattern, which are interpreted as necessary conditions in order for the evidence $m_1(\mathbf{A}_1, T_1), \ldots, m_n(\mathbf{A}_n, T_n)$ to be abstracted into $h_\psi(\mathbf{A}_h, T_h^b, T_h^e)$; and $\pi(\mathbf{A}_1, T_1, \ldots, \mathbf{A}_n, T_n)$ is an observation procedure that gives as a result an observation of $h_\psi(\mathbf{A}_h, T_h^b, T_h^e)$ from a set of observations for $m_1(\mathbf{A}_1, T_1), \ldots, m_n(\mathbf{A}_n, T_n)$.

To illustrate this concept, consider the sequence of observations in Figure 1.1. Each of these observations is an instance of an observable we call *point* (p), represented as $p(\mathbf{A} = \{V\}, T)$, where T determines the temporal location of the observation and V is a value attribute.



V	3.4	17.6	12.9	2.6	-17.5	-20	-10.5	-0.8	7.8	17.5	19.4	19.4	17.6	7.8	-14.9	-16.3	-11.6	15.6	17.1	-15.8	7.7	13.7	2.7	-19.8	-19.6	-3.1	0.2	8.1	9.6	0
T	1	4	8	10	14	15	19	21	22	24	26	27	28	30	34	35	40	45	47	55	64	70	73	78	80	83	84	85	86	94

Figure 1.1: Initial temporal observations.

If we analyze these observations visually, we may hypothesize the presence of an underlying sinusoidal process. Let us define an observable *sinus* for such a sinusoidal process, with two attributes: the amplitude of the process (α) and its frequency (ω) —for simplicity, the phase component is dismissed—. The knowledge necessary to conjecture this hypothesis is collected in the following abstraction pattern:

$$h_{sinus}(\{\alpha, \omega\}, T_h^b, T_h^e) = \pi(V_1, T_1, \ldots, V_n, T_n)) \quad abstracts\ p(V_1, T_1), \ldots, p(V_n, T_n)$$
$$\{C(\alpha, \omega, T_h^b, T_h^e, V_1, T_1, \ldots, V_n, T_n)\}$$

We can estimate the attribute values $(\alpha, \omega, T_h^b, T_h^e)$ of this process by a simple observation procedure π, such that $\alpha = max(|V_i|)$, for $1 \leq i \leq n$, i.e., the amplitude α is obtained as the maximum absolute value of the observations; $\omega = \pi / mean(T_j^{peak} - T_{j-1}^{peak})$, where T_j^{peak} are *point* observations representing a peak, satisfying $(V_j^{peak} = V_k, T_j^{peak} = T_k) \wedge sign(V_k - V_{k-1}) \neq sign(V_{k+1} - V_k)$, so that the frequency ω is obtained as the inverse of the mean temporal separation between consecutive peaks in the sequence of observations; and $T_h^b = T_1, T_h^e = T_n$, i.e., the temporal support of the hypothesis is the time interval between the first and the last evidence points.

We can impose the following constraint $C(\alpha, \omega, T_h^b, T_h^e, V_1, T_1, \ldots, V_n, T_n)$ for every pair (V_i, T_i) in the sequence:

$$max(|\alpha \cdot sin(\omega \cdot T_i) - V_i|) \leq \varepsilon,$$

This constraint provides a model of a sinusoidal process and a measure of how well it fits a set of observations by means of a maximum error ε. Figure 1.2 shows the continuous representation of the abstracted process, whose resulting observation is $h_{sinus}(\alpha = 20, \omega = 0.3, T_h^b = 1, T_h^e = 94)$. A value of $\alpha/3$ has been chosen for ε.

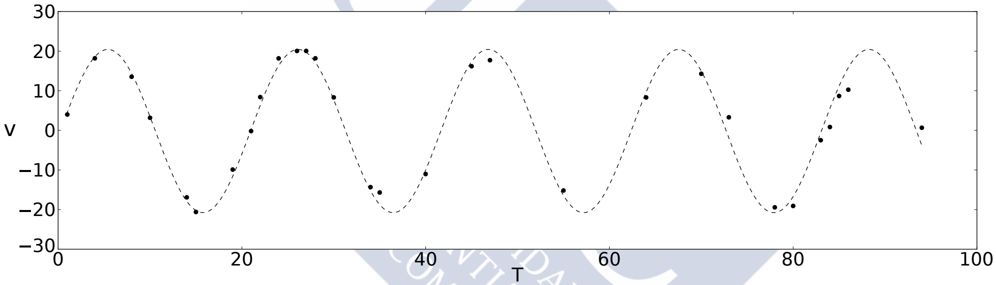


Figure 1.2: Abstracted sinusoidal process.

Of course, various observation procedures can be devised in order to estimate the same or different characteristics of the process being guessed. These procedures can provide one or several valid estimations in terms of their consistency with the abovementioned necessary constraints. In addition, different processes can be guessed from the same set of observations, all of them being valid in terms of their consistency. Hence, further criteria may be needed in order to rank the set of interpretations.

This simple example summarizes the common approach to the interpretation of experimental results in science and technology, when the knowledge is available as a model or a set of models. The challenge is to assume that this knowledge is not available in an analytical but in a declarative form, as a pattern or a set of patterns, and that the interpretation task is expected to mimic certain mechanisms of human perception.

1.2 Case study: ECG interpretation

A paradigmatic example of a time series interpretation problem in which humans show a higher performance than machines is the interpretation of electrocardiogram (ECG) signals, resulting from the recording on the body surface of the electrical activity of the heart as it changes along time. Usually this recording is performed simultaneously at different points on the skin, generating a set of time series, each one of them providing a different "perspective" of the heart activity. Since its proposal as a clinical tool in the late 19th century [Einthoven, 1893], ECG became soon a simple and low cost test recommended for the study of cardiopathies, and immediately elicited a lot of interest because of the mortality of cardiovascular diseases. Nowadays it is the primary method for the study and diagnosis of cardiac problems, including electrical conduction blocks, enlargement of heart muscle, insufficient blood flow, or heart rate regularity abnormalities, among others.

In the ECG signal we can identify several distinctive waveforms, corresponding to the electrical activation-recovery processes of the different parts of the heart. The P wave represents the activation of the atria, and it is the first wave of the cardiac cycle. The next group of waves recorded is the QRS complex, representing the simultaneous activation of the right and left ventricles. Finally, the wave that represents the ventricular recovery is called the T wave. An example of a normal ECG strip is shown in Figure 1.3, where these particular waveforms are highlighted in the second cardiac cycle.

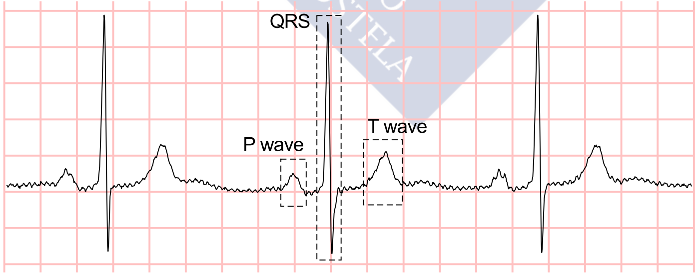


Figure 1.3: Basic waveforms of the ECG

Much of the information provided by the ECG is contained in the P wave, the QRS complex, and the T wave. The identification and analysis of these waveforms, considering their morphology, duration and amplitude, is the first necessary step to provide an effective interpretation of the processes taking place in the heart. But this analysis only provides information about the heart function during one heartbeat. There are pathophysiological processes that are manifested in a sequence of beats, showing particular rhythm patterns. Therefore, it is also necessary an analysis at a higher level, what is commonly called arrhythmia interpretation.

Since Pipberger et al. [1960] digitized for the first time the ECG and designed the first computer programs for its analysis, a lot of efforts have been made from multiple fields of science and engineering in order to explain the cardiac conduction system behavior from the ECG signal.

Learning to interpret the ECG involves the acquisition of perceptual skills from an extensive bibliography with interpretation criteria and worked examples. In particular, pattern recognition is specially important in order to build a bottom-up representation of the cardiac phenomena in multiple abstraction levels. Scientific literature has been gathering in the last years an increasing number of proposals for the computational analysis of the ECG signal, formulated as a set of classification problems, with one particular solution for each of them: beat detection, P wave identification, ST segment characterization, QT segment characterization, beat recognition by origin, or arrhythmia identification, among others. This wide range of proposals can be divided into two groups:

1. Knowledge based approaches: The provided solution tries to simulate the reasoning process performed by the cardiologist [Kors and Bemmel, 1990; Lavrac et al., 1985]. To achieve this, some kind of representation of the expert knowledge is adopted, most of the times based in a predicate logic. If the simple form of propositional logic is chosen the representation may be implemented by decision trees. On the other hand, the knowledge provided by the cardiologist is often affected by uncertainty, and it is commonly expressed in vague terms, which have led to a representation of the knowledge based on the fuzzy set theory [Bortolan and Degani, 1983]. Beyond the use of general models of knowledge representation, the proposal to encode different ECG patterns as symbol strings and the design of specific syntaxes through formal grammars for representing normal and abnormal patterns has achieved some significance. Thus, we find in the bibliography some classifiers formalized as regular and context-free grammars [Trahanias and Skordalakis, 1990].

2. Learning based approaches: The main aim is to obtain a classifier from some examples previously labeled by an expert, collected in a training set. Once the classifier has been

built, it has to be validated with a test set. These approaches are not incompatible with knowledge based approaches, since the learning process can end with a set of explicit classification criteria, for example in decision tree induction. But in most of learning based approaches, discriminating knowledge is not easily separable from discriminating algorithms. Here we can enumerate a large set of techniques from statistics, like Fischer Linear Discriminant Analysis, Bayesian Networks, Hidden Markov Models, Support Vector Machines; and techniques from artificial intelligence, like Artificial Neural Networks, Evolutionary Algorithms, or the Adaptive Resonance Theory [Clifford et al., 2006].

Another issue to be solved by any proposal for the computational analysis of the ECG is how to represent the ECG signal. There are three common solutions:

a) **Signal**. The classifier takes as input an ECG fragment as a digital signal, that is, a discrete-time signal. The signal may have been previously filtered to reduce artifacts like powerline noise or baseline drift, among others [Sörnmo and Laguna, 2005]. This representation ideally contains all the information from the physiological processes generating the ECG signal, but it is very exposed to the presence of noise and artifacts which are not always easy to remove.

b) **Features**. A segmentation of the ECG is performed to identify the constituent waveforms of each heartbeat: P, QRS, and T. Their amplitude and duration is measured, as long as some other important features like the PQ segment, the QT interval, or the ST segment deviation, among others [Laguna et al., 1994]. A classifier is then defined or trained to split this feature space into the target classes. The advantage of this representation is that it makes an abstraction on the signal in the same terms (waves, segments, beats) that the expert uses to interpret the ECG. In contrast, the presence of noise and artifacts often prevents a correct segmentation, hindering the posterior analysis.

c) **Bases of functions**. The first representation of the ECG by a base of functions uses the Laguerre orthogonal polynomials [Young and Huggins, 1963]. Later the Karhunen-Loeve expansion was used, which provides an optimal representation by means of the mean square error [Ahmed et al., 1975], and the wavelet transform, which allows representation in multiple resolution levels [Saxena et al., 2002]. Recently the Hermite orthogonal functions were proposed, which include a width parameter that provides an efficient representation of beats with significant differences in the QRS complex duration [Laguna et al., 1996]. In all these cases, the classifier takes as input a vector with the coefficients that allow to reconstruct the ECG signal from the combination of the basis functions. The main advantage of this representation is its robustness to noise and artifacts. However, the intuitive meaning of the classification criteria is lost, and this prevents its use unless the benefits are outstanding.

Most of the proposals we can find in the bibliography combine both knowledge-based and learning-based approaches, frequently in an heterogeneous way. Despite new proposals appear incessantly, computer interpretation of the ECG is still considered an open problem, and it seems we are far from provide sufficiently satisfactory solutions to be transferred to clinical routine, integrated in the bedside instrumentation or in the emergent home monitoring. A number of difficulties can be identified: 1) the variability of the physiological processes between different patients, or even in the same patient over time; 2) the stochastic nature of these processes; 3) the simultaneous occurrence of multiple physiological processes that can interact in different ways; 4) the presence of noise and artifacts in the signal which mask the physiological processes; and 5) the tacit, subjective, and hardly formalizable knowledge that constitutes the experience of the cardiologist.

In general, all the state-of-the-art approaches respond to the classification paradigm, so once the classifier is obtained it behaves as a logical level as a deductive system from a data vector representing the signal. As an evidence of our initial hypothesis, we will try to prove that temporal abductive reasoning provides a more appropriate framework for the computational interpretation of the ECG. Several arguments support this claim:

1. The complementariness between bottom-up and top-down processing is essential to provide robust explanations consistent with the evidence in noisy environments. A deductive arrhythmia classifier can not correctly identify a fragment in which the beats were not properly detected. An abductive classifier can conjecture the presence and morphology of a beat from its context, like a human can reconstruct a speech despite failing to identify all its constituent sounds.

2. An important limitation of the traditional approaches to the interpretation of the ECG is the assumption that all the physiological processes underlying the ECG signal are mutually exclusive, which is against all medical evidence. As discussed above, an abductive interpretation incorporates in a natural way several causes to the final explanation, while

a classical approximation needs to define a new class, different from the classes that identify separate physiological processes.

3. Most of the approaches found in the bibliography usually behave with a fixed picture of the classification parameters, putting aside temporal information that is of paramount importance for the correct interpretation of the ECG. This approach proves to be particularly wrong when ECG signal must be interpreted together with other signals, representing complex interactions among different physiological processes, and when the ECG signal must be interpreted with respect to contextual information, like drug administration. In both cases, temporal information is necessary in order to get a real picture of what is happening behind the scenes.

Along this thesis, the ECG domain was selected to illustrate all the concepts and notions defined by the proposed interpretation framework. Even though the reader has no experience in electrocardiography, the examples can be fully understood with the basic notions introduced in this section, and all the domain knowledge that is formalized in the abstraction patterns can be found in a common ECG handbook, such as [Wagner, 2008]. Both theoretical and practical results demonstrate how this framework can outperform well-established classical automatic approaches for ECG analysis, and even most that need the intervention of human experts to provide a final result.

1.3 Thesis outline

The rest of the work presented in this dissertation is organized as follows:

- Chapter 2 formalizes the proposed interpretation framework. This includes the definition of observables and observations as basic representation entities and abstraction patterns as the knowledge description primitives. Then, fundamental notions like abstraction relation, hypothesis, and interpretation are formalized, along with a definition of an interpretation problem and its solution from a static set-cover perspective. Finally, the computational complexity of the problem is studied, proving the NP-hardness of finding the optimal solution.

- Chapter 3 describes CONSTRUE, the proposed algorithm for the effective resolution of interpretation problems. This algorithm addresses an interpretation problem as a heuristic search problem, and it dynamically builds an interpretation by means of a hypothesize-and-test cycle that combines different bottom-up and top-down reasoning modes under the guidance of an attentional mechanism. First, an *off-line* version of the algorithm is described, where all the evidence is assumed to be available at the beginning of the interpretation. The description focuses on the hypothesize-and-test cycle, showing how this strategy is consistent with the intuition about how humans interpret temporal data in multiple abstraction levels. Then, the CONSTRUE-ONLINE algorithm is presented, which generalizes CONSTRUE by supporting the continuous acquisition of evidence during the interpretation. Also, a set of optimization strategies are presented to improve the efficiency of the interpretation, including a pruning mechanism to limit hypothesis exploration, a new search strategy exploiting necessary and sufficient abstraction conditions, and an efficient formalism for the representation of temporal knowledge.

- Chapter 4 presents the main contributions and results of this thesis through two experiments proving the practical capabilities of the framework by addressing two well-known problems in computer ECG processing: The first one is the QRS detection problem, which emphasizes the ability of a non-monotonic reasoning scheme to amend previous conclusions. The second problem is heartbeat classification by origin; we prove that an interpretation in multiple abstraction levels allows to describe the same set of abstract features used by human experts in the classification task, and then a simple set of qualitative rules is enough to provide better results than state-of-the-art automatic classifiers, even improving most of existing approaches that require the assistance of a human expert.

- In Chapter 5 the main theoretical aspects of the proposed interpretation framework are discussed and illustrated with real examples from the ECG domain, with emphasis in those features leading to overcoming some of the most important weaknesses of traditional deductive classifiers. These features are related to the combination of multiple hypotheses without requiring a casuistry-based heuristic, the ability to cope with ignorance and to look for missing evidence, and the interpretability of the reasoning process and the results. The chapter finishes with some semantical considerations on the main concepts defined by the framework.

- The conclusion chapter presents a synthesis of the key results and conclusions drawn from the present research, and also provides new directions for future research.
- Appendix A describes the full abstraction model defined for the interpretation of multi-lead ECG signals in multiple abstraction levels, including signal deviations, electrical activation-recovery of the heart chambers, and cardiac rhythms patterns. In addition, with the aim of supporting reproducible research, the full source code implementing this abstraction model and the interpretation algorithms has been published under an Open Source License.
- Appendix B describes an original algorithm for the delineation of QRS complexes in multi-lead ECG signals. This algorithm is based on a path-simplification strategy, and it is specially designed for working with noisy signals. It is used by the ECG abstraction model as an observation procedure in the QRS abstraction pattern.


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 2

A PROCESS-BASED INTERPRETATION FRAMEWORK

2.1 Definitions

This section formally defines the main pieces of the proposed interpretation framework: *observables* and *observations* for representing the interpretation of the system under study, and *abstraction patterns* for representing the knowledge about this system.

2.1.1 Representation entities

An *observation* is the result of applying an observation procedure. Such a procedure makes it possible to observe something with the quality of being *observable*. We call $\mathcal{Q} = \{q_0, q_1, \ldots, q_n\}$ the set of observables of a particular domain.

Definition 1. We define an **observable** as a tuple $q = \langle \psi, \mathbf{A}, T^b, T^e \rangle$, where ψ is a name representing the underlying process being observable, $\mathbf{A} = \{A_1, \ldots, A_{n_q}\}$ is a set of attributes to be valued, and T^b and T^e are two temporal variables representing the beginning and the end of the observable.

We call $V_q(A_i)$ the domain of possible values for the attribute A_i. We assume a discrete representation of the time domain τ, isomorphic to the set of natural numbers $\mathbb{N}$. For any observable, we implicitly assume the constraint $T^b < T^e$. In the case of an instantaneous observable, this is represented as $q = \langle \psi, \mathbf{A}, T \rangle$. Some observables can be dually represented from the temporal perspective, as either an observable supported by a temporal interval or as an observable supported by a temporal instant, according to the task to be carried out. A paradigmatic example is found in representing the heart beat, since it can be represented as a domain entity with a temporal extension comprising its constituent waves, and it can also be represented as an instantaneous entity for measuring heart rate.

Example 2.1.1: In the ECG signal, several distinctive waveforms can be identified, corresponding to the electrical activation-recovery cycle of the different heart chambers. The so-called P wave represents the activation of the atria, and is the first wave of the cardiac cycle. The next group of waves recorded is the QRS complex, representing the simultaneous activation of the right and left ventricles. Finally, the wave that represents the ventricular recovery is called the T wave. Together, these waveforms devise the characteristic pattern of the heart cycle, which is repeated in a normal situation with every beat [Wagner, 2008].
According to this description, the observable $q_{Pw} = \langle \texttt{atrial\_activation}, \{\texttt{amplitude}\}, T^b, T^e \rangle$ represents a P wave resulting from an atrial activation process with an unknown amplitude localized in a still unknown temporal interval.

Definition 2. We define an **observation** as a tuple $o = \langle q, \mathbf{v}, t^b, t^e \rangle$, an instance of the observable q resulting from applying an observation procedure that assigns a specific value to each attribute and to the temporal variables, where $\mathbf{v} = (v_1, \ldots, v_{n_q})$ is the set of attribute values, such that $\mathbf{v} \in V_q(A_1) \times \ldots \times V_q(A_{n_q})$; and t^b, $t^e \in \tau$ are two precise instants limiting the beginning and the end of the observation.

We also use the notation $(A_1 = v_1, \ldots, A_{n_q} = v_{n_q})$ to represent the assignment of values to the attributes of the observable and $T^b = t^b$ and $T^e = t^e$ for representing the assignment of temporal limits to the observation.

Example 2.1.2: The tuple $o = \langle q_{Pw}, \texttt{0.17mV}, \texttt{12:16.977}, \texttt{12:17.094} \rangle$ represents the particular occurrence of the P wave observable highlighted in Figure 2.1.

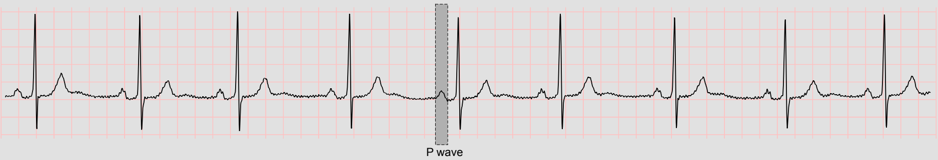


Figure 2.1: Example of a P wave observation. [Source: MIT-BIH arrhythmia DB [Goldberger et al., 2000], recording: 123, between 12:11.900 and 12:22.400]

Some notions involving observables and observations are defined below that will be useful in describing certain properties and constraints of the domain concepts, as well as in temporally arranging the interpretation process.

Definition 3. Given a set of observables $\mathcal{Q}$, a **generalization relation** can be defined between two different observables $q = \langle \psi, \mathbf{A}, T^b, T^e \rangle$ and $q' = \langle \psi', \mathbf{A}', T'^b, T'^e \rangle$, denoted by q' `is a` q, meaning that q generalizes q' if and only if $\mathbf{A} \subseteq \mathbf{A}'$ and $V_{q'}(A_i) \subseteq V_q(A_i)\ \forall A_i \in \mathbf{A}$.

The generalization relation is reflexive, antisymmetric and transitive. The inverse of a generalization relation is a specification relation, which captures the common notion of *Subtyping* in Type theory [Pierce, 2002]. From a logical perspective, a generalization relation can be read as an implication $q' \rightarrow q$, meaning that q' is more specific than q. It holds that every observation $o = \langle q', \mathbf{v}, t^b, t^e \rangle$ of the observable q' is also an observation of q.

Example 2.1.3: A common example of a generalization relation results from a domain partition of an attribute. For example, the observable $q_1 = \langle \texttt{Sinus\_Rhythm}, \{\texttt{RR} \in [200ms, 4000ms]\}, T^b, T^e \rangle$ is a generalization of the observables:

$$q_2 = \langle \texttt{Sinus\_Tachycardia}, \{\texttt{RR} \in [200ms, 600ms]\}, T^b, T^e \rangle$$
$$q_3 = \langle \texttt{Normal\_Rhythm}, \{\texttt{RR} \in [600ms, 1000ms]\}, T^b, T^e \rangle$$
$$q_4 = \langle \texttt{Sinus\_Bradycardia}, \{\texttt{RR} \in [1000ms, 4000ms]\}, T^b, T^e \rangle$$

The `RR` attribute represents the measure, in milliseconds, of the mean distance between consecutive beats, while q_2, q_3 and q_4 represent the normal cardiac rhythm denominations according to the heart rate [Wagner, 2008].

Example 2.1.4: Another usual way to define a generalization relation is by grouping common attributes of several observables in a more general observable. For example, the observable $q_5 = \langle \texttt{QRS}, \{\texttt{amplitude}, \texttt{shape}\}, T^b, T^e \rangle$, representing a QRS complex, generalizes:

$$q_6 = \langle \texttt{PacedQRS}, \{\texttt{amplitude}, \texttt{shape}, \texttt{spike}\}, T^b, T^e \rangle$$
$$q_7 = \langle \texttt{VentQRS}, \{\texttt{amplitude}, \texttt{shape}, \texttt{axis}\}, T^b, T^e \rangle$$

representing a paced QRS complex with an attribute referring the pacemaker spike and a ventricular QRS complex with an attribute describing the axis of the complex, respectively.

Definition 4. Given a set of observables $\mathcal{Q}$, an **exclusion relation** can be defined between two observables $q = \langle \psi, \mathbf{A}, T^b, T^e \rangle$ and $q' = \langle \psi', \mathbf{A}', T'^b, T'^e \rangle$, denoted by q `excludes` q', meaning that these are mutually exclusive and cannot be observed over two overlapping intervals, i.e., every two observations $o = \langle q, \mathbf{v}, t^b, t^e \rangle$ and $o' = \langle q', \mathbf{v}', t'^b, t'^e \rangle$ satisfy either $t^e < t'^b$ or $t'^e < t^b$.

The exclusion relation describes a temporal property affecting the observation of the system, and its explanation lies in the nature of the underlying processes and mechanisms. In the domain of electrocardiography, we assume that the exclusion relation is reflexive, symmetric and transitive, based on the nature of the electrical activation and recovery processes in cardiac physiology. Reflexivity represents that any of the physiological processes must end before it can start again; for example, a P wave cannot be observed over two overlapping intervals and, as a consequence, two such observations cannot be present in the same interpretation. Moreover, reflexivity makes it impossible for two different observations of the same observable to occur at the same time, that is, no attribute may take different values simultaneously: $(q = q') \wedge (\mathbf{v} \neq \mathbf{v}') \Rightarrow (t^e < t_j'^b) \vee (t_j'^e < t_i^b)$. This exclusion relation splits the set of observables $\mathcal{Q}$ into a set of equivalence classes, each one containing a set of observables corresponding to competing processes; for example, the equivalence class of heartbeat categories (normal, supraventricular, ventricular, paced, ...) brings together different, mutually exclusive types of beats according to their origin in the cardiac tissue.

It should be noted that the generalization relation may hide the exclusion relation, since the observation of multiple non-exclusive observables over overlapping intervals can appear as the observation of a single observable which is a generalization of the above in multiple overlapping intervals:

Example 2.1.5: The R-on-T phenomenon can be observed as an overlapping of two observations of the observable wave as a generalization of the R and T waves [Wagner, 2008]. An example is shown in Figure 2.2, where the last highlighted T wave overlaps the R wave of the next heartbeat.

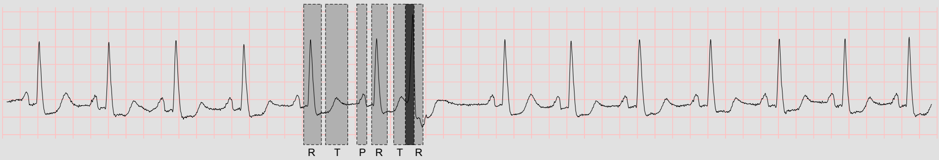


Figure 2.2: Example of the R-on-T phenomenon. [Source: MIT-BIH arrhythmia DB, recording: 105, between 00:10.800 and 00:21.300]

We call $\mathcal{O}$ the set of observations available for the observables in $\mathcal{Q}$. In order to index this set of observations, they will be represented as a sequence by defining an order relation between them. This ordering aims to prioritize the interpretation of the observations as they appear.

Definition 5. Let $<$ be an **order relation** between two observations $o_i = \langle q_i, \mathbf{v}_i, t_i^b, t_i^e \rangle$ and $o_j = \langle q_j, \mathbf{v}_j, t_j^b, t_j^e \rangle$ such that $(o_i < o_j) \Leftrightarrow (t_i^b < t_j^b) \vee ((t_i^b = t_j^b) \wedge (t_i^e < t_j^e)) \vee ((t_i^b = t_j^b) \wedge (t_i^e = t_j^e) \wedge (q_i < q_j))$, assuming a lexicographical order between observable names.

A sequence of observations is an ordered set of observations $\mathcal{O} = (o_1, \ldots, o_i, \ldots)$ where for all $i < j$ then $o_i < o_j$. Every subset of a sequence of observations is also a sequence. The q-sequence of observations from $\mathcal{O}$, denoted as $O(q)$, is the subset of the observations for the observable q. The exclusion relation forces that any two observations $o_i = \langle q, \mathbf{v}_i, t_i^b, t_i^e \rangle$ and $o_j = \langle q, \mathbf{v}_j, t_j^b, t_j^e \rangle$ in $O(q)$ satisfy $o_i < o_j \Rightarrow t_i^e < t_j^b$ for the current application domain. By $\mathtt{succ}(o_i)$ we denote the successor of the observation o_i in the sequence $\mathcal{O}$, according to the order relation $<$. By q-$\mathtt{succ}(o_i)$ we denote the successor of the observation $o_i \in O(q)$ in its q-sequence $O(q)$. Conversely to this notation, we denote by $q(o_i)$ the observable corresponding to the o_i observation.

2.1.2 Abstraction patterns

We model an abstraction process as an abduction process, based on the conjectural relation $m \leftarrow h$, which can be read as *'the observation of the finding m allows us to conjecture the observation of h as a possible explanatory hypothesis'*. For example, a very prominent peak in the ECG signal allows us to conjecture the observation of a heartbeat. A key aspect of the present proposal is that both the hypothesis and the finding are observables, and therefore formally identical, i.e., there exists $q_i, q_j \in \mathcal{Q}$, with $q_i \neq q_j$, such that $h \equiv q_i = \langle \psi_i, \mathbf{A}_i, T_i^b, T_i^e \rangle$ and $m \equiv q_j = \langle \psi_j, \mathbf{A}_j, T_j^b, T_j^e \rangle$. In general, an abstraction process can involve a number of different findings, even multiple findings of the same observable, and a set of constraints among them; thus, for example, a regular sequence of normal heartbeats allows us to conjecture the observation of a sinus rhythm. Additionally, an observation procedure is required in order to produce an observation of the hypothesis from the observation of those findings involved in the abstraction process.

We devise an abstraction process as a knowledge-based reasoning process, supported by the notion of abstraction pattern, which brings together those elements required to perform an abstraction. Formally:

Definition 6. An **abstraction pattern** $P = \langle h, M_P, C_P, \pi_P \rangle$ consists of a hypothesis h, a set of findings $M_P = \{m_1, \ldots, m_n\}$, a set of constraints $C_P = \{C_1, \ldots, C_t\}$ among the findings and the hypothesis, and an observation procedure $\pi_P(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_n, T_n^b, T_n^e) \in O(h)$.

Every constraint $C_i \in C_P$ is a relation defined on a subset of the set of variables taking part in the set of findings and the hypothesis $X = \{\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_n, T_n^b, T_n^e\}$. Thus, a constraint is a subset of the Cartesian product of the respective domains, and represents the simultaneously valid assignments to the variables involved. We will denote a set of constraints making reference to the set of variables being constrained, as in $C_P(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_n, T_n^b, T_n^e)$ for the abstraction pattern.

An abstraction pattern establishes, through the set C_P, those conditions for conjecturing the observation of h from a set of findings M_P, and through the observation procedure π_P, the calculations for producing a new observation $o_h \in O(h)$ from the observation of these findings. We call $M_P^q = \{m_1^q, m_2^q, \ldots, m_s^q\}$ the set of findings of the observable q in P, being $M_P = \bigcup_{q \in Q} M_P^q$. Thus, a set of findings allows the elements of a multiset of observables to be distinguished. The interpretation procedure will choose, as we will see later, from the available observations for every observable q satisfying the constraints C_P, which are to be assigned to the findings in M_P^q in order to calculate o_h.

The set of findings M_P is divided into two sets A_P and E_P, being $A_P \cap E_P = \varnothing$, where A_P is the set of findings that is said to be abstracted in o_h, and E_P is the set of findings that constitute the observation environment of o_h, that is, the set of findings needed to properly conjecture o_h, but which are not synthesized in o_h.

A temporal covering assumption can be made as a *default assumption* [Poole, 1990] on a hypothesis $h = \langle \psi_h, \mathbf{A}_h, T_h^b, T_h^e \rangle$ with respect to those findings $m = \langle \psi_m, \mathbf{A}_m, T_m^b, T_m^e \rangle$ appearing in an abstraction pattern:

Default Assumption 1. (*Temporal covering*) *Given an abstraction pattern P, it holds that $T_h^b \leq T_m^b$ and $T_m^e \leq T_h^e$, for all $m \in A_P \subseteq M_P$.*

The temporal covering assumption allows us to define the exclusiveness of an interpretation as the impossibility of including competing abstractions in the same interpretation.

Example 2.1.6: According to [CSE, 1985], in the electrocardiography domain a *"wave"* is a discernible deviation from a horizontal reference line called baseline, where at least two opposite slopes can be identified. The term discernible means that both the amplitude and the duration of the deviation must exceed some minimum values, agreed as 20 µV and 6 ms respectively. A wave can be completely described by a set of attributes: its amplitude (A), voltage polarity ($VP \in \{+,-\}$) and its main turning point T^{tp}, resulting in the following observable:

$$q_{wave} = \langle \texttt{electrical\_activity}, \{A, VP, T^{tp}\}, T^b, T^e \rangle$$

Let us consider the following abstraction pattern:

$$P_{wave} = \langle wave, M_P = \{m_0^{ECG}, \ldots, m_n^{ECG}\}, C_{P_{wave}}, \texttt{wave\_observation()} \rangle$$

where m_i^{ECG} is a finding representing an ECG sample, with a single attribute V_i representing the sample value and a temporal variable T_i representing its time point. We set the onset and end of a wave to the time of the second m_1^{ECG} and second-to-last m_{n-1}^{ECG} samples, considering m_0^{ECG} and m_n^{ECG} as environmental observations which are used to check the presence of a slope change just before and after the wave; thus $E_{P_{wave}} = \{m_0^{ECG}, m_n^{ECG}\}$, and $A_{P_{wave}} = \{m_1^{ECG}, \ldots, m_{n-1}^{ECG}\}$.
A set of temporal constraints are established between the temporal variables: $C_1 = \{T^e - T^b \geq 6ms\}$, $C_2 = \{T^b = T_1\}$, $C_3 = \{T^e = T_{n-1}\}$ and $C_4 = \{T^b < T^{tp} < T^e\}$. Another set of constraints limit the amplitude and slope changes of the samples included in a wave: $C_5 = \{sign(V_1 - V_0) \neq sign(V_2 - V_1)\}$, $C_6 = \{sign(V_n - V_{n-1}) \neq sign(V_{n-1} - V_{n-2})\}$, $C_7 = \{sign(V_{tp} - V_{tp-1}) = -sign(V_{tp+1} - V_{tp})\}$ and $C_8 = \{min\{|V_{tp} - V_1|, |V_{tp} - V_{n-1}|\} \geq 20\mu V\}$.
Once a set of ECG samples has satisfied these constraints, they support the observation of a wave: $o_{wave} = \langle q_{wave}, (a, vp, t^{tp}), t^b, t^e \rangle$. The values of t^b and t^e are completely determined by the constraints C_2 and C_3, while the observation procedure `wave_observa– tion()` provides a value for the attributes as follows: $vp = sign(V_{tp} - V_1)$, $a = max\{|V_{tp} - V_1|, |V_{tp} - V_{n-1}|\}$, and $t^{tp} = t^b + tp$, where $tp = arg\,min_k\{V_k | 1 \leq k \leq n-1\}$, if $V_1 < V_0$, or $tp = arg\,max_k\{V_k | 1 \leq k \leq n-1\}$, if $V_1 > V_0$. This pattern is illustrated in Figure 2.3.

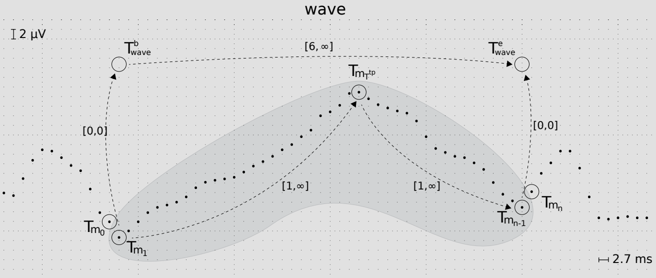


Figure 2.3: Wave abstraction pattern

2.1.3 Abstraction grammars

According to the definition, an abstraction pattern is defined over a fixed set of evidence findings M_P. In general, however, an abstraction involves an undetermined number of pieces of evidence -in the case of an ECG wave, the number of samples-. Hence we provide a procedure for dynamically generating abstraction patterns, based on formal language theory. The set $\mathcal{Q}$ of observables can be considered as an alphabet. Given an alphabet $\mathcal{Q}$, the special symbols $\varnothing$ (empty set), and λ (empty string), and the operators $|$ (union), $\cdot$ (concatenation), and $*$ (Kleene closure), a formal grammar G denotes a pattern of symbols of the alphabet, describing a language $L(G) \subseteq \mathcal{Q}^*$ as a subset of the set of possible strings of symbols of the alphabet.

Let G^{ap} be the class of formal grammars of abstraction patterns. An *abstraction grammar* $G \in G^{ap}$ is syntactically defined as a tuple (V_N, V_T, H, R). For the production rules in R the expressiveness of right-linear grammars is adopted [Hopcroft et al., 2001]:

$$\begin{aligned} H &\rightarrow qD \\ D &\rightarrow qF \mid q \mid \lambda \end{aligned}$$

H is the initial symbol of the grammar, and this plays the role of the hypothesis guessed by the patterns generated by G. V_N is the set of non-terminal symbols of the grammar, satisfying $H \in V_N$, although H cannot be found on the right-hand side of any production rule, since a hypothesis cannot be abstracted by itself. V_T is the set of terminal symbols of the grammar, representing the set of observables $Q_G \subseteq \mathcal{Q}$ that can be abstracted by the hypothesis.

Given a grammar $G \in G^{ap}$, we devise a constructive method for generating a set of abstraction patterns $P_G = \{P_1, \ldots, P_i, \ldots\}$. Since a formal grammar is simply a syntactic specification of a set of strings, every grammar $G \in G^{ap}$ is semantically extended to an attribute grammar [Aho et al., 2006], embedded with a set of actions to be performed in order to incrementally build an abstraction pattern by the application of production rules. An abstraction grammar is represented as $G = ((V_N, V_T, H, R), B, BR)$, where $B(\alpha)$ associates each grammar symbol $\alpha \in V_N \cup V_T$ with a set of attributes, and $BR(r)$ associates each rule $r \in R$ with a set of attribute computation rules. An abstraction grammar associates the following attributes: *i)* $P(attern)$, with each non-terminal symbol of the grammar; this will be assigned an abstraction pattern; *ii)* $A(bstracted)$, with each terminal symbol corresponding to an observable $q \in Q_G$; this allows us to assign each finding either to the set A_P or E_P, depending on its value of true or false; *iii)* $C(onstraint)$, with each terminal symbol corresponding to an observable; this will be assigned a set of constraints. There are proposals in the bibliography dealing with BNF descriptions of Constraint Satisfaction Problems and their semantic expression in different formalisms [Barro et al., 1994; Félix et al., 1999]; however, we will focus on the result as a set of constraints among the variables involved.

In the following, the set of attribute computation rules associated with the grammar productions is specified to provide a formal method for building abstraction patterns $P \in P_{G_h}$ from a grammar $G_h \in G^{ap}$. P_{G_h} gathers the set of abstraction patterns that share the same observable h as a hypothesis; thus, these represent the different ways to conjecture h. Using this method, the application of every production incrementally adds a new observable as a finding and a set of constraints between this finding and previous entities, as follows:

1. The initial production $H \to qD$ entails:

$$
\begin{aligned}
&P_H := \langle h, M_H = \varnothing, C_H = \varnothing, \pi_H = \varnothing \rangle \\
&C_q := C(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, T_1^b, T_1^e) \\
&A_q \in \{true, false\} \\
&P_D := \langle h, M_D = M_H \cup \{m_1^q\}, C_D = C_H \cup C_q, \pi_D(\mathbf{A}_1, T_1^b, T_1^e) \rangle
\end{aligned}
$$

2. All the productions of the form $D \to qF$ entail:

$$
\begin{aligned}
&P_D := \langle h, M_D, C_D, \pi_D(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_k, T_k^b, T_k^e) \rangle \\
&C_q := C(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, \ldots, \mathbf{A}_{k+1}, T_{k+1}^b, T_{k+1}^e) \\
&A_q \in \{true, false\} \\
&P_F := \langle h, M_F = M_D \cup \{m_{k+1}^q\}, C_F = C_D \cup C_q, \pi_F(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_{k+1}, T_{k+1}^b, T_{k+1}^e) \rangle
\end{aligned}
$$

3. Productions of the form $D \to q$ conclude a string of terminal symbols, and consequently the generation of a pattern $P \in P_{G_h}$:

$$
\begin{aligned}
&P_D := \langle h, M_D, C_D, \pi_D(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_k, T_k^b, T_k^e) \rangle \\
&C_q := C(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, \ldots, \mathbf{A}_{k+1}, T_{k+1}^b, T_{k+1}^e) \\
&A_q \in \{true, false\} \\
&P := \langle h, M_P = M_D \cup \{m_{k+1}^q\}, C_P = C_D \cup C_q, \pi_P(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_{k+1}, T_{k+1}^b, T_{k+1}^e) \rangle
\end{aligned}
$$

4. Productions of the form $D \to \lambda$ also conclude a string and the generation of a pattern:

$$
\begin{aligned}
&P_D := \langle h, M_D, C_D, \pi_D(\mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_k, T_k^b, T_k^e) \rangle \\
&P := P_D
\end{aligned}
$$

This constructive method enables the incremental addition of new constraints as new findings are included in the representation of the abstraction pattern, providing a dynamic mechanism for knowledge assembly by language generation. Moreover, it is possible to design an adaptive observation procedure as new evidence becomes available, since the observation procedure may be different at each step.

In the case that no temporal constraints are attributed to a production, a 'hereafter' temporal relationship will be assumed by default to exist between the new finding and the set of previous findings. For instance, a production of the form $D \rightarrow qF$ entails that $C_F = C_P \cup \{T_i^b \leq T_{k+1}^b \mid m_i \in M_P\}$.

Hence, in the absence of any temporal constraint, an increasing temporal order among consecutive findings in every abstraction pattern is assumed. Moreover, every temporal constraint must be consistent with this temporal order.

According to the limitation imposed on observations of the same observable which prevents two different observations from occurring at the same time, an additional constraint is added on any two findings of the same observable, and thus $\forall m_i^q, m_j^q \in M_P^q, i < j \rightarrow T_i^e < T_j^b$.

Several examples of abstraction pattern grammars modeling common knowledge in electrocardiography are given below, in order to illustrate the expressiveness of the G^{ap} grammars.

Example 2.1.7: The grammar $G_N = (V_N, V_T, H, R)$ is designed to generate an abstraction pattern for a *normal cardiac cycle*, represented by the observable q_N, including the descriptions of common durations and intervals [Wagner, 2008]. In this grammar, $V_N = \{H, D, E\}$, $V_T = \{q_{Pw}, q_{QRS}, q_{Tw}\}$, and R is given by:

$$
\begin{array}{ll}
H \rightarrow q_{Pw}D & \{P_H := \langle q_N, M_H=\varnothing, C_H=\varnothing, \pi_H=\varnothing\rangle \\
 & C_{Pw} := \{T_N^b=T_{Pw}^b;\ 50ms \leq T_{Pw}^e - T_{Pw}^b \leq 120ms\} \\
 & A_{Pw} := true \\
 & P_D := \langle q_N, M_D=\{m^{Pw}\}, C_D=C_{Pw}, \pi_D=\varnothing\rangle \\
 & \} \\
D \rightarrow q_{QRS}E & \{P_D := \langle q_N, M_D=\{m^{Pw}\}, C_D=C_{Pw}, \pi_D=\varnothing\rangle \\
 & C_{QRS} := \{50ms \leq T_{QRS}^e - T_{QRS}^b \leq 150ms;\ 100ms \leq T_{QRS}^b - T_{Pw}^b \leq 210ms\} \\
 & A_{QRS} := true \\
 & P_E := \langle q_N, M_E=M_D \cup \{m^{QRS}\}, C_E=C_D \cup C_{QRS}, \pi_E=\varnothing\rangle \\
 & \}
\end{array}
$$

$$E \rightarrow q_{Tw} \qquad \{P_E{:=}\ \langle q_N, M_E{=}\{m^{Pw}, m^{QRS}\}, C_E, \pi_E{=}\varnothing\rangle$$

$$C_{Tw}{:=}\ \{80ms{\le}T^b_{Tw}{-}T^e_{QRS}{\le}120ms;\ T^e_{Tw}{-}T^b_{QRS}{\le}520ms;\ T^e_N{=}T^e_{Tw}\}$$

$$A_{Tw}{:=}\ true$$

$$P{:=}\ \langle q_N, M_P{=}M_E{\cup}\{m^{Tw}\}, C_P{=}C_E{\cup}C_{Tw}, \pi_P{=}\varnothing\rangle$$

$$\}$$

This grammar generates a single abstraction pattern, which allows us to interpret the sequence of a P wave, a QRS complex, and a T wave as the coordinated contraction and relaxation of the heart muscle, from the atria to the ventricles. Some additional temporal constraints are required and specified in the semantic description of the corresponding productions. In this case, an observation procedure π is not necessary since the attributes of the hypothesis are completely determined by the constraints in the grammar, and do not require additional calculus. Figure 2.4 shows the abstraction pattern and the full set of constraints.

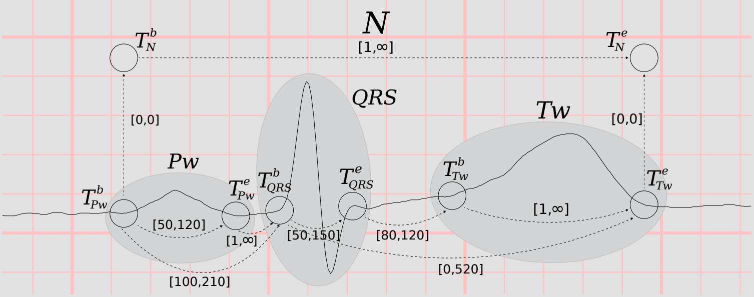


Figure 2.4: Illustration of the *Normal cardiac cycle* abstraction pattern

The next example shows the ability of an abstraction grammar to generate abstraction patterns dynamically with an undefined number of findings.

Example 2.1.8: A *bigeminy* is a heart arrhythmia in which there is a continuous alternation of long and short heart beats. Most often this is due to ectopic heart beats occurring so frequently that there is one after each normal beat, typically premature ventricular contractions (PVCs) [Wagner, 2008]. For example, a normal beat is followed by a PVC, which is then followed by a pause. The normal beat then returns, only to be followed by another PVC. The grammar $G_{VB} = (V_N, V_T, H, R)$ generates a set of abstraction patterns for *ventricular bigeminy*, where $V_N = \{H, D, E, F\}$, $V_T = \{q_N, q_V\}$, and R is given by:

$H \rightarrow q_N D$

$\{P_H := \langle q_{VB}, M_H = \emptyset, C_H = \emptyset, \pi_H = \emptyset \rangle$

$C_N := \{T^b_{VB} = T_1\}$

$A_N := true$

$P_D := \langle q_{VB}, M_D = \{m^N_1\}, C_D = C_N, \pi_D = \emptyset \rangle$

$\}$

$D \rightarrow q_V E$

$\{P_D := \langle q_{VB}, M_D = \{m^N_1\}, C_D = C_N, \pi_D = \emptyset \rangle$

$C_V := \{200ms \leq T_2 - T_1 \leq 800ms\}$

$A_V := true$

$P_E := \langle q_{VB}, M_E = M_D \cup \{m^V_2\}, C_E = C_D \cup C_V, \pi_E = \emptyset \rangle$

$\}$

$E \rightarrow q_N F$

$\{P_E := \langle q_{VB}, M_E = \{m^N_1, \ldots, m^V_{k-1}\}, C_E, \pi_E = \emptyset \rangle$

$C_N := \{1.5 \cdot 200ms \leq T_k - T_{k-1} \leq 4 \cdot 800ms\}$

$A_N := true$

$P_F := \langle q_{VB}, M_F = M_E \cup \{m^N_k\}, C_F = C_E \cup C_N, \pi_F = \emptyset \rangle$

$\}$

$F \rightarrow q_V E$

$\{P_F := \langle q_{VB}, M_F = \{m^N_1, m^V_2, \ldots, m^N_k\}, C_F, \pi_F = \emptyset \rangle$

$C_V := \{200ms \leq T_{k+1} - T_k \leq 800ms\}$

$A_V := true$

$P_E := \langle q_{VB}, M_E = M_F \cup \{m^V_{k+1}\}, C_E = C_F \cup C_V, \pi_F = \emptyset \rangle$

$\}$

$F \rightarrow q_V$

$\{P_F := \langle q_{VB}, M_F = \{m^N_1, m^V_2, \ldots, m^N_{n-1}\}, C_F, \pi_F = \emptyset \rangle$

$C_V := \{200ms \leq T_n - T_{n-1} \leq 800ms;\ T^e_{VB} = T_n\}$

$A_V := true$

$P := \langle q_{VB}, M_P = M_F \cup \{m^V_n\}, C_P = C_F \cup C_V, \pi_P = \emptyset \rangle$

$\}$

For simplicity, we have referenced each N and V heart beat with a single temporal variable. Thus T_i represents the time point of the ith heart beat, and is a normal beat if i is odd, and a PVC if i is even. With the execution of these production rules, an unbounded sequence of alternating normal and premature ventricular QRS complexes is generated, described above as ventricular bigeminy. Note that in terms of the $\{N, V\}$ symbols the G_{VB} grammar is syntactically equivalent to the regular expression $NV(NV)^+$.

In this example, as in 2.1.7, an observation procedure π_P is not necessary, since the constraints in the grammar completely determine the temporal endpoints of the hypothesis and there are no more attributes to be valued. Figure 2.5 shows an example of a ventricular bigeminy pattern.

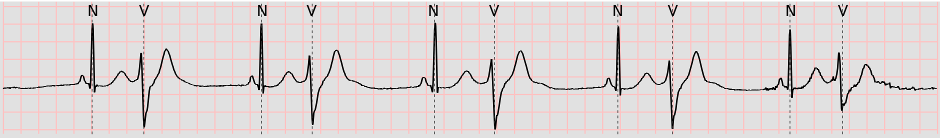


Figure 2.5: Example of ventricular bigeminy. [Source: MIT-BIH arrhythmia DB, recording: 106, between 25:06.350 and 25:16.850]

2.2 An interpretation framework

In this section, we define and characterize an interpretation problem. Informally, an interpretation problem arises from the availability of a set of initial observations from a given system, and of domain knowledge formalized as a set $\mathcal{G} = \{G_{q_1}, \ldots, G_{q_n}\}$ of G^{ap} grammars. Every abstraction grammar $G_h \in \mathcal{G}$ generates a set of abstraction patterns that share the same hypothesis h. The whole set of abstraction patterns that can be generated by $\mathcal{G}$ is denoted by $\mathcal{P}$.

Definition 7. Let $\mathcal{Q}$ be a set of observables and $\mathcal{G}$ a set of abstraction grammars. We say $\mathcal{G}$ induces an **abstraction relation** in $\mathcal{Q} \times \mathcal{Q}$, denoted by $q_i \ltimes q_j$ if and only if there exists an abstraction pattern P generated by some $G_h \in \mathcal{G}$ such that:

1. $q_j = h$
2. $M_P^{q_i} \cap A_P \neq \varnothing$
3. $q_i \not\ltimes^+ q_i$, where $\ltimes^+$ is the transitive closure of $\ltimes$

The relation $q_i \ltimes q_j$ is a sort of conjectural relation that allows us to conjecture the presence of q_j from the observation of q_i. The transitive closure of the abstraction relation is a strict partial order relation between the domain observables, such that $q_i < q_j \Leftrightarrow q_i \ltimes^+ q_j$; that is, if and only if $\exists q_{k_0}, \ldots, q_{k_n} \in \mathcal{Q}$ such that $q_{k_0} = q_i$, $q_{k_n} = q_j$ and for all m, with $0 \leq m < n$, it holds that $q_{k_m} \ltimes q_{k_{m+1}}$. We denote by $q_i = q_{k_0} \ltimes q_{k_1} \ltimes \ldots \ltimes q_{k_n} = q_j$ an *abstraction sequence* in n steps that allows the conjecture of q_j from q_i. This order relation defines an abstraction hierarchy among the observables in $\mathcal{Q}$. From the definition of a strict partial order, there must be at the base of this hierarchy at least one observable we call q_0, corresponding in the domain of electrocardiography to the digital signal.

Example 2.2.1: Let $\mathcal{Q} = \{q_{Pw}, q_{QRS}, q_{Tw}, q_N, q_V, q_{VB}\}$ and $\mathcal{G} = \{G_N, G_{VB}\}$, containing the knowledge represented in examples 2.1.7 and 2.1.8. From the derived abstraction relation it follows that $q_{Pw}, q_{QRS}, q_{Tw} \mathrel{|\!\!<} q_N$, and $q_N, q_V \mathrel{|\!\!<} q_{VB}$. Intuitively, we can see that this relation splits the observables into three abstraction levels: the wave level, describing the activation/recovery of the different heart chambers; the heartbeat level, describing each cardiac cycle by its origin in the muscle tissue; and the rhythm level, describing the dynamic behavior of the heart over multiple cardiac cycles. These levels match those commonly used by experts in electrocardiogram analysis [Wagner, 2008].

It is worth noting that the abstraction relation is only established between observables in the A_P set. This provides flexibility in defining the evidence forming the context of a pattern, as this may belong to different abstraction levels.

Definition 8. We define an **abstraction model** as a tuple $\mathcal{M} = \langle \mathcal{Q}, \mathrel{|\!\!<}, \mathcal{G} \rangle$, where $\mathcal{Q}$ is the set of domain observables, $\mathrel{|\!\!<}$ is an abstraction relation between such observables, and $\mathcal{G}$ is the available knowledge as a set of abstraction grammars.

The successive application of the available abstraction grammars results in a series of observations organized in a hierarchy of abstraction, according to the order relation between observables as described above. We are able to define an interpretation problem as follows.

Definition 9. We define an **interpretation problem** as a tuple $IP = \langle \mathcal{O}, \mathcal{M} \rangle$, where $\mathcal{O} = (o_1, o_2, \ldots, o_i, \ldots)$ is a sequence of observations requiring interpretation and $\mathcal{M}$ is an abstraction model of the domain.

It is worth mentioning that this definition of an abductive interpretation problem differs from the common definition of an abductive diagnosis problem, where the difference between normal and faulty behaviors is explicit, leading to the role of faulty manifestations. Only when a faulty manifestation is detected is the abductive process of diagnosis started. In contrast, in the present framework all the observations have the same status, and the objective of the interpretation process is to provide an interpretation of what is observed at the highest possible abstraction level in terms of the underlying processes. As we will see later, some observables may stand out amongst others regarding the efficiency of the interpretation process, as salient features that can draw some sort of perceptual attention.

As discussed above, any observable $q \in Q_P$ can appear multiple times as different pieces of evidence for an abstraction pattern P, in the form of findings collected in the set M_P. As a consequence, P can predict multiple observations of the set $\mathcal{O}$ for a given observable $q \in Q_P$, each of these corresponding to one of the findings of the set M_P through a matching relation. This matching relation is a matter of choice for the agent in charge of the interpretation task, by selecting from the evidence the observation corresponding to each finding in a given pattern.

Definition 10. Given an interpretation problem IP, a **matching relation** for a pattern $P \in \mathcal{P}$ is an injective relation in $M_P \times \mathcal{O}$, defined by $m^q \leftarrowtail o$ if and only if $o = \langle q, \mathbf{v}, t^b, t^e \rangle \in O(q) \subseteq \mathcal{O}$ and $m^q = \langle \psi, \mathbf{A}, T^b, T^e \rangle \in M_P$, such that $(A_1 = v_1, \ldots, A_{n_q} = v_{n_q})$, $T^b = t^b$ and $T^e = t^e$.

A matching relation makes an assignment of a set of observations to a set of findings of a certain pattern, leading us to understand the interpretation problem as a search within the available evidence for a valid assignment for the constraints represented in an abstraction pattern.

From the notion of matching relation we can design a mechanism for abductively interpreting a subset of observations in $\mathcal{O}$ through the use of abstraction patterns. Thus, a matching relation for a given pattern allows us to hypothesize new observations from previous ones, and to iteratively incorporate new evidence into the interpretation by means of a hypothesize-and-test cycle. The notion of abstraction hypothesis defines those conditions that a subset of observations must satisfy in order to be abstracted by a new observation, and makes it possible to incrementally build an interpretation from the incorporation of new evidence.

Definition 11. Given an interpretation problem IP, we define an **abstraction hypothesis** as a tuple $\hbar = \langle o_h, P, \leftarrowtail \rangle$, where $P = \langle h, M_P, C_P, \pi_P \rangle \in \mathcal{P}$, $\leftarrowtail \subseteq M_P \times \mathcal{O}$, and we denote $O_\hbar = codomain(\leftarrowtail)$, satisfying:

1. $o_h \in O(h)$.
2. $o_h = \pi_P(O_\hbar)$.
3. $C_P(\mathbf{A}_h, T_h^b, T_h^e, \mathbf{A}_1, T_1^b, T_1^e, \ldots, \mathbf{A}_n, T_n^b, T_n^e)|_{o_h, o_j \in O_\hbar}$ is satisfied.

These conditions entail: (1) an abstraction hypothesis guesses an observation of the observable hypothesized by the pattern; (2) a new observation is obtained from the application of the observation procedure to those observations being assigned to the set of findings M_P by the matching relation; and (3) the observations taking part in an abstraction hypothesis must satisfy those constraints of the pattern whose variables are assigned a value by the observations.

Example 2.2.2: Figure 2.6 shows an example of an abstraction hypothesis using the normal cardiac cycle pattern described by the grammar in example 2.1.7. The matching relation established between the three observations and the findings predicted by the pattern allows to conjecture a new observation of the normal cardiac cycle observable.

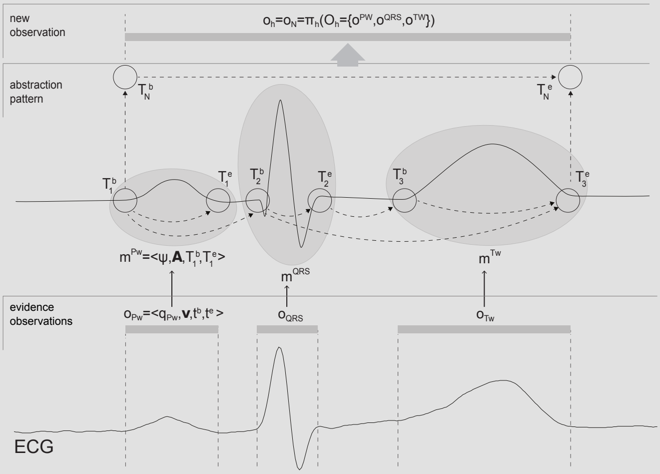


Figure 2.6: Example of an abstraction hypothesis.

Even though the matching relation is a matter of choice, and therefore a conjecture in itself, some additional constraints may be considered as default assumptions. An important default assumption in the abstraction of periodic processes states that consecutive observations are related by taking part in the same hypothesis, defining the basic period of the process. This assumption functions as a sort of operative hypothesis of the abstraction task:

Default Assumption 2. (*Basic periodicity) Periodic findings in an abstraction pattern must be assigned consecutive observations by any matching relation:*

$$\forall m_i^q, m_{i+1}^q \in M_P^q, m_i^q \leftarrowtail o_j \wedge \mathtt{q{-}succ}(o_j) \in O_\hbar \Rightarrow m_{i+1}^q \leftarrowtail \mathtt{q{-}succ}(o_j)$$

This default assumption allows us to avoid certain combinations of abstraction hypotheses that, although formally correct, are meaningless from an interpretation point of view. For example, without the assumption of basic periodicity, a normal rhythm fragment might be abstracted by two alternating bradycardia hypotheses, as shown in Figure 2.7.

The set of observations that may be abstracted in an interpretation problem is $O(domain(\Bbbk))$, that is, observations corresponding to observables involved in the set of findings to be abstracted

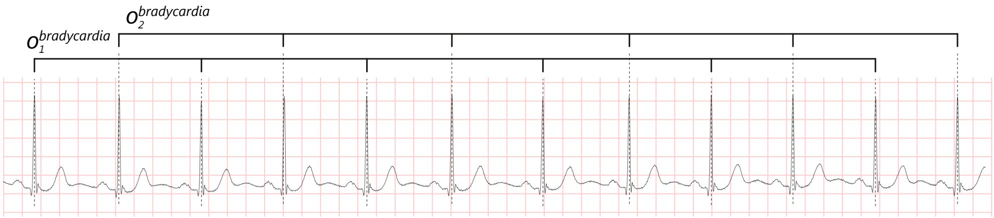


Figure 2.7: Motivation for the assumption of basic periodicity. [Source: MIT-BIH arrhythmia DB, recording: 103, between 00:40.700 and 00:51.200]

by some abstraction pattern. An abstraction hypothesis defines in the set of observations $\mathcal{O}$ a counterpart of the subsets A_P and E_P of the set of findings M_P of a pattern P, resulting from the selection of a set of observations $O_\hbar \subseteq \mathcal{O}$ by means of a matching relation, satisfying those requirements shown in the definition 11.

Definition 12. Given an abstraction hypothesis $\hbar = \langle o_h, P, \leftarrowtail \rangle$, we define the following sets of observations:

- $\texttt{abstracted\_by}(o_h) = \{o \in O_\hbar \mid m_i^q \leftarrowtail o \wedge m_i^q \in A_P\}$.
- $\texttt{environment\_of}(o_h) = \{o \in O_\hbar \mid m_i^q \leftarrowtail o \wedge m_i^q \in E_P\}$.
- $\texttt{evidence\_of}(o_h) = \texttt{abstracted\_by}(o_h) \cup \texttt{environment\_of}(o_h)$.

We denote by $\texttt{abstracted\_by}(o_h)$ the set of observations abstracted by o_h and which are somehow its constituents, while $\texttt{environment\_of}(o_h)$ denotes the evidential context of o_h. We denote by $\texttt{evidence\_of}(o_h)$ the set of all observations supporting a specific hypothesis. Since the matching relation is injective, it follows that $\texttt{abstracted\_by}(o_h) \cap \texttt{environment\_of}(o_h) = \varnothing$.

The definition of these sets can be generalized to include as arguments a set of observations $O = \{o_{h_1}, \ldots, o_{h_m}\}$ from a set of abstraction hypotheses $\hbar_1, \ldots, \hbar_m$:

- $\texttt{abstracted\_by}(O) = \bigcup_{o_h \in O} \texttt{abstracted\_by}(o_h)$
- $\texttt{environment\_of}(O) = \bigcup_{o_h \in O} \texttt{environment\_of}(o_h)$.
- $\texttt{evidence\_of}(O) = \bigcup_{o_h \in O} \texttt{evidence\_of}(o_h)$.

As a result of an abstraction hypothesis, a new observation o_h is generated which can be included in the set of domain observations, so that $\mathcal{O} = \mathcal{O} \cup \{o_h\}$. In this way, an interpretation can be incrementally built from the observations, by means of the aggregation of abstraction hypotheses.

Definition 13. Given an interpretation problem IP, an **interpretation** is defined as a set of abstraction hypotheses $I = \{\hbar_1, \ldots, \hbar_m\}$.

An interpretation can be rewritten as $I = \langle O_I, P_I, \leftarrow\!\!\prec_I \rangle$, where $O_I = \{o_{h_1}, \ldots, o_{h_m}\}$ is the set of observations guessed by performing multiple abstraction hypotheses; $P_I = \{P_1, \ldots, P_m\}$ is the set of abstraction patterns used in the interpretation; and $\leftarrow\!\!\prec_I = \leftarrow\!\!\prec_{\hbar_1} \cup \ldots \cup \leftarrow\!\!\prec_{\hbar_m} \subseteq (M_1 \cup \ldots \cup M_m) \times \mathcal{O}$ is the global matching relation. We should note that the global matching relation $\leftarrow\!\!\prec_I$ is not necessarily injective, since some observations may simultaneously belong to both the `abstracted_by`() and `environment_of`() sets of different observations.

From a given interpretation problem IP, multiple interpretations can be abductively proposed through different sets of abstraction hypotheses. Indeed, the definition of interpretation is actually weak, since even an empty set $I = \varnothing$ is formally a valid interpretation. Thus, we need additional criteria in order to select the solution to the interpretation problem as the best choice among different possibilities [Peng and Reggia, 1990].

Definition 14. Given an interpretation problem IP, an interpretation I is a **cover** of IP if the set of observations to be interpreted $O(domain(\mathrel{|\!\!<})) \subseteq \mathcal{O}$ is included in the set of observations abstracted by I, that is, $O(domain(\mathrel{|\!\!<})) \subseteq \texttt{abstracted\_by}(O_I)$.

Definition 15. Given an interpretation problem IP, two different abstraction hypotheses $\hbar$ and $\hbar'$ of the mutually exclusive observables q_h and $q_{h'}$ are **alternative hypotheses** if and only if $\texttt{abstracted\_by}(o_h) \cap \texttt{abstracted\_by}(o_{h'}) \neq \varnothing$.

Example 2.2.3: A *ventricular trigeminy* is an infrequent arrhythmia very similar to *ventricular bigeminy*, except that the ectopic heart beats occur after every pair of normal beats instead of after each one. The grammar for hypothesizing a ventricular trigeminy q_{VT} would therefore be very similar to that described in example 2.1.8, with the difference that each q_V finding would appear after every pair of q_N findings. These two processes are mutually exclusive, insofar as the heart can develop just one of these activation patterns at a given time. For this reason, in the event of an observation of q_V, this may be abstracted by either a q_{VB} or a q_{VT} hypothesis, but never by both simultaneously.

Definition 16. Given an interpretation problem IP, a cover I for IP is **exclusive** if and only if it contains no alternative hypotheses.

Thus, two or more different hypotheses of mutually exclusive observables abstracted from the same observation will be incompatible in the same interpretation, since inferring both a statement and its negation is logically prevented, and therefore only one of them can be selected.

On the other hand, a parsimony criterion is required, in order to disambiguate the possible interpretations to select as the most plausible those of which the complexity is minimum [Peng and Reggia, 1990]. We translate this minimum complexity in terms of minimal cardinality.

Definition 17. Given an interpretation problem IP, a cover I for IP is **minimal**, if and only if its cardinality is the smallest among all covers for IP.

Minimality introduces a parsimony criterion on hypothesis generation, promoting temporally maximal hypotheses, that is, those hypotheses of a larger scope rather than multiple equivalent hypotheses of smaller scope. For example, consider an abstraction pattern that allows the conjecture of a regular cardiac rhythm from the presence of three or more consecutive heart beats. Without a parsimony criterion, a sequence of nine consecutive beats could be abstracted by up to three consecutive rhythm observations, even when a single rhythm observation would be sufficient and better.

Definition 18. The **solution** of an interpretation problem IP is the set of all minimal and exclusive covers of IP.

2.3 Computational complexity

Abduction, generally described as inference to the best explanation, has been formulated under different frameworks according to the task to be addressed, but has always been found an intractable problem in the general case [Josephson and Josephson, 1994]. The next theorem proves that an interpretation problem is also an intractable problem.

Theorem 1. *Finding the solution to an interpretation problem is NP-hard.*

Proof: We will provide a polynomial-time reduction of the well-known set covering problem to an interpretation problem. Given a set of elements $U = \{u_1, \ldots, u_m\}$ and a set S of subsets of U, a cover is a set $C \subseteq S$ of subsets of S whose union is U. In terms of complexity analysis, two different problems of interest are identified:

- A set covering decision problem, stating that given a pair (U,S) and an integer k the question is whether there is a set covering of size k or less. This decision version of set covering is NP-complete.
- A set covering optimization problem, stating that given a pair (U,S) the task is to find a set covering that uses the fewest sets. This optimization version of set covering is NP-hard.

We will therefore reduce the set covering problem to an interpretation problem by means of a polynomial-time function φ. Thus, we shall prove that $\varphi(U,S)$ is an interpretation problem, and there is a set covering of $\varphi(U,S)$ of size k or less if and only if there is a set covering of U in S of size k or less.

Given a pair (U,S), let $\varphi(U,S) = \langle \mathcal{O}, \mathcal{M} \rangle$ where:

1. $\mathcal{O} = U = \{u_1, \ldots, u_m\}$, such that $u_i = \langle q, true, i \rangle$ and $q = \langle \psi, present, T \rangle$.
2. $\mathcal{M} = \langle Q, \mathrel{|\!<}, \mathcal{P} \rangle$, such that $domain(\mathrel{|\!<}) = q$.
3. $\forall s = \{u_{i_1}, \ldots, u_{i_n}\} \in S$, $\exists P \in \mathcal{P}$, being $P = \langle q_P, M_P, C_P, \pi_P \rangle$, where:
   - $q \mathrel{|\!<} q_P$ and $P \neq P' \Rightarrow q_P \neq q_{P'}$.
   - $M_P = A_P = M_P^q = \{m_1^q = \langle \psi, present, T_1 \rangle, \ldots, m_n^q\}$.
   - $C_P = \{\bigwedge_{k=1}^{n}(m_k^q.T = k); T_h^b = min\{m_k^q.T\}; T_h^e = max\{m_k^q.T\}\}$.
   - $q_P.present = \pi_P(m_1^q, \ldots, m_n^q) = \bigwedge_{k=1}^{n} m_k^q.present$.

Thus, $\varphi(U,S)$ is an interpretation problem according to this definition. On the other hand, $\varphi(U,S)$ can be built in polynomial time. In addition, for all $s \in S$ there exists an abstraction hypothesis $\hbar = \langle o_h, P, \leftarrowtail \rangle$ such that:

1. $o_h = \langle h, true, min_{u_i \in s}\{i\}, max_{u_i \in s}\{i\} \rangle$.
2. $u_i \in s \Rightarrow u_i \in codomain(\leftarrowtail)$.
3. $\leftarrowtail$ provides a valid assignment, since the set of observations satisfying $\pi_P = true$ also satisfies the constraints in C_P.

Since each abstraction hypothesis involves a different abstraction pattern there are no alternative hypotheses in any interpretation of $\varphi(U,S)$.

Suppose there is a set covering $C \subseteq S$ of U of size k or less. For all $u \in U$ there exists $c_i \in C - \{\varnothing\}$ such that $u \in c_i$ and, by the above construction, there exists $\hbar_i \in I$ such that $\texttt{abstracted\_by}(o_{h_i}) = \{u \in codomain(\leftarrowtail_{\hbar_i})\} = \{u \in c_i\} = c_i$, and therefore, $O(domain({\mid}\!{<})) \subseteq \bigcup_{\hbar_i \in I} \texttt{abstracted\_by}(o_{h_i}) = \bigcup_i c_i = C$. That is, the set of abstraction hypotheses I is an exclusive cover of the interpretation problem $\varphi(U,S)$ of size k or less.

Following the same reasoning as for the set covering optimization problem, finding a minimal and a exclusive cover of an interpretation problem $\varphi(U,S)$ is NP-hard, since we can use the solution of this problem to check whether there is an exclusive cover of the interpretation problem of size k or less, and this has been proven above to be NP-complete. ☐


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CHAPTER 3

INTERPRETATION ALGORITHMS

The solution set for an interpretation problem IP consists of all exclusive covers of IP having the minimum possible number of abstraction hypotheses. Obtaining this solution set can be stated as a search on the set of interpretations of IP. The major source of complexity of searching for a solution is the local selection, from the available evidence in $\mathcal{O}$, of the most appropriate matching relation for a number of abstraction hypotheses that can globally shape a minimal and exclusive cover of IP.

Nevertheless, the whole concept of solution must be revised in practical terms, due to the intractability of the task and the incompleteness of the abstraction model, that is, of the available knowledge. Indeed, we assume that any realistic abstraction model can hardly provide a cover for every possible interpretation problem. Hence the objective should shift from searching for a solution to searching for an approximate solution.

Certain principles applicable to the interpretation problem can be exploited in order to approach a solution in an iterative way, bounding the combinatorial complexity of the search. These principles can be stated as a set of heuristics that make it possible to evaluate and discriminate some interpretations against others from the same base evidence:

- A *coverage principle*, which states the preference for interpretations explaining more initial observations.
- A *simplicity principle*, which states the preference for interpretations with fewer abstraction hypotheses.

- An *abstraction principle*, which states the preference for interpretations involving higher abstraction levels.
- A *predictability principle*, which states the preference for interpretations that properly predict future evidence.

The coverage and simplicity principles are used to define a cost measure for the heuristic search process [Edelkamp and Schrödl, 2011], while the abstraction and predictability principles are used to guide the reasoning process, in an attempt to emulate the same shortcuts used by humans.

In order to search for solutions of an interpretation problem we will exploit the behavior of a formal grammar as a language generator. Let us assume for simplicity that only observations can evoke a hypothesis in the reasoning; this excludes the possibility of temporal information from triggering hypothesis formation. From this point of view, temporal information is only used in consistency checking and to delimit the observation interval for a predicted finding.

An interpretation procedure must be designed according to the temporal constraints over the interpretation task: distinguishing between an *off-line* interpretation, performed on a collection of time series previously recorded, and *on-line* interpretation, performed on a system while it is operating and it is being monitored [Brusoni et al., 1998]. In its on-line version, interpretation must be performed continuously, usually under some temporal constraints on when the results of the interpretation are needed, e.g., arrhythmia identification must be done in time, to apply effective treatment. These temporal constraints define a temporal window specifying the maximum delay in providing an interpretation from the monitoring time.

3.1 Off-line interpretation

Given an interpretation problem IP, a heuristic vector for a certain interpretation I can be defined to guide the search, as $\varepsilon(I) = (1.0 - \varsigma(I), \kappa(I))$, where $\varsigma(I) = |\mathtt{abstracted\_by}(O_I)| / |O(domain(\Bbbk))|$ is the *covering ratio* of I, and $\kappa(I) = |O_I|$ is the *complexity* of I, a counterpart of simplicity. The main goal of the search strategy is to approach a solution with a maximum covering ratio and a minimum complexity, which is equivalent to the minimization of the heuristic vector. The covering ratio will be considered the primary heuristic, and complexity will be considered for ranking interpretations with the same covering ratio. The $\varepsilon(I)$ heuristic is intuitive and very easy to calculate, but as a counterpart it is a non-admissible heuristic, since it is not monotone and may underestimate or overestimate the true goal covering. Therefore optimality cannot be guaranteed and we require an algorithm efficient with this type of heuristic. We propose the CONSTRUE() algorithm, whose pseudocode is shown in Algorithm 1. This algorithm is a variation of the K-Best First Search algorithm [Felner et al., 2003], with partial expansion to reduce the number of explored nodes.

Algorithm 1 CONSTRUE search algorithm.

```
function CONSTRUE($IP$)
    var $I_0 = \varnothing$
    var $K = max(|\{q_j \in \mathcal{Q} \mid q_i \,|\!< q_j, q_i \in \mathcal{Q}\}|)$
    SET_FOCUS($I_0, o_1$)
    var $open$ = SORTED($[\langle \varepsilon(I_0), I_0 \rangle]$)
    var $closed$ = SORTED($[]$)
    while $open \neq \varnothing$ do
        for all $I \in open[0 \ldots K]$ do
            $I'$ = NEXT(GET_DESCENDANTS($I$))
            if $I'$ is null then
                $open = open - \{\langle \varepsilon(I), I \rangle\}$
                $closed = closed \cup \{\langle \varepsilon(I), I \rangle\}$
            else if $\varsigma(I') = 1.0$ then
                return $I'$
            else
                $open = open \cup \{\langle \varepsilon(I'), I' \rangle\}$
            end if
        end for
    end while
    return $min(closed)$
end function
```

The CONSTRUE() algorithm takes as its input an interpretation problem *IP*, and returns the first interpretation found with full coverage, or the interpretation with the maximum covering ratio and minimum complexity if no covers are found, using the abstraction and predictability principles in the searching process. To do this, it manages two ordered lists of interpretations, named *open* and *closed*. Each interpretation is annotated with the computed values of the heuristic vector. The *open* list contains those partial interpretations that can further evolve by (1) appending new hypotheses or (2) extending previously conjectured hypotheses to subsume or predict new evidence. This *open* list is initialized with the trivial interpretation $I_0 = \varnothing$. The *closed* list contains those interpretations that cannot explain more evidence.

At each iteration, the algorithm selects the K most promising interpretations according to the heuristic vector (line 8), and partially expands each one of them to obtain the next descendant node I'. If this node is a solution, then the process ends by returning it (line 13), otherwise it is added to the *open* list. The partial expansion ensures that the *open* list grows at each iteration by at most K new nodes, in order to save memory. When a node cannot expand further, it is added to the *closed* list (line 12), from which the solution is taken if no full coverages are found (line 20).

The selection of a value for the K parameter depends on the problem at hand. We select its value as $K = max(|\{q_j \in \mathcal{Q} \mid q_i \Bbbk q_j, q_i \in \mathcal{Q}\}|)$, that is, as the maximum number of observables that can be abstracted from any observable q_i. The intuition behind this choice is that at any point in the interpretation process, and with the same heuristic values, the same chance is given to any plausible abstraction hypothesis in order to explain a certain observation.

In order to expand the current set of interpretations, the GET_DESCENDANTS() function relies on different reasoning modes, that is, different forms of abduction and deduction, which are brought into play under the guidance of an attentional mechanism. Since searching for a solution finally involves the election of a matching relation, both observations and findings should be included in the scope of this mechanism. Hence, a focus of attention can be defined to answer the following question: which is the next observation or finding to be processed? The answer to this question takes the form of a hypothesize-and-test cycle: if the attention focuses on an observation, then an abstraction hypothesis explaining this observation should be generated (hypothesize); however, if the attention focuses on a finding predicted by some hypothesis, an observation should be sought to match such finding (test). Thus, the interpretation problem is solved by a reasoning strategy that progresses incrementally over time, coping with new evidence through the dynamic generation of abstraction patterns from a finite number of abstraction grammars, and bounding the theoretical complexity by a parsimony criterion.

To illustrate and motivate the reasoning modes implemented in building interpretations and supporting the execution of the CONSTRUE() algorithm, we use a simple, but complete, interpretation problem.

Example 3.1.1: Let $\mathcal{Q} = \{q_{wave}, q_{Pw}, q_{QRS}, q_{Tw}, q_N\}, \mathcal{G} = \{G_w, G_N, G_{Tw}\}$, where G_w models the example 2.1.6, G_N is described in example 2.1.7, and $G_{Tw} = (\{H, D\}, \{q_{QRS}, q_{wave}\}, H, R)$ describes the knowledge to conjecture a T wave with the following rules:

$$\begin{array}{ll}
H \rightarrow q_{QRS}D & \{P_H := \langle q_{Tw}, M_H = \emptyset, C_H = \emptyset, \pi_H = \emptyset \rangle \\
 & C_{QRS} := \{80ms \leq T^b_{Tw} - T^e_{QRS} \leq 120ms;\ T^e_{Tw} - T^b_{QRS} \leq 520ms\} \\
 & A_{QRS} := false \\
 & P_D := \langle q_{Tw}, M_D = \{m^{QRS}\}, C_D = C_{QRS}, \pi_D = \emptyset \rangle \\
 & \} \\
D \rightarrow q_{wave} & \{P_D := \langle q_{Tw}, M_D = \{m^{QRS}\}, C_D = C_{QRS}, \pi_D = \emptyset \rangle \\
 & C_{wave} := \{T^b_{Tw} = T^b_{wave};\ T^e_{Tw} = T^e_{wave};\ max(diff(sig[m^{wave}])) \leq 0.7 \cdot max(diff(sig[m^{QRS}]))\} \\
 & A_{wave} := true \\
 & P := \langle q_{Tw}, M_P = M_D \cup \{m^{wave}\}, C_P = C_D \cup C_{wave}, \pi_P = \texttt{Tw\_delin}(T^b_{QRS}, T^e_{QRS}, T^b_{wave}, T^e_{wave}) \rangle \\
 & \}
\end{array}$$

This grammar hypothesizes the observation of a T wave from a wave appearing shortly after the observation of a QRS complex, requiring a significant decrease in the maximum slope of the signal (in the constraint definition C_{wave}, the expression "$max(diff(sig[m]))$" stands for the maximum absolute value of the derivative of the ECG signal between T^b_m and T^e_m). Let us note that the observation of a T wave is not deduced from the observation of a QRS complex since it is not a necessary condition for the observation of a QRS complex. The observation procedure of the generated pattern is denoted as `Tw_delin()`, and may be any of the methods described in the literature for the delineation of T waves, such as in [Laguna et al., 1994].

In addition to the P_{wave} pattern generated by G_w and detailed in example 2.1.6, G_N and G_{Tw} generate the following abstraction patterns:

$$P_N = \langle q_N, A_{P_N} = \{m^{Pw}, m^{QRS}, m^{Tw}\} \cup E_{P_N} = \emptyset, C_{P_N}, \pi_{P_N} = \emptyset \rangle$$
$$P_{Tw} = \langle q_{Tw}, A_{P_{Tw}} = \{m^{wave}\} \cup E_{P_{Tw}} = \{m^{QRS}\}, C_{QRS} \cup C_{wave}, \texttt{Tw\_delin()} \rangle$$

Finally, let $\mathcal{O} = \{o^{wave}_1 = \langle q_{wave}, \emptyset, 0.300, 0.403 \rangle,\ o^{wave}_2 = \langle q_{wave}, \emptyset, 0.463, 0.549 \rangle,\ o^{Pw} = \langle q_{Pw}, \emptyset, 0.300, 0.403 \rangle,\ o^{QRS} = \langle q_{QRS}, \emptyset, 0.463, 0.549 \rangle\}$ be a set of initial observations including a P wave and a QRS complex abstracting two wave observations located at specific time points.

Given this interpretation problem, Figure 3.1 at page 64 shows the starting point for the interpretation, where the root of the interpretation process is the trivial interpretation I_0, and the attention is focused on the first observation. The sequence of reasoning steps towards the resolution of this interpretation problem will be explained in the following subsections.

3.1.1 Focus of attention

The focus of attention is modeled as a stack; thus, once the focus is set on a particular observation (or finding), any observation that was previously under focus will not return to be focused on until the reasoning process on the current observation is finished. Algorithm 2 shows how the different reasoning modes are invoked based on the content of the focus of attention, resulting in a hypothesize-and-test cycle.

Algorithm 2 Method for obtaining the descendants of an interpretation using different reasoning modes based on the content of the focus of attention.

```
 1: function GET_DESCENDANTS($I$)
 2:     var $focus$ = GET_FOCUS($I$).TOP()
 3:     var $desc = \emptyset$
 4:     if IS_OBSERVATION($focus$) then
 5:         if $focus = o_h \mid \hbar \in I$ then
 6:             $desc$ = DEDUCE($I, focus$)
 7:         end if
 8:         $desc = desc \cup$ ABDUCE($I, focus$) $\cup$ ADVANCE($I, focus$)
 9:     else if IS_FINDING($focus$) then
10:         $desc$ = SUBSUME($I, focus$) $\cup$ PREDICT($I, focus$)
11:     end if
12:     return $desc$
13: end function
```

Lines 4-8 generate the descendants of an interpretation I when there is an observation at the top of the stack. These descendants are the result of two possible reasoning modes: the deduction of new findings, performed by the DEDUCE() function, provided that the observation being focused on is an abstraction hypothesis; and the abduction of a new hypothesis explaining the observation being focused on, performed by the ABDUCE() function. A last descendant is obtained using the ADVANCE() function, which simply restores the previous focus of attention by means of a POP() operation. If the focus is then empty, ADVANCE() inserts the next observation to explain, which may be selected by temporal order in the general case, or by some domain-dependent saliency criterion to prioritize certain observations over others. By removing the observation at the top of the focus of attention, the ADVANCE() function sets aside that observation as unintelligible in the current interpretation, according to the available knowledge.

If the top of the stack contains a finding, then Algorithm 2 obtains the descendants of the interpretation from the SUBSUME() and PREDICT() functions (line 10). The first of these functions looks for an existing observation satisfying the constraints on the finding focused on, while the second makes predictions about observables that have not yet been observed. All of these reasoning modes are described separately and detailed below; we will illustrate how the CONSTRUE() algorithm combines these in order to solve the interpretation problem in Example 3.1.1.

3.1.2 Building an interpretation: Abduction

Algorithm 3 enables the abductive generation of new abstraction hypotheses. It is applied when the attention is focused on an observation that can be abstracted by some abstraction pattern, producing a new observation at a higher level of abstraction.

Algorithm 3 Moving forward an interpretation through abduction.

```
 1: function ABDUCE($I, o_i$)
 2:     var $desc = \varnothing$
 3:     for all $G_h = \langle V_N, V_T, H, R \rangle \in \mathcal{G} \mid q(o_i) |\!< h$ do
 4:         for all $(U \rightarrow qV) \in R \mid q(o_i)$ is_a $q \wedge A_q = true$ do
 5:             $P_V = \langle h, M_V = \{m^q\}, C_V = C_q, \pi_V \rangle$
 6:             $\hbar = \langle o_h, P_V, \leftarrowtail_{\hbar} = \{m^q \leftarrowtail o_i\} \rangle$
 7:             $L_{\hbar} = [(U \rightarrow qV)]; B_{\hbar} = U; E_{\hbar} = V$
 8:             $I' = \langle O_I \cup \{o_h\}, P_I \cup \{P_V\}, \leftarrowtail_I \cup \leftarrowtail_{\hbar} \rangle$
 9:             $\mathcal{O} = \mathcal{O} \cup \{o_h\}$
10:             GET_FOCUS($I'$).POP()
11:             GET_FOCUS($I'$).PUSH($o_h$)
12:             $desc = desc \cup \{I'\}$
13:         end for
14:     end for
15:     return $desc$
16: end function
```

The result of ABDUCE() is a set of interpretations I', each one adding a new abstraction hypothesis with respect to the parent interpretation I. To generate these hypotheses, we iterate through those grammars that can make a conjecture from the observation o_i under focus (line 3). Then, for each grammar, each production including the corresponding observable $q(o_i)$ (line 4) initializes an abstraction pattern with a single finding of this observable (line 5), and a new hypothesis is conjectured with a matching relation involving both the observation under focus and the finding (line 6). A list structure $L_\hbar$ and two additional variables $B_\hbar$ and $E_\hbar$ are initialized to trace the sequence of productions used to generate the findings in the abstraction pattern; these will play an important role in subsequent reasoning steps (line 7). Finally the new hypothesis opens a new interpretation (lines 8-9) focused on this hypothesis (line 11).

In this way, the ABDUCE() function implements, from a single piece of evidence, the hypothesize step of the hypothesize-and-test cycle. Below we explain the reasoning modes involved in the test step of the cycle.

Example 3.1.2: Let us consider the interpretation problem set out in example 3.1.1 and the interpretation I_0 shown in Figure 3.1. According to Algorithm 2, the ABDUCE() function is used to move forward the interpretation, since the focus of attention points to an observation o^{Pw}. The abstraction pattern that supports this operation is P_N, and a matching relation is established with the m^{Pw} finding. As a result, the following hypothesis is generated:

$$\hbar_1 = \langle o^N, P_N, \{m^{Pw} \leftarrowtail o^{Pw}\}\rangle$$

Figure 3.1 shows the result of this reasoning process, in a new interpretation called I_1. Note that the focus of attention has been moved to the newly created hypothesis (lines 10-11 of the ABDUCE() function).

3.1.3 Building an interpretation: Deduction

This reasoning mode is applied when the attention is focused on an observation o_h previously conjectured as part of an abstraction hypothesis $\hbar$ (see Algorithm 4). The DEDUCE() function takes the evidence that has led to conjecture o_h and tries to extend it with new findings which can be expected, i.e., deduced, from the abstraction grammar G_h used to guess the observation. The key point is that this deduction process follows an iterative procedure, as the corresponding abstraction pattern is dynamically generated from the grammar. Hence the DEDUCE() function aims to extend a partial matching relation by providing the next finding to be tested, as part of the test step of the hypothesize-and-test cycle.

Since the first finding leading to conjecture o_h does not necessarily appear at the beginning of the grammar description, the corresponding abstraction pattern will not, in general, be generated incrementally from the first production of the grammar. Taking as a starting point the production used to conjecture o_h (line 4 in Algorithm 3), the goal is to add a new finding by applying a new production at both sides, towards the beginning and the end of the grammar, using the information in the $L_\hbar$ list. The $B_\hbar$ variable represents the non-terminal at the left-

Algorithm 4 Moving forward an interpretation through deduction of new findings.

```
 1: function DEDUCE($I, o_h$)
 2:     var $desc = \varnothing$
 3:     if $B_{\hbar} \neq H$ then
 4:         for all $(X \rightarrow qB_{\hbar}) \in R$ do
 5:             $P_{B_{\hbar}} = \langle h, M_{B_{\hbar}} = \{m^q\}, C_{B_{\hbar}} = C_q, \pi_{B_{\hbar}} \rangle$
 6:             for all $(U \rightarrow q'V) \in L_{\hbar}$ do
 7:                 $P_V = \langle h, M_U \cup \{m^{q'}\}, C_U \cup C_{q'}, \pi_V \rangle$
 8:             end for
 9:             $\hbar = \langle o_h, P_{E_{\hbar}}, \leftarrow\!\prec_{\hbar} \rangle$
10:             $I' = \langle O_I, P_I \cup \{P_{E_{\hbar}}\}, \leftarrow\!\prec_I \rangle$
11:             INSERT($L_{\hbar}, (X \rightarrow qB_{\hbar}), begin$);$B_{\hbar} = X$
12:             GET_FOCUS($I'$).PUSH($m^q$)
13:             $desc = desc \cup \{I'\}$
14:         end for
15:     else
16:         for all $(E_{\hbar} \rightarrow qX) \in R$ do
17:             $P_X = \langle h, M_{E_{\hbar}} \cup \{m^q\}, C_{E_{\hbar}} \cup C_q, \pi_X \rangle$
18:             $\hbar = \langle o_h, P_X, \leftarrow\!\prec_{\hbar} \rangle$
19:             $I' = \langle O_I, P_I \setminus \{P_{E_{\hbar}}\} \cup \{P_X\}, \leftarrow\!\prec_I \rangle$
20:             INSERT($L_{\hbar}, (E_{\hbar} \rightarrow qX), end$);$E_{\hbar} = X$
21:             GET_FOCUS($I'$).PUSH($m^q$)
22:             $desc = desc \cup \{I'\}$
23:         end for
24:     end if
25:     return $desc$
26: end function
```

hand side of the first production in $L_{\hbar}$, while $E_{\hbar}$ represents the non-terminal at the right-hand side of the last production in $L_{\hbar}$. Hence, this list has the form $L_{\hbar} = [(B_{\hbar} \rightarrow q'V'), (V' \rightarrow q''V''), \ldots, (V'^{n-1} \rightarrow q'^{n}E_{\hbar})]$. In case $L_{\hbar}$ is empty, both variables $B_{\hbar}$ and $E_{\hbar}$ represent the H non-terminal. With this information the sequence of findings supporting the hypothesis $\hbar$ can be updated in two opposite directions:

- Towards the beginning of the grammar (lines 3-14): we explore the set of observables that may occur before the first finding according to the productions of the grammar (line 4), and a new finding is deduced for each of these in different descendant interpretations. A new pattern $P_{B_{\hbar}}$ associated with the $B_{\hbar}$ non-terminal is initialized with the new finding (line 5), and by moving along the sequence of productions generating the previous set of findings (lines 6-8) the pattern associated to the rightmost non-terminal $P_{E_{\hbar}}$ is updated with a new set of findings containing m^q. Consequently, the hypothesis and the interpretation are also updated (lines 9 and 10), and the applied production is inserted at the beginning of $L_{\hbar}$ (line 11). Finally the newly deduced finding is focused on (line 12).

- Towards the end of the grammar (lines 15-23): for each one of the observables that may occur after the last finding, a new finding m^q is deduced, expanding the abstraction pattern associated with the new rightmost non-terminal X. After updating the hypothesis $\hbar$, the previous pattern $P_{E_{\hbar}}$ in the resulting interpretation I' is replaced by the new one, P_X, and the applied production is inserted at the end of $L_{\hbar}$. Finally, the new finding is focused on (line 21).

Example 3.1.3: Let us consider the interpretation problem set out in example 3.1.1 and the interpretation I_1 shown in Figure 3.1. Remember that the grammar used to generate the hypothesis in the focus of attention, G_N, has the following form:

$$H \to q_{Pw}D$$
$$D \to q_{QRS}E$$
$$E \to q_{Tw}$$

In this situation, it is possible to deduce new findings from the o^N hypothesis. Following Algorithm 3 we can check that $B_{\hbar} = H$ and $E_{\hbar} = D$, since the only finding in the matching relation is m^{Pw}. Deduction then has to be performed after this last finding, using the production $D \to q_{QRS}E$. After constraint checking, the resulting finding is as follows:

$$m^q_{n+1} = m^{QRS} = \langle q_{QRS}, \varnothing, T^b_{QRS} \in [0.400, 0.520], T^e_{QRS} \in [0.450, 0.660] \rangle$$

Figure 3.1 illustrates the outcome of this reasoning process and the uncertainty in the temporal limits of the predicted finding, which is now focused on in the interpretation I_2.

3.1.4 Building an interpretation: Subsumption

Subsumption is performed when the attention is focused on a finding previously deduced from some abstraction grammar (see Algorithm 5). This reasoning mode avoids the generation of a new hypothesis for every piece of available evidence if it can be explained by a previous hypothesis. The SUBSUME() function explores the set of observations $\mathcal{O}$ and selects those consistent with the constraints on the finding in the focus of attention (line 3), expanding the matching relation of the corresponding hypothesis in different descendant interpretations (line 4). The focus of attention is then restored to its previous state (line 5), allowing the deduction of new findings from the same hypothesis. The SUBSUME() function clearly enforces the simplicity principle.

Algorithm 5 Moving forward an interpretation through subsumption.

```
1: function SUBSUME($I, m_i$)
2:     var $desc = \varnothing$
3:     for all $o_j \in \mathcal{O} \mid m_i \leftarrow\!\prec o_j$ do
4:         $I' = \langle O_I, P_I, \leftarrow\!\prec_I \cup \{m_i \leftarrow\!\prec o_j\}\rangle$
5:         GET_FOCUS($I'$).POP($m_i$)
6:         $desc = desc \cup \{I'\}$
7:     end for
8:     return $desc$
9: end function
```

Example 3.1.4: Let us consider the interpretation I_2 shown in Figure 3.1. If we apply the subsumption procedure, it is possible to set a matching relation between o^{QRS} and m^{QRS}, since this observation satisfies all the constraints on the finding. The result is shown in the interpretation I_3. Note that the uncertainty in the end time of the o^N hypothesis is now reduced after the matching, having $T_N^e \in [0.631, 1.030]$. Following this, the attention focuses once again on this hypothesis, and a new deduction operation may therefore be performed.

3.1.5 Building an interpretation: Prediction

This reasoning mode is also performed when the attention is focused on a finding deduced from some abstraction grammar (see Algorithm 6). In this case, if a finding previously deduced has not yet been observed, it will be predicted.

The goal of the PREDICT() function is to conjecture a new observation to match the focused finding. For this, the abstraction model is explored and those grammars whose hypothesized observable is more specific than the predicted observable are selected (line 3). Then, a new pattern is initialized with no evidence supporting it, and a new abstraction hypothesis with an empty matching relation is generated (lines 4-5). Finally, the attention focuses on the observation being guessed (lines 9-10) to enable the DEDUCE() function to start a new test step at a lower abstraction level. Since $L_\hbar$ is initialized as an empty list (line 6), $B_\hbar$ and $E_\hbar$ point to

Algorithm 6 Moving forward an interpretation through the prediction of non-available evidence.

```
function PREDICT($I, m_i$)
    var $desc = \varnothing$
    for all $G_h = \langle V_N, V_T, H, R \rangle \in \mathcal{G} \mid h$ is_a $q(m_i)$ do
        $P_H = \langle h, M_H = \varnothing, C_H = \varnothing, \pi_H = \varnothing \rangle$
        $\hbar = \langle o_h, P_H, \leftarrowtail_{\hbar} = \varnothing \rangle$
        $L_{\hbar} = \varnothing; B_{\hbar} = E_{\hbar} = H$
        $I' = \langle O_I \cup \{o_h\}, P_I \cup \{P_H\}, \leftarrowtail_I \cup \{m_i \leftarrowtail o_h\} \rangle$
        $\mathcal{O} = \mathcal{O} \cup \{o_h\}$
        GET_FOCUS($I'$).POP($m_i$)
        GET_FOCUS($I'$).PUSH($o_h$)
        $desc = desc \cup \{I'\}$
    end for
    return $desc$
end function
```

the initial symbol of the grammar, and the corresponding abstraction pattern will be generated only towards the end of the grammar.

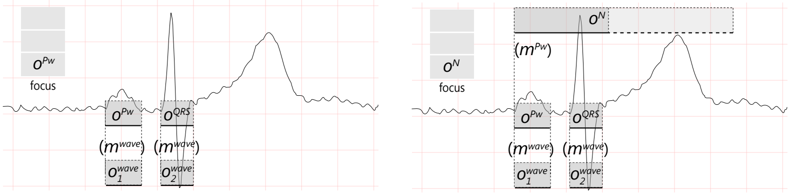
Example 3.1.5:
Starting from the I_3 interpretation shown in Figure 3.1, the next step we can take to move forward the interpretation is a new deduction on the o^N hypothesis, generating a new finding m^{Tw} and leading to the I_4 interpretation. Since there is no available observation of the T wave, a matching with this new finding m^{Tw} cannot be made by the SUBSUME() function, thus, the only option for moving forward this interpretation is through prediction. Following the PREDICT() function, the G_{Tw} grammar can be selected, and a new observation o^{Tw} can be conjectured, generating the I_5 interpretation.
From I_5 we can continue the deduction on the o^{Tw} hypothesis. If we apply the DEDUCE() function we obtain the $m^{QRS'}$ finding from the environment, shown in Figure 3.1 as I_6. To move on, we can apply the SUBSUME() function, establishing the matching relation $\{m^{QRS'} \leftarrowtail o^{QRS}\}$. This leads to the I_7 interpretation, in which the uncertainty on the o^{Tw} observation is reduced; however, the evidence for the P_{Tw} pattern is not yet complete. A new DEDUCE() step is necessary, which deduces the m^{wave} necessary finding in the I_8 interpretation. This finding is also absent, so another PREDICT() step is required. In this last step, the P_{wave} pattern can be applied to observe the deviation in the raw ECG signal, generating the o_3^{wave} observation and completing the necessary evidence for the o^{Tw} observation and thus also for o^N. Constraint solving assigns the value of t^b_{Tw}, t^e_{Tw} and t^e_N, so the result is a cover of the initial interpretation problem in which all the hypotheses have a necessary and sufficient set of evidence. This solution is depicted in I_9.

It is worth noting that in this example the global matching relation $\leftarrowtail_I$ is not injective, since $m^{QRS} \leftarrowtail o^{QRS}$ and $m^{QRS'} \leftarrowtail o^{QRS}$. Also note that each interpretation only generates one descendant; in a more complex scenario, however, the possibilities are numerous, and the responsibility of finding the proper sequence of reasoning steps lies with the CONSTRUE() algorithm.

3.1.6 Improving the efficiency of interpretation through saliency

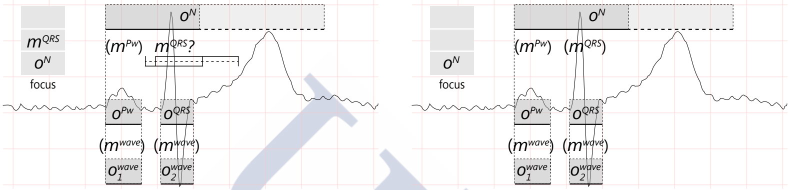
Starting a hypothesize-and-test cycle for every single sample is not feasible for most of the time series interpretation problems. Still, many problems may benefit from certain saliency features that can guide the attention focus to some limited temporal fragments that can be easily interpretable. Thus, the interpretation of the whole time series can pivot on a reduced number of initial observations, thereby speeding up the interpretation process.

A saliency-based attentional strategy can be devised from the definition of abstraction patterns using a subset of their constraints as a coarse filter to identify a set of plausible observations. For example, in the ECG interpretation problem the most common strategy is to begin the analysis by considering a reduced set of time points showing a significant slope in the signal, consistent with the presence of QRS complexes [Zong et al., 2003b]. This small set of evidence allows us to focus the interpretation on the promising signal segments, in the same way that a cardiologist focuses on the prominent peaks to start the analysis [Wagner, 2008]. It should be noted that this strategy is primarily concerned with the behavior of the focus of attention, and that it does not discard the remaining, non-salient observations, as these are included later in the interpretation by means of the subsumption and prediction reasoning modes.



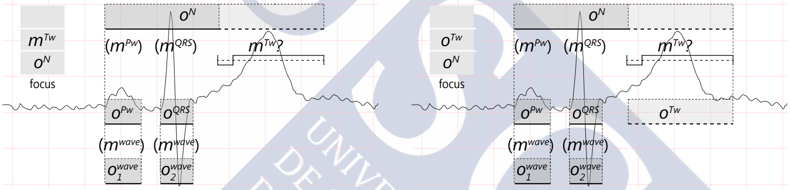
I_0 = Initial evidence

$I_1 = \text{ABDUCE}(I_0, o^{Pw})$



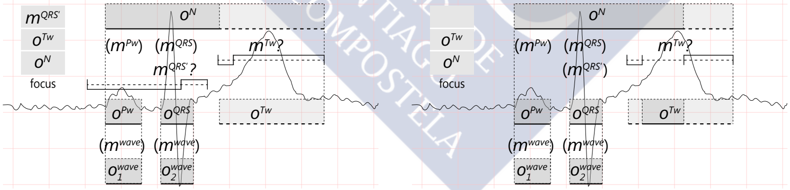
$I_2 = \text{DEDUCE}(I_1, o^N)$

$I_3 = \text{SUBSUME}(I_2, m^{QRS})$



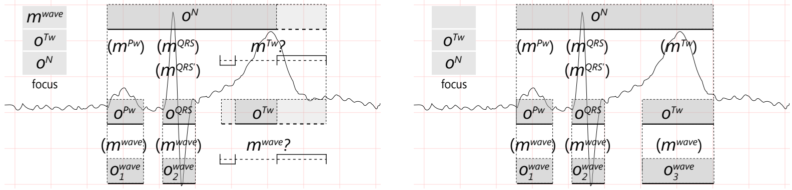
$I_4 = \text{DEDUCE}(I_3, o^N)$

$I_5 = \text{PREDICT}(I_4, m^{Tw})$



$I_6 = \text{DEDUCE}(I_5, o^{Tw})$

$I_7 = \text{SUBSUME}(I_6, m^{QRS'})$



$I_8 = \text{DEDUCE}(I_7, o^{Tw})$

$I_9 = \text{PREDICT}(I_8, m^{wave})$

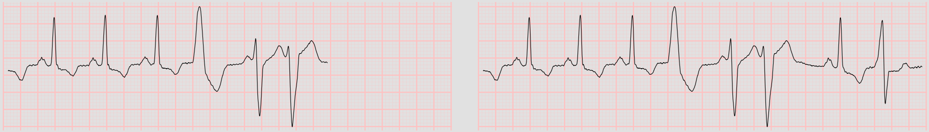
Figure 3.1: Sequence of reasoning steps for solving a simple interpretation problem.

3.2 On-line interpretation

Up to now, an interpretation problem is determined by a set $\mathcal{O}$ of initial observations that progressively grows by including new abstraction hypotheses as a result of the interpretation process. In this context, all the evidence that has to be explained must be available at the very beginning of the interpretation. This condition is not satisfied in the problem of continuous signal interpretation, since new evidence appears steadily, and it must be explained either by the existing hypotheses or by guessing new ones.

The CONSTRUE() algorithm described in Algorithm 1 is complete, since it traverses the full search space until it finds a cover for the interpretation problem. If no covers are found, it returns the interpretation with the highest coverage factor, so the algorithm is also optimal with respect to the coverage principle. However, this completeness poses an important practical problem if the abstraction model does not admit any cover for the initial evidence, due to the exponential growth of the search space. This is illustrated in Example 3.2.1.

Example 3.2.1: Let us consider the abstraction model described in Appendix A and the 3.5 seconds ECG fragment shown in Figure 3.2a. For this problem the size of the full search space is 4593 interpretations, of which 430 are consistent. But if the length of the fragment is extended to 5 seconds, as shown in Figure 3.2b, then the size of the new search space is 15436 interpretations, of which 2019 are consistent. This gives us an idea of the scale of the problem, forcing the adoption of strategies to reduce the time and memory consumption of the interpretation process, even at the cost of losing completeness.



(a) From 00:25.600 to 00:29.100 **(b)** From 00:25.600 to 00:30.600

Figure 3.2: Signal fragments showing the complexity growth of the search space [Source: MIT-BIH arrhythmia DB, recording: 233]

In this section, an evolution of the CONSTRUE() algorithm is proposed for obtaining effective interpretations of time series with an affordable computational cost, and enabling a continuous acquisition process to include new evidence over time. This new algorithm is named CONSTRUE-ONLINE(), and its pseudocode is shown in Algorithm 7.

Algorithm 7 On-line version of the CONSTRUE search algorithm.

```
function CONSTRUE-ONLINE($IP$)
    var $I_0 = \varnothing$
    var $K = max(|\{q_j \in \mathcal{Q} \mid q_i \Bumpeq q_j, q_i \in \mathcal{Q}\}|)$
    var $tlast = 0, tev = 0$
    $\text{SET\_FOCUS}(I_0, o_1)$
    var $open = \text{SORTED}([\langle \varepsilon(I_0, tlast), I_0 \rangle])$
    var $closed = \text{SORTED}([])$
    while $open \neq \varnothing$ do
        $\mathcal{O} = \mathcal{O} \cup \{o_k \mid tev < t_k^e \leq \text{TIME}()\}$
        $tev = max(t_k^e \mid o_k \in \mathcal{O})$
        for all $I \in open[0 \dots K]$ do
            $I' = \text{NEXT}(\text{GET\_FIRM\_DESCENDANTS}(I))$
            $tlast = max(T(I'), tlast)$
            if $I'$ is null then
                $open = open - \{\langle \varepsilon(I, T(I)), I \rangle\}$
                $closed = \{\langle \varepsilon(I'', tlast), I'' \rangle \mid I'' \in closed\} \cup \{\langle \varepsilon(I, tlast), I \rangle\}$
            else if $\varsigma(I', \infty) = 1.0$ then
                return $I'$
            else
                $open = open \cup \{\langle \varepsilon(I', T(I')), I' \rangle\}$
            end if
        end for
        if $tev - last > \Delta$ then
            $open = min_K(\{\langle \varepsilon(I'', tlast), I'' \rangle \mid I'' \in open\})$
        end if
    end while
    return $min(closed)$
end function
```

A number of strategies are described in the following subsections to improve the efficiency of the interpretation process and to support the dynamic inclusion of new evidence. These strategies concern both the definition of adequate search heuristics for on-line interpretation and a pruning mechanism of the search space, as well as the exploitation of the knowledge structure to guide the hypothesize-and-test cycle. Finally, a specific formalism for temporal knowledge representation based on the *Simple Temporal Problem* (STP) [Dechter et al., 1991] is proposed in order to improve the efficiency of temporal reasoning in the interpretation process.

3.2.1 Dynamic inclusion of new evidence

A dynamic interpretation process requires the inclusion of new evidence after the interpretation has started. In the CONSTRUE-ONLINE() algorithm, the inclusion is performed before each cycle of expansion of best K nodes, by adding new observations in temporal order (line 9). This inclusion of new evidence demands a redefinition of the heuristics used for interpretation comparison, as $domain(\Bbbk)$ is constantly updated and also the coverage ratio. This leads to the definition of the *covering ratio until time t* of an interpretation I as follows:

$$\varsigma(I,t) = \frac{|\{o_k \in \texttt{abstracted\_by}(O_I) | t_k^e \leq t\}|}{|\{o_k \in O(domain(\Bbbk)) | t_k^e \leq t\}|}$$

The heuristic vector is also redefined as $\varepsilon(I,t) = (1.0 - \varsigma(I,t), \kappa(I))$. In addition, we define the *temporal frontier* of an interpretation I, denoted by $T(I)$, as the latest ending time of the evidence explained by the interpretation:

$$T(I) = max\{t_k^e \mid o_k \in \texttt{abstracted\_by}(O_I)\}$$

By default, the heuristic vector of a new interpretation I' is calculated as $\varepsilon(I', T(I'))$ (line 20), so looking for higher coverages takes precedence over advancing the interpretation over time. This is consistent with the heuristic principles guiding the search, but in practice makes it difficult to provide valuable results for hard problems in affordable time, particularly if no full coverages exist. The following subsections detail the strategies adopted to constrain the execution time of the algorithm while keeping the attachment to the search principles.

3.2.2 Pruning the search space

As we illustrated in the example in Figure 3.2, a complete traversal of the search space is not feasible in most ECG interpretation problems. Therefore, some strategy has to be adopted in order to reduce the interpretation time. We have opted for an approach following the principles of the K-Beam algorithm [Edelkamp and Schrödl, 2011], pruning the *open* list under certain conditions. These conditions may refer to the consumption of time, memory, or any combination of resources we need to limit.

During the interpretation process, two different time points are continuously tracked (line 4): the *last explanation point*, *tlast*, which is the most advanced temporal frontier of the interpretations generated so far (updated in line 13); and the *evidence acquisition point*, *tev*, that is the ending time of the latest observation included as base evidence (updated in line 10). Example 3.2.2 shows these two time points in a particular interpretation.

Example 3.2.2: Consider the ECG strip shown in Figure 3.3. In an *on-line* interpretation, the signal is acquired concurrently with the interpretation process, and the *evidence acquisition point* determines the last time for which the signal is available (in this example, $tev = 10{:}48.200$). Let us now consider the outlined partial interpretation, which we will call I, built by the CONSTRUE-ONLINE() algorithm from the abstraction model described in Appendix A. This interpretation describes the ECG fragment in three abstraction levels: The *Deflection* level contains observations of signal deviations consistent with the presence of electrical activity from the heart (gray areas). These observations are abstracted into *P wave* (green areas), *QRS complex* (blue areas) and *T wave* (yellow areas) observations at the *Conduction* level. Finally, conduction phenomena are abstracted into rhythm hypotheses at the *Rhythm* abstraction level. In this particular example, the first five beats are abstracted by a normal *Sinus rhythm* hypothesis (green area), while the following four beats are abstracted by a *Tachycardia* hypothesis (red area). The *Tachycardia* hypothesis is still open and its specific end time is not yet known (this is represented by a sloping line between the earliest and the latest possible end times), so more heartbeats may be abstracted by it. The *temporal frontier* of this partial interpretation is the end time of the last *T wave* observation, so $T(I) = 10{:}46.080$, and assuming this is the most advanced interpretation built by CONSTRUE-ONLINE() so far, the *last explanation point* is also this instant.

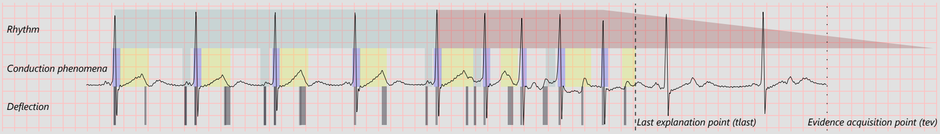


Figure 3.3: Time references used by the pruning mechanism. [Source: MIT-BIH arrhythmia DB, recording: 220, between 10:40.000 and 10:48.200]

The evolution of these two variables will guide the search strategy regarding performance management. It can be assumed that variable tev evolves at a constant rate in real time. On the other hand, the rate of change of the $tlast$ variable cannot be predicted in advance, and it depends on the complexity of the problem. Also, there might be temporal fragments that are more difficult to explain than others, which will affect this rate of change. A simple but effective method to narrow the interpretation time involves tracking the evolution of the difference between the tev and $tlast$ variables. When this difference exceeds a maximum value Δ, the *open* list is reduced and only the K best interpretations with respect to the most advanced interpretation time $tlast$ are considered in the next iteration (line 24).

Finally, interpretations in the *closed* list are always ranked by taking the heuristic value for the latest interpretation time $tlast$, and the full list has to be updated every time an interpretation is closed (line 16).

Example 3.2.3: To illustrate the convenience of this pruning mechanism in ECG interpretation problems, let us consider the ECG fragment shown in Figure 3.4. This fragment is difficult to interpret due to the presence of atrial fibrillation with polymorphic QRS complexes, baseline drift, and a final episode of a possible incipient ventricular tachycardia. The ECG abstraction model cannot provide a full coverage for this fragment, and the best explanation found by the interpretation algorithms is depicted in the figure. In this interpretation, the first heartbeat cannot be explained and therefore is considered as unintelligible. Then, the five following beats are incorrectly interpreted as a *Sinus rhythm* episode with a *Rhythm block* between the third and fourth beat (brown area). After that, a proper *Atrial fibrillation* episode with seven beats is conjectured (blue area), and the two final beats are also correctly interpreted as a *Tachycardia* (red area). The coverage ratio of this interpretation is 0.9811.

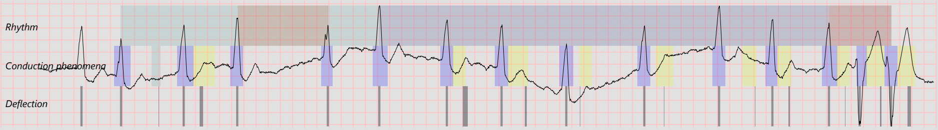


Figure 3.4: ECG fragment to illustrate the convenience of the pruning strategy. [Source: MIT-BIH arrhythmia DB, recording: 203, between 28:30.000 and 28:40.500]

Using the CONSTRUE algorithm without pruning, a total number of 1040799 interpretations are explored to obtain this interpretation, taking 2549.568 seconds in a regular PC. With the pruning mechanism and setting $\Delta = 2500$ ms, the same interpretation is provided after exploring 16582 interpretations, taking 44.765 seconds to finish in the same computer.

3.2.3 Necessary and sufficient abstraction conditions

The knowledge structure of an abstraction model can be exploited by identifying certain shortcuts that can be used to increase the efficiency of the interpretation procedure, by guiding and focusing the search for the best explanation. These shortcuts appear as a set of necessary and sufficient conditions attached to the different hypotheses that can be guessed by the abstraction model: necessary conditions avoid some hypotheses that can be pruned because these are inconsistent with data; sufficient conditions form a solid basis for further explanation. Both types of conditions can be compiled from the abstraction model before solving an interpretation problem, aiming at developing a more efficient interpretation strategy [Console et al., 1996].

Necessary conditions

An abstraction grammar $G \in G^{ap}$ describes all the different ways of guessing an observation for a given observable h as a set of abstraction patterns $P_G = \{P_1, P_2, \ldots, P_i, \ldots\}$. Given a pattern $P = \langle h, M_P, C_P, \pi_P \rangle \in P_G$, C_P gathers all the necessary conditions for the hypothesis h being guessed by P. Thus, $C_{P_1} \vee C_{P_2} \vee \ldots \vee C_{P_i} \vee \ldots$ gathers the whole set of necessary conditions described by the grammar G for the hypothesis h. The relevant property of necessary conditions is that a hypothesis can be rejected when its necessary conditions are inconsistent with observations, and this property can play an important role in the hypothesize-and-test cycle: whenever an abstraction hypothesis on h is conjectured from the available evidence, necessary conditions should be checked first in order to reject h as soon as it appears inconsistent with data. This is a consequence of the falsification property of abductive reasoning [Flach, 1996]. More than one necessary condition can be attached to each hypothesis; for instance, an alternative set of necessary conditions for h can be any of the elements C_{P_i} in the aforementioned disjunction. The most interesting set of necessary conditions is the most specific one, i.e., the one implied by all the others. Our approach seeks to identify a set of *operational* conditions, concerning constraints that help to reduce the search space [Console et al., 1996]. The notion of operational necessary conditions for a hypothesis $h \in \mathcal{Q}$ is defined as follows:

Definition 19. Given an abstraction grammar $G \in G^{ap}$, the **operational necessary conditions** for the hypothesis h is the largest set of constraints C_{NC} such that for every abstraction pattern $P = \langle h, M_P, C_P, \pi_P \rangle \in P_G$ it holds that $C_{NC} \subseteq C_P$.

Thus, C_{NC} is the set of constraints that must be satisfied whenever h is observed, and can guide the interpretation algorithm by dismissing inconsistent hypotheses as early as possible. Operational necessary conditions should be considered in the design of abstraction grammars, as they can dramatically influence the performance of the interpretation procedure.

Any abstraction grammar, as a regular grammar, can be represented by a nondeterministic finite automaton, and this automaton can in turn be represented as a graph. This graph representation is a suitable format to identify necessary conditions in a grammar. For example, consider the *Trigeminy* abstraction grammar generating the set of patterns for trigeminy arrhythmias and which is described in section A.3.6 (page 127). The equivalent finite automaton for this grammar is depicted in Figure 3.5.

If we analyze the transitions in this graph and the associated constraints definition, we can check that the operational necessary conditions for this abstraction grammar involve a first

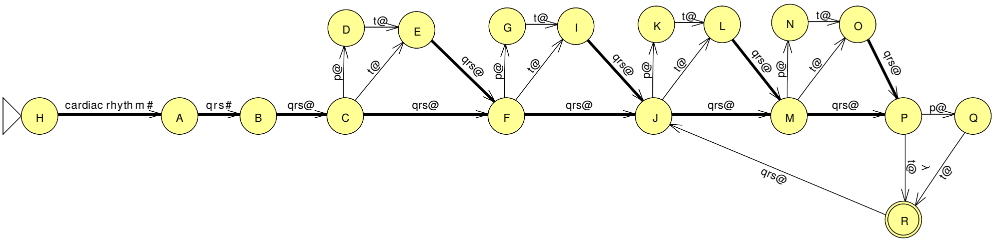


Figure 3.5: Finite automaton equivalent to the *Trigeminy* abstraction grammar, with highlighted necessary conditions.

environment *Cardiac Rhythm* observation and six *QRS complexes*, as well as the constraints defined from state H to P affecting these observations. Thus, if the grammar rules defining these constraints were described in sequence from state H, the operational necessary conditions could be checked in at most seven reasoning steps. However, in the form they are described in the example grammar, the number of reasoning steps to check these conditions could be up to 81, corresponding to the number of simple paths between H and P.

A subset of the operational necessary conditions can be automatically obtained from the edge analysis of the equivalent automaton of an abstraction grammar. In this manner, constraints associated to the edges of the graph whose removal prevents any path from the initial state H to a final state are necessary conditions.

For example, consider the G_{VB} grammar in Example 2.1.8 (page 39). The graph representing the equivalent finite automaton is shown in Figure 3.6.

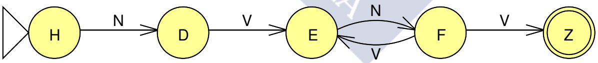


Figure 3.6: Finite automaton equivalent to the G_{VB} abstraction grammar.

From the edge analysis of this graph, we can conclude that the constraints defined by the rules $H \to q_N D$, $D \to q_V E$, $E \to q_N F$ and $F \to q_V$ are necessary conditions. However, it should be noted that the incremental building of abstraction patterns prevents to check a set of constraints that is not generated by a sequence of consecutive rules of the grammar. Also, constraints defined in rules that may be applied after a cycle in the graph cannot be automatically included in the set of operational necessary conditions since they can involve an undetermined number of findings. Therefore, the set of necessary conditions obtained from the edge analysis of the finite automaton has to be restricted to the largest set generated by sequentially connected rules before any cycle. In the case of G_{VB}, this set corresponds to the constraints defined by the rules $H \rightarrow q_N D$ and $D \rightarrow q_V E$. In general, the set of necessary conditions that can be automatically obtained is quite restricted, and it usually corresponds to the first set of constraints that is already checked by the CONSTRUE() algorithm in its execution from the first symbol of the grammar. Still, this analysis is useful to guide the design of abstraction grammars, as illustrated above.

Sufficient conditions

Sufficient conditions can be used to avoid exploring parts of the search space that would lead to redundant interpretations. Any abstraction pattern $P = \langle h, M_P, C_P, \pi_P \rangle \in P_G$ contains the necessary conditions C_P that also behave as sufficient conditions for guessing h. Thus given a set $P_G = \{P_1, P_2, \ldots, P_i, \ldots\}$ of abstraction patterns, $C_{P_1} \vee C_{P_2} \vee \ldots \vee C_{P_i} \vee \ldots$ gathers the whole set of sufficient conditions described by the grammar G for the hypothesis h. If sufficient conditions for a hypothesis h are found to be true, h cannot be refuted later, as long as its supporting evidence remains intact. Thus, if a sufficient condition for a hypothesis h is true, search may be avoided on further interpretations of the same piece of evidence. An efficient checking of sufficient conditions leads us to identify a set of *operational* sufficient conditions:

Definition 20. Given an abstraction grammar $G \in G^{ap}$, the **operational sufficient conditions** for the hypothesis h is the minimum set $\{C_{SC}^1 = C_{P_i}, \ldots, C_{SC}^n = C_{P_j}\}$ such that for every abstraction pattern $P = \langle h, M_P, C_P, \pi_P \rangle \in P_G$ there is a set of sufficient conditions C_{SC}^i such that $C_{SC}^i \subseteq C_P$.

Operational sufficient conditions gather the minimum set of conditions that have to be checked in order to ensure that a hypothesis cannot be refuted, which can be understood as a form of confirmatory reasoning. For example, for the *Trigeminy* grammar shown in Figure 3.5, the operational sufficient conditions contains a single set of conditions, corresponding to the constraints of the abstraction pattern generated from the shortest path between states H and R.

Lemma 1. *For every grammar $G \in G^{ap}$, the set of operational sufficient conditions is finite.*

Proof: We can distinguish two types of abstraction grammars:

i. Grammars generating a finite set of patterns $\{P_1, \ldots, P_n\}$: In this case, the proof is trivial, since in the worst case the operational sufficient conditions gather the constraints of all patterns: $\{C_{P_1}, \ldots, C_{P_n}\}$, and this set is finite of size n.

ii. Grammars generating an infinite set of patterns: Since a grammar has by definition a finite set of rules R, these grammars are necessarily recursive. Let us call $P_0 = \langle h, M_{P_0}, C_{P_0}, \pi_{P_0} \rangle$ the abstraction pattern resulting from applying the rules in a recursion exactly once. Then, for every pattern $P_+ = \langle h, M_{P_+}, C_{P_+}, \pi_{P_+} \rangle$ generated by applying the rules in the recursion more than once, it follows that $C_{P_0} \subset C_{P_+}$, satisfying the definition of the operational sufficient conditions. The same reasoning applies if more than one recursion exists. Therefore, and again in the worst case, the set of constraints obtained by applying rules in a recursion exactly once constitute a finite set of operational sufficient conditions for h. $\square$

From Lemma 1 it follows that even in grammars generating an infinite number of abstraction patterns, it is possible to check that at least one pattern can be observed in finite time. A case of special interest occurs when the set of operational sufficient conditions has only one element. In this situation, it follows that the operational sufficient conditions match the operational necessary conditions $\{C_{SC} = C_{NC}\}$. Thus, grammars with this property allow a very efficient verification or refutation of the hypothesis h, requiring to check a single set of constraints C_{NC}.

On the other hand, the generation of new interpretations during the hypothesize-and-test cycle as implemented in the GET_DESCENDANTS() function (Algorithm 2) is too memory consuming, particularly due to ABDUCE() and PREDICT() reasoning modes (Algorithms 3 and 6). These modes generate a number of new open interpretations with all the possible explanations for a given observation or a predicted finding. Most of these interpretations will found to be inconsistent, but consistency checking requires an exploration leading to new predictions, resulting in a wide search tree that will be mostly discarded.

To address this weakness, a new strategy for the traversal of the search space has been developed, based on operational sufficient conditions. This strategy includes a greedy search step after a new hypothesis has been conjectured, trying to discard it or confirm it as soon as possible. As a consequence, the CONSTRUE-ONLINE() algorithm traverses only a subspace of the full interpretation tree, what is called the space of *firm* interpretations.

Definition 21. Given an interpretation problem IP, we say an interpretation $I = \{\hbar_1, \ldots, \hbar_m\}$ is **firm** if and only if for every hypothesis $\hbar = \langle o_h, P_\hbar, \leftarrowtail_\hbar \rangle \in I$ it holds that $C^i_{SC} \subseteq C_{P_\hbar}$.

This definition states that every hypothesis has to be supported by a set of evidence satisfying a set of constraints in the operational sufficient conditions of h. This means that the hypotheses in a firm interpretation will not be rejected regardless future observations.

The traversal of this subspace is performed by the GET_FIRM_DESCENDANTS() function (line 12), detailed in Algorithm 8. The algorithm is a simple greedy depth-first search that only returns the interpretations that are firm. In this way, every time a new hypothesis is conjectured the interpretation focuses on the satisfaction of a set of sufficient conditions before moving on to another hypothesis. It should be noted that the heuristic comparison between different interpretations is still performed in the CONSTRUE-ONLINE() algorithm, but considering only firm interpretations.

Algorithm 8 Method to iterate through the subspace of firm interpretations, using greedy depth-first search between consecutive nodes.

```
function GET_FIRM_DESCENDANTS($I$)
    var $stack = [I]$
    while $stack \neq \varnothing$ do
        $I = stack[-1]$
        $I' =$ NEXT(GET_DESCENDANTS($I$))
        if IS_FIRM($I'$) then
            return $I'$
        else if $I'$ is null then
            POP($stack$)
        else
            PUSH($stack, I'$)
        end if
    end while
end function
```

3.2.4 An efficient temporal representation

Temporal constraints in abstraction patterns describe the particular temporal arrangement that has to be satisfied by the evidence in order to be abstracted by the hypothesis. Up to now, this has been assumed to be a subset of the whole constraints C_P, that can be formalized as a boolean function $\mathcal{TC}_P(T_h^b, T_h^e, T_1^b, T_1^e, \ldots, T_n^b, T_n^e) \in \{True, False\}$. This model makes it computationally unfeasible to look for specific solutions for the variables $\{T_h^b, T_h^e, T_1^b, \ldots, T_n^e\}$ in the general case, and constraints can only be checked after the values of the variables are known. Hence, a more bounded representation formalism for temporal knowledge description is proposed, which will allow us to delimit a temporal fragment of the time series to find the solution of temporal constraints in an abstraction pattern. This formalism describes the temporal arrangement of the observables in an abstraction pattern by means of a temporal network based on the *Simple Temporal Problem* (STP) [Dechter et al., 1991]. The STP allows us to represent temporal information as a network of metric temporal constraints between a set of temporal variables, each one representing a time instant. The STP formalism has been successfully applied on temporal reasoning, scheduling or planning tasks, and in deductive, abductive [Brusoni et al., 1997, 1998] and inductive [Alvarez et al., 2013] inference frameworks. Furthermore, it represents a good balance between expressiveness and complexity, supporting the usual processing tasks in polynomial time. A STP defines a temporal constraint $\tau c(T_i,T_j)$ between every two temporal variables T_i and T_j as a closed interval $\tau c_{ij} = [a_{ij}, b_{ij}]$, where a_{ij} and b_{ij} are integer numbers constraining the possible values of the duration of the interval between both temporal variables, so that $a_{ij} \leq T_j - T_i \leq b_{ij}$. Formally:

Definition 22. A **temporal network** $N = \langle \mathcal{T}, \mathcal{TC} \rangle$ consists of a set of temporal variables $\mathcal{T} = \{T_1, \dots, T_n\}$, and a set of temporal constraints $\mathcal{TC} = \{\tau c(T_i,T_j);\ 1 \leq i,j \leq n\}$ between the variables in $\mathcal{T}$.

A temporal network can be represented as a directed graph, where each node is associated to one of the variables in $\mathcal{T}$, and each arc from the node T_i to the node T_j is associated to the constraint $\tau c(T_i,T_j)$. Given a constraint $\tau c(T_i,T_j) = [a_{ij}, b_{ij}]$, its symmetrical constraint $\tau c(T_j,T_i) = -\tau c(T_i,T_j) = [-b_{ij}, -a_{ij}]$ contains the same information, so both are redundant and we will just represent one of them. The absence of any explicit constraint between a pair of variables T_i and T_j is equivalent to the constraint $\tau c(T_i,T_j) = \tau c_U = (-\infty, \infty)$, defined as the *universal constraint*, that is, the constraint that do not constrain the possible values of duration of the interval between both temporal variables. We will not represent the universal constraints in the graph.

Given a temporal network $N = \langle \mathcal{T}, \mathcal{TC} \rangle$, a set $\mathcal{TC}'$ of new temporal constraints on $\mathcal{T}$ can be added later by the usual *intersection* set operation, resulting in a new temporal network $N' = \langle \mathcal{T}, \mathcal{TC} \cap \mathcal{TC}' \rangle$ where $\mathcal{TC} \cap \mathcal{TC}' = \{\tau c(T_i,T_j) \cap \tau c'(T_i,T_j);\ 1 \leq i,j \leq n\ \wedge\ \tau c(T_i,T_j) \in \mathcal{TC}\ \wedge\ \tau c'(T_i,T_j) \in \mathcal{TC}'\}$.

Definition 23. Given a temporal network $N = \langle \mathcal{T}, \mathcal{TC} \rangle$, its **extension over** the set $\mathcal{T}'$, being $\mathcal{T} \subseteq \mathcal{T}'$, is a new temporal network $N^{\uparrow\mathcal{T}'} = \langle \mathcal{T}', \mathcal{TC}^{\uparrow\mathcal{T}'} \rangle$, where $\mathcal{TC}^{\uparrow\mathcal{T}'} = \{\tau c^{\uparrow\mathcal{T}'}(T_i,T_j);\ 1 \leq i,j \leq n\}$ such that $\tau c^{\uparrow\mathcal{T}'}(T_i,T_j) = \tau c(T_i,T_j)$ if $T_i, T_j \in \mathcal{T}$, and $\tau c^{\uparrow\mathcal{T}'}(T_i,T_j) = \tau c_U$ otherwise.

Definition 24. Given two temporal networks $P = \langle \mathcal{T}_p, \mathcal{TC}_p \rangle$ and $Q = \langle \mathcal{T}_q, \mathcal{TC}_q \rangle$, their **combination**, $P \bowtie Q$, is a new temporal network $R = \langle \mathcal{T}_r, \mathcal{TC}_r \rangle$, where $\mathcal{T}_r = \mathcal{T}_p \cup \mathcal{T}_q$ and $\mathcal{TC}_r = \{\tau c_r(T_i,T_j) = \tau c_p^{\uparrow\mathcal{T}_r}(T_i,T_j) \cap \tau c_q^{\uparrow\mathcal{T}_r}(T_i,T_j);\ \tau c_p^{\uparrow\mathcal{T}_r}(T_i,T_j) \in \mathcal{TC}_p^{\uparrow\mathcal{T}_r}\ \wedge\ \tau c_q^{\uparrow\mathcal{T}_r}(T_i,T_j) \in \mathcal{TC}_q^{\uparrow\mathcal{T}_r}\}$.

We will denote by $\Lambda = \langle \varnothing, \varnothing \rangle$ the empty temporal network, and it satisfies that, $\Lambda \bowtie N = N \bowtie \Lambda = N$, for every temporal network N. The combination operation is used to incrementally build temporal networks from abstraction grammars, including in each step the temporal variables of a new observable and the constraints from this observable to those generated previously.

Definition 25. An n-tuple $(t_1, \ldots, t_n) \in \tau^n$ is a **solution** of the temporal network N if and only if the assignment $T_1 = t_1, \ldots, T_n = t_n$ satisfies all the constraints in the network.

A temporal network N is considered *consistent* if and only if there exists at least one solution. It can be demonstrated that any k-tuple of assignments $(t_{i_1}, \ldots, t_{i_k})$, with $1 \leq k \leq n$, satisfying the corresponding constraints $\{\tau c(T_{i_h}, T_{i_g});\ i_1 \leq i_h, i_g \leq i_k\}$ of a consistent network N can be extended to a solution of N. We say that such k-tuple is consistent with N.

Given a temporal network N as a STP, it always exists the minimal network M equivalent to N, which corresponds with the most explicit representation of N. In order to analyze the consistency of N and obtain its minimal representation, the Floyd-Warshall's ALL-PAIRS-SHORTEST-PATHS algorithm [Dechter et al., 1991] can be used, which calculates the intersection of each constraint $\tau c(T_i, T_j)$ with all the induced constraints of the form $\tau c(T_i, T_k) \oplus \tau c(T_k, T_j) \equiv [a_{ik} + a_{kj}, b_{ik} + b_{kj}]$, being $\tau c(T_i, T_k) = [a_{ik}, b_{ik}]$ and $\tau c(T_k, T_j) = [a_{kj}, b_{kj}]$, for every pair of nodes T_i and T_j of the graph.

Definition 26. A temporal network $N = \langle \mathcal{T}, \mathcal{TC} \rangle$ is **minimal** if and only if every temporal constraint $\tau c(T_i, T_j) \in \mathcal{TC}$ holds that:

$$\tau c(T_i, T_j) \subseteq \cap_{\forall k} (\tau c(T_i, T_k) \oplus \tau c(T_k, T_j))$$

The adoption of this STP-based temporal formalism enables separation of constraints in an abstraction pattern P into two disjoint constraint subsets $C_P = \mathcal{TC}_P \cup \mathcal{VC}_P$, where $\mathcal{TC}_P$ is a set of temporal constraints between the temporal variables in h and M_P, and $\mathcal{VC}_P$ represents a set of non temporal constraints $\mathcal{VC}_P = \{\upsilon c_1(\mathbf{A}_h, \mathbf{A}_1, \ldots, \mathbf{A}_n), \ldots, \upsilon c_n(\mathbf{A}_h, \mathbf{A}_1, \ldots, \mathbf{A}_n)\}$. This separation boosts the performance of deductive reasoning operations, since the range of values for the temporal variables of the findings are bounded by constraint propagation in the minimization operation of $\mathcal{TC}_P$. Additionally, this model is naturally included in the incremental building of abstraction patterns by means of the aforementioned combination operation, as follows: $C_i = \mathcal{TC}_i \cup \mathcal{VC}_i; C_j = \mathcal{TC}_j \cup \mathcal{VC}_j \Rightarrow C_i \cup C_j = (\mathcal{TC}_i \bowtie \mathcal{TC}_j) \cup (\mathcal{VC}_i \cup \mathcal{VC}_j)$.

Example 3.2.4: Consider the partial interpretation of the ECG fragment shown in Figure 3.7, in which the first two beats are abstracted by a normal *Sinus rhythm* hypothesis, while the following five beats are abstracted by a *Bigeminy* hypothesis (yellow area). This hypothesis is still open and more heartbeats may be abstracted by it, although at this point these have not yet been observed.

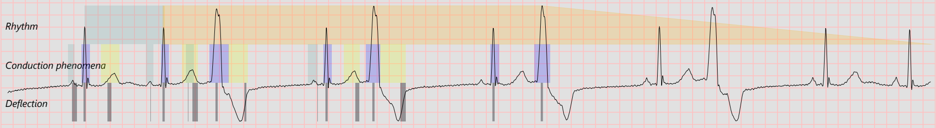


Figure 3.7: Partial interpretation with a *Bigeminy* rhythm hypothesis. [Source: MIT-BIH arrhythmia DB, recording: 119, between 04:51.000 and 05:01.500]

Let us assume that during the hypothesize-and-test cycle the focus of attention is on the *Bigeminy* hypothesis, and therefore a new deduction operation may be performed. The DEDUCE() function is then applied, generating a new finding for the *QRS complex* observable. There are no observations to be matched with this finding in a subsumption operation, so without any additional information we would be forced to predict all possible *QRS complex* observations and check which of them would satisfy the constraints defined by the *Bigeminy* pattern. However, if we use the STP formalism for temporal constraints description, we would be able to focus the prediction to a reduced temporal interval, as shown in Figure 3.8.

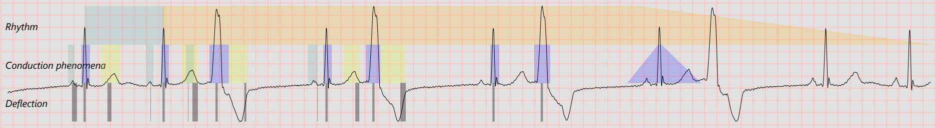


Figure 3.8: Consistent temporal location of the newly deduced *QRS complex* finding.

Following the hypothesize-and-test cycle, the predicted *QRS complex* is conjectured, and eventually a consistent observation is found in the *Deflection* level, so the *QRS complex* is observed and the specific time limits can be calculated. The result is shown in Figure 3.9. After that, the *Bigeminy* hypothesis is focused again, and a new deduction operation can be performed.

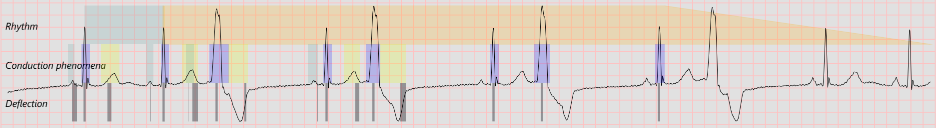


Figure 3.9: Resulting interpretation when the deduced finding has been observed.

The improvements described in this section reinforce the applicability of the CONSTRUE() algorithm in real interpretation problems. In the next chapter, this applicability is demonstrated through various experiments dealing with well-known problems in automatic ECG processing.

CHAPTER 4

RESULTS

In order to assess the validity of our proposal, we have designed and built an abstraction model for the interpretation of ECG signals, detailed in Appendix A. This model defines three abstraction levels: The first one corresponds to signal deviation phenomena, and includes a set of intervals or time points labeled as *deflections*, consistent with the presence of electrical activity from the cardiac muscle. Observations from this level constitute the initial evidence for the abductive interpretation process. The second abstraction level corresponds to intracardiac conduction phenomena, and defines the *P wave*, *QRS complex*, and *T wave* observables. These observables can be conjectured taking as evidence those observations from the first abstraction level, and they provide a description of the ECG as a sequence of waves corresponding to the atrial activation, ventricular activation and ventricular recovery processes, respectively. Finally, the third abstraction level collects the patterns characterizing the *cardiac rhythm* under different physiological conditions [Wagner, 2008], including normal rhythm and arrhythmias.

This means that for any input ECG signal, the interpretation algorithm is able to provide a sequence of cardiac rhythms covering the entire temporal interval of the signal. These rhythms are supported by a sequence of cardiac cycles with their corresponding muscle activation/recovery phenomena, which in turn are supported by signal deviation observations. This result corresponds to the interpretation that is usually performed by medical experts.

At this moment, the interpretation algorithm generates as output an annotation file that locates the onset, the offset, and the fiducial mark of every QRS complex, as well as the onset and offset of all identified P and T waves. Additionally, the file contains a series of rhythm annotations identifying the cardiac rhythm in which each heartbeat is framed.

To evaluate and compare the results with existing approaches, some experiments have been conducted to demonstrate the practical results that can be achieved with an actual implementation of the framework. Each section of this chapter outlines an experiment in which an abductive approach may confer an advantage, the results and its implications are discussed, and a comparison with existing approaches is carried out as far as possible.

4.1 Improving a state-of-the-art QRS detector

This experiment was the first practical test intended for the validation of the interpretation framework capability to correct previous conclusions by exploiting non-monotonic reasoning. For this, the simple problem of QRS detection was considered, and the theoretical capabilities to cope with ignorance and to look for missing evidence were exploited to improve the results of the `WQRS` state-of-the-art QRS detection algorithm [Zong et al., 2003b].

The knowledge base used for this experiment was a small subset of the full ECG abstraction model described in Appendix A. This subset includes the *R-Deflection* observable for the original annotations given by the `WQRS` algorithm, the *QRS complex* observable for the delineation of the QRS complexes, and the *Sinus Rhythm, Bradycardia, Tachycardia* and *Extrasystole* observables to interpret the abstracted QRS complexes. Regarding the abstraction grammars, the *QRS complex* grammar was included to abstract the *R-Deflections* comprising the initial evidence, the *Regular rhythm* grammars were included to conjecture *Sinus rhythms*, *Bradycardias* and *Tachycardias*, and finally the *Extrasystole* grammar was also included to conjecture the presence of ectopic beats. Thus, this reduced abstraction model allows to interpret a ECG signal at the higher abstraction level as a sequence of regular cardiac rhythms with the possible presence of isolated ectopic beats.

Following the CONSTRUE() algorithm, the initial set of *R-Deflections* can be modified during the interpretation by including new predicted beats or by discarding the annotations that are considered false positives as they are declared unintelligible according to the available knowledge. Figure 4.1 shows an example of an ECG trace in which the initial set of *R-Deflections* has both false positive and false negative detections. The interpretation at rhythm level concludes that best hypothesis explaining the fragment is normal sinus rhythm, and therefore the third annotation is discarded as a false positive, and the actual third and fourth QRS complexes are predicted from the smaller signal deviations that appear consistently with the rhythm temporal constraints.

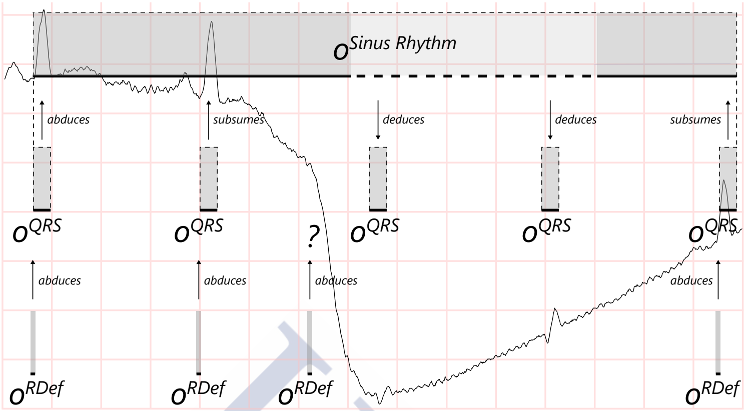


Figure 4.1: How abductive interpretation can fix errors in the initial evidence [Source: MIT-BIH Arrhythmia DB, record: 121, between 27:32.500 and 27:36.000. *R-Deflections* obtained with `WQRS`.]

The validation dataset consisted of a selection of ECG recordings showing regular rhythms and extrasystoles, comprising all the 18 recordings of 24 hours duration of the Normal Sinus Rhythm (NSR) database and 20 recordings of 30 minutes duration from the MIT-BIH Arrhythmia database, both from the Physionet initiative [Goldberger et al., 2000]. For each of these records, the `WQRS` algorithm was executed, and the resulting annotations were converted to *R-Deflection* observations. Then, these observations were interpreted by the CONSTRUE() algorithm, and a new set of annotations was generated with the *QRS* observations abstracted by some rhythm pattern in the resulting best explanation.

Table 4.1 shows the comparative results of the original algorithm and the corrected output through abduction, in terms of sensitivity, positive predictivity, and the combined F1-score. As can be noted, the abstraction process can slightly decrease the overall sensitivity, but is compensated with an improvement in the positive predictivity to get in the majority of cases an advance in the combined F1-score. Still, cases such as records MIT-103 or MIT-109 show that the abstraction process is also able to correct sensitivity failures. In order to prove that the improvement on the detection is statistically significant, the Wilcoxon statistical test was applied to the differences on the F1-score, obtaining a *p-value* of 0.008 on the records of the Normal Sinus Rhythm database, a *p-value* of 0.033 on the records of the MIT-BIH Arrhythmia database, and a combined *p-value* of 0.001 on the full set of records. These results were published in [Teijeiro et al., 2014].

	`WQRS`			`WQRS` + Abduction		
Record	**Se**	**P+**	**F1**	**Se**	**P+**	**F1**
NSR-16265	100.00	99.74	**99.87**	99.97	99.76	99.86
NSR-16272	98.24	89.79	93.83	97.74	93.66	**95.66**
NSR-16273	99.99	99.93	99.96	99.95	99.98	**99.96**
NSR-16420	99.98	99.79	99.88	99.92	99.92	**99.92**
NSR-16483	99.98	99.88	99.93	99.98	99.97	**99.97**
NSR-16539	99.97	99.79	**99.88**	99.76	99.92	99.84
NSR-16773	99.99	99.96	**99.97**	99.96	99.95	99.95
NSR-16786	100.00	99.97	99.98	100.00	99.98	**99.99**
NSR-16795	99.99	99.87	**99.93**	99.96	99.88	99.92
NSR-17052	99.98	99.52	99.75	99.95	99.68	**99.81**
NSR-17453	99.98	99.72	99.85	99.94	99.89	**99.91**
NSR-18177	99.98	99.63	99.80	99.91	99.74	**99.82**
NSR-18184	99.99	99.55	99.77	99.98	99.79	**99.88**
NSR-19088	100.00	98.29	99.14	99.98	98.37	**99.17**
NSR-19090	99.99	99.70	99.84	99.99	99.79	**99.89**
NSR-19093	100.00	99.87	**99.93**	99.99	99.88	**99.93**
NSR-19140	100.00	99.82	99.91	100.00	99.84	**99.92**
NSR-19830	99.93	98.74	99.33	99.86	98.99	**99.42**
NSR-Gross	**99.90**	**99.08**	**99.49**	**99.83**	**99.40**	**99.61**
MIT-100	100.00	99.95	**99.97**	99.95	99.95	99.95
MIT-101	99.93	99.80	99.86	99.80	100.00	**99.90**
MIT-102	100.00	100.00	**100.00**	100.00	100.00	**100.00**
MIT-103	99.94	100.00	99.97	100.00	100.00	**100.00**
MIT-104	100.00	97.58	98.78	99.25	98.93	**99.09**
MIT-105	99.72	91.72	95.55	98.00	97.24	**97.62**
MIT-107	99.89	98.13	99.00	99.66	99.94	**99.80**
MIT-108	99.59	86.20	92.41	95.54	97.79	**96.65**
MIT-109	99.86	100.00	99.93	100.00	100.00	**100.00**
MIT-111	99.94	99.94	**99.94**	99.77	100.00	99.88
MIT-112	100.00	99.91	99.95	100.00	100.00	**100.00**
MIT-113	100.00	100.00	**100.00**	100.00	99.67	99.83
MIT-115	100.00	100.00	**100.00**	100.00	100.00	**100.00**
MIT-117	100.00	100.00	**100.00**	100.00	100.00	**100.00**
MIT-122	100.00	99.95	99.97	100.00	100.00	**100.00**
MIT-123	100.00	100.00	**100.00**	99.84	99.06	99.45
MIT-209	100.00	99.53	99.76	100.00	99.88	**99.94**
MIT-212	100.00	99.91	99.95	100.00	100.00	**100.00**
MIT-230	100.00	99.95	99.97	100.00	100.00	**100.00**
MIT-234	100.00	100.00	**100.00**	100.00	100.00	**100.00**
MIT-Gross	**99.95**	**98.58**	**99.26**	**99.62**	**99.64**	**99.63**

Table 4.1: QRS detection correction results

4.2 Heartbeat classification using abstract features from the abductive interpretation of the ECG

Heartbeat classification from electrocardiogram signals is a valuable tool for the study of the cardiac arrhythmia, and it is one of the challenges that has raised more efforts in the field of biosignal analysis [Luz et al., 2016]. Heartbeat classifiers have traditionally failed to apply to new patients those categories learned from a training set, no matter how large this training set may be. Some sort of automatic adaptation to the specific characteristics of each new patient could be thought of as the solution, but results have not been satisfactory enough. In fact, recent bibliography shows that most of the state-of-the-art approaches rely on some form of expert assistance for processing every new patient: some of them by combining a general classifier, previously trained with large collections of ECG recordings, with a local one, trained specifically with an annotated fragment of the ECG record of the new patient [de Chazal and Reilly, 2006; Hu et al., 1997; Kiranyaz et al., 2016]; others by following a clustering strategy that requires from the expert the final assignment of a class label to each beat morphology [Llamedo and Martinez, 2012; de Chazal, 2014; Oster et al., 2015].

This experiment aims to return to the challenge of performing an autonomous classification, that is, without requiring expert intervention after the classifier has been designed. To this end, the classification strategy is aligned with the following four principles, supported by the abductive interpretation framework:

- *The knowledge will be explicitly represented*, the kind of knowledge that can be found in an ECG handbook, so that it can be understood and validated by experts; the feature set is limited to those used by experts in clinical practice, such as wave durations, intervals, amplitudes, etc. Interpretability of the classification model is therefore a natural consequence.

- *Ad-hoc thresholds and values will be avoided*, since they may be suspicious of being overfitted to a training database. For this reason, all classification features will be qualitative, and the separation between values will be performed either by purely electrocardiographic criteria, or by generic criteria if no quantifiable knowledge is available.

- *Ignorance will be admitted*, in case of a beat cannot be successfully assigned to any predefined class. Ignorance has the ability to clearly show those weaknesses in the

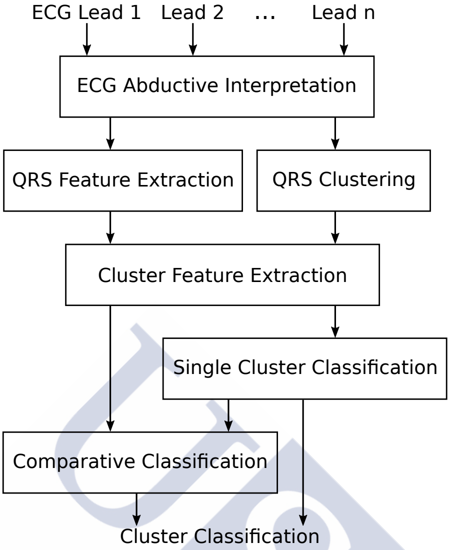


Figure 4.2: Overview of the proposed classification method

knowledge base, providing a direction for further improvement. Moreover, the credibility of the results is enhanced, as long as it comes together with a good performance.

- *Robustness to variability will be demanded*, even to the kind of noise present in the ECG. The interpretation cannot involve any single piece of evidence in isolation, but it should take into account contextual information, combining bottom-up and top-down processing in order to provide a more informed result.

Classification starts from the results of the abductive interpretation process, which works as a temporal abstraction stage [Moskovitch and Shahar, 2015] describing the myocardial behavior at the conduction and rhythm abstraction levels. This provides the same signal features used by cardiologists in ECG analysis [Wagner, 2008]. A context-based adaptive QRS clustering method is then applied to cope with within-patient variability, obtaining a reduced number of groups representing the cardiac activity [Castro et al., 2015]. Finally, a simple knowledge-based classification procedure assigns a label to each cluster. Figure 4.2 shows an overview of the strategy.

4.2.1 Validation Database

The validation of the proposed approach was performed using the MIT-BIH Arrhythmia Database [Moody and Mark, 2001] from the Physionet initiative [Goldberger et al., 2000]. This database can be considered the gold standard for arrhythmia and heartbeat classification, and it has been used in most of the published research. All beats present in the database were annotated by at least two expert cardiologists, and assigned a class label using a 17 label set. This label set was converted to the five standard beat classes defined by the AAMI, following the recommendations of the American National Standard for Ambulatory ECGs (ANSI/AAMI EC38:2007) [AAMI, 2007], detailed in table 4.2. All the comparisons were performed using the `bxb` application from Physiotools [Goldberger et al., 2000]. It is worth mentioning that many works in the bibliography, particularly those following the methodology introduced by de Chazal *et al.* [de Chazal et al., 2004], are not fully compliant with this standard, since MIT-BIH ’j’ and ’e’ classes are grouped in the AAMI ’N’ class instead of ’S’ class. This incorrect labeling affects 245 beats in the database, representing 8% of the total supraventricular beats.

AAMI	MIT-BIH equivalent classes
N	N, L, R, B
S	a, J, A, S, j, e, n
V	V, E
F	F
Q	/, f, Q

Table 4.2: Equivalence between beat classes in the MIT-BIH Arrhythmia Database and the standard AAMI classes

4.2.2 Abductive interpretation of the ECG

The interpretation of the ECG aims to identify and characterize the physiological processes underlying signal behavior by building a representation of the cardiac phenomena in multiple abstraction levels. To this end, the ECG abstraction model described in Appendix A was employed, considering as initial evidence a set of *R-Deflection* observations that can be obtained by any QRS detection algorithm.

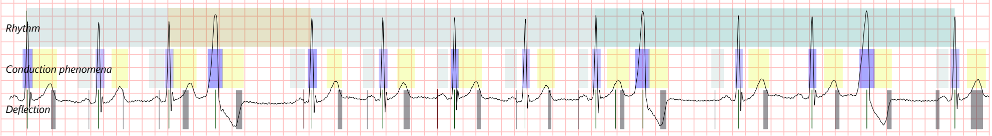


Figure 4.3: Result of the abductive interpretation of an ECG fragment. [Source: MIT-BIH Arrhythmia, record: 116, lead: MLII, between 09:24.000 and 09:34.500]

When the interpretation procedure begins, the first *R-Deflection* observation is focused, and the hypothesize-and-test cycle starts by trying to abduce it in a *QRS complex* observation. Eventually, this new hypothesis may lead to a consistent rhythm hypothesis, which will allow us to deduce new *QRS complexes* as long as the *P wave* and *T wave* observations corresponding to each cardiac cycle. The final target is to obtain a set of rhythm hypotheses explaining all the initial *R-Deflections* as *QRS complexes*, and maximizing the ECG *coverage* with the associated *P wave* and *T wave* observations abstracting the observed *Deflections*. The reason to focus first on *R-Deflections* is that it gives more importance to the interpretation of relevant signal deviations. Furthermore, as shown in Section 4.1, the initial set of *R-Deflections* can be modified during the interpretation process, so QRS detection errors do not necessarily lead to interpretation errors.

Figure 4.3 shows the full interpretation of an ECG fragment following the procedure described above. At the rhythm level, the fragment is explained as a *Sinus Rhythm* episode (light green areas) interrupted by an *Extrasystole* (light brown area) in the fourth beat, a return to *Sinus Rhythm* during five beats and a final *Trigeminy* episode (dark green area) during five beats. At the conduction level, all *QRS complexes* (blue areas) and *T waves* (yellow areas) are properly delineated. With respect to *P waves* (light green areas), the detection is accurate in all cases except the third beat in the trigeminy pattern in which it is missed. As can be seen, the final result of the interpretation procedure is an explanation of the physiological processes observed in the ECG record in the same terms used by experts. This explanation in abstraction levels will make it possible to adapt the sort of knowledge that can be found in any ECG handbook to a reduced set of classification rules, as described below.

Table 4.3 shows the time spent by the CONSTRUE-ONLINE() algorithm for the interpretation of each record in the MIT-BIH Arrhythmia database. All records have a duration of 30 minutes, and we can see that in 34 of them the interpretation time is lower, thus satisfying real-time constraints. However, this condition cannot be generally guaranteed, and for some records such as 203 the interpretation takes up to 3 hours to complete. In aggregate, the entire database takes 26.5 hours to be interpreted, which slightly exceeds the 24 hours of signal duration.

Record	Time (s)	Record	Time (s)	Record	Time(s)
100	548.53	117	438.87	212	772.25
101	528.38	118	907.35	213	1705.05
102	518.21	119	1381.98	214	1609.17
103	545.37	121	490.13	215	2627.02
104	1235.47	122	669.67	217	2514.64
105	1662.71	123	479.86	219	3221.25
106	2576.77	124	1026.79	220	790.81
107	1198.38	200	3575.75	221	6164.96
108	1506.41	201	4008.89	222	3781.80
109	663.32	202	2474.65	223	1998.43
111	650.10	203	11111.71	228	1422.56
112	833.92	205	1751.76	230	581.96
113	606.20	207	1204.81	231	656.24
114	895.52	208	7254.64	232	1803.75
115	506.21	209	1420.19	233	8032.65
116	901.17	201	3271.11	234	769.34

Table 4.3: Interpretation times for the records in the MIT-BIH Arrhythmia DB

4.2.3 QRS clustering

Classification relies on a previous clustering task in order to exploit the high similarity between large number of beats that is invariably observed in ECG signals. A clustering algorithm $\zeta : O(q_{QRS}) \rightarrow \mathbf{P}(O(q_{QRS}))$ should find a partition of the set of QRS observations $O(q_{QRS}) = \{o_1^{QRS}, \ldots, o_n^{QRS}\}$, satisfying the following requirements:

1. Maximum cluster purity. This will be achieved if no beats from different classes belong to the same cluster. All purity reductions below 100% will be directly translated to classification errors.

2. Minimum number of clusters. Ideally, the number of clusters should be equal to the number of beat classes present in the ECG recording. A higher number of clusters increase the uncertainty of the classification features and the number of decisions the classifier has to make.

In this work, we use the clustering method presented in [Castro et al., 2015] which performs an adaptive multi-lead context-based clustering of QRS complexes in real time. This method creates a dynamic number of clusters represented by evolving templates to obtain sets of QRS complexes with similar rhythm and morphology. A high robustness to noise and morphological variability is achieved by means of a segment-based QRS complex characterization inspired by dominant points detection, a similarity measure that performs non-linear alignment using Dynamic Time Warping, and a noise-cluster proliferation control mechanism. As a result, the method provides good cluster purity while minimizing the number of clusters.

4.2.4 Beat classification

As a result of a QRS clustering we obtain a set $\zeta(O(q_{QRS})) = \{Q_1, \ldots, Q_k\}$ of clusters. In the following we explain how to assign a label to each Q_i, classifying all the observations $o_j^{QRS} \in Q_i$ with the information resulting from the abductive interpretation of the ECG.

Heartbeat classification actually involves the determination of two different properties of the beat nature: 1) The physiological origin, i.e. the cardiac muscle area where the electrical activation begins; and 2) the temporal location of the beat with respect to the underlying cardiac rhythm. Usually, the origin distinguishes between supraventricular beats (generated at the atria or the atrioventricular node) and ventricular beats (generated at the ventricles), and the distinction is made based on morphological criteria. On the other hand, the temporal location distinguishes premature beats, escape beats and regular beats, that can be identified from rhythm analysis.

In general, practical electrocardiography makes a distinction between: 1) the interpretation of the "normal" electrocardiogram, and 2) the characterization of the possible transient anomalies [Wagner, 2008]. The goal of the first task is to examine the most common situation observed in the subject's ECG, identifying the possible presence of permanent disorders. The goal of the second task is to study punctual or temporary changes with respect to the "normality" characterized in the first task.

This proposal reproduces this strategy by splitting the label assignment procedure in two stages detailed below. The first one considers a reduced set of individual features of each cluster Q_i and applies a set of general rules to determine the class of the beats in Q_i. But if the value of the individual features is not significant enough to decide the origin, the set of features is extended with additional features that are calculated by comparing the Q_i cluster with an already classified cluster Q_N that is assumed to represent a “normality” situation.

Single cluster classification

The first stage for cluster classification considers a reduced set of five qualitative features obtained from the interpretation results: *Heart Rate, QRS Duration, Heart Axis, P Wave* and *Rhythm*. Each QRS observation in the cluster has a specific value for each feature, and the values for the cluster are obtained by aggregating the values of all the QRS in the cluster for rhythm-related and P Wave features, and by taking the value of a QRS considered the representative of the cluster for morphological features. This representative is selected as the QRS observation with minimum distance to the mean duration, amplitude and heart axis of the cluster. Following we describe these features individually, and Table 4.4 details the possible qualitative values each feature may take.

1. *Heart Rate*: This feature is calculated from the time distance of a QRS observation with respect to the previous QRS observation in the interpretation, what is called the instantaneous RR interval. For a cluster, this feature is calculated as the mean heart rate of all beats in the cluster.

2. *QRS Duration*: Represents the time interval between the onset and the offset of a QRS observation. If the delineation information is available in more than one ECG lead, the duration is taken as the distance between the earliest onset and the latest offset. For a cluster, the duration is calculated as the duration of the representative QRS of the cluster.

3. *Heart Axis*: Represents the mean direction of cardiac depolarization. To be accurately determined, it is necessary the information of all limb leads, specially I, II, and AVF. In this work, since the validation database only has one limb lead available (lead II), we obtain an approximation of the heart axis using only that lead, by calculating the amplitude relation between positive and negative waves in the QRS. The value of this approximation is in the range [-90º, +90º]. Fig. 4.4 shows an example of each possible

qualitative value for this feature, and its corresponding numerical value. For a cluster, the heart axis is taken as the axis of the representative QRS. In the absence of lead II information, a balanced heart axis is assumed.

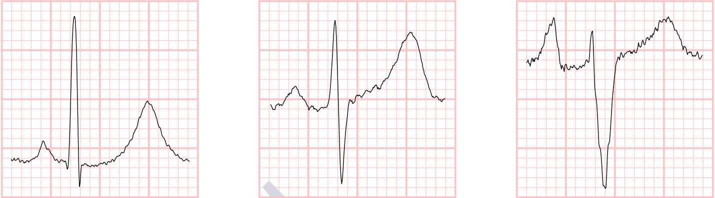

Normal (+70º) Balanced (+5º) Inverted (-56º)

Figure 4.4: Examples of heartbeats with different heart axis

4. *P Wave*: This feature indicates whether a QRS complex is preceded by a P wave in the interpretation result, and therefore has a sinus origin. To determine the presence or absence of a P wave in a cluster, for each available lead an amplitude histogram of all *P wave* observations is calculated, with the bins set in 0 µV, 50 µV, 100 µV, 500 µV and 1 mV. Then, the peak of the histogram is taken as the P wave amplitude in the lead, and finally it is required the mean amplitude of all leads to be higher than 50 µV. For example, consider the interpretation in Fig. 4.3 and suppose a cluster will all but the 4th, 10th, and 13th heartbeats. The amplitude of the P waves associated to the clustered heartbeats are 0µV (absent P wave), 42µV, 58µV, 56µV, 49µV, 54µV, 54µV, 58µV, 58µV, 0µV, 49µV. If we calculate the amplitude histogram, the bin [0µV, 50µV] has 5 elements, and the bin [50µV, 100µV] has 6 elements, so in the absence of another lead the *P Wave* feature has a value of 1 for this cluster.

5. *Rhythm*: This feature is obtained from the highest abstraction level of the interpretation procedure: the rhythm hypothesis in which a QRS complex is enclosed. The *Regular* value is assigned to all beats interpreted in a regular rhythm (Bradycardia, Sinus rhythm or Tachycardia), to the odd beats in Bigeminisms, to the first and second beats of each triple in Trigeminisms, and to the beat after a compensatory pause in Extrasystoles and Couplets. The *Advanced* value is assigned to beats interpreted as Extrasystoles or Couplets, as long as the even beats in Bigeminisms and the third beat of each triple in Trigeminisms. The *Atrial fibrillation* value is assigned to all beats interpreted in an atrial fibrillation episode. Finally, the *Delayed* value is assigned to the beats interpreted

as Asystole or as Rhythm Block. For a cluster, this feature is calculated as the most repeated rhythm. If the *Advanced* or *Delayed* tags have a significant frequency in the cluster (over 20%), a correction factor is applied based on the number of *Regular* beats with a higher heart rate (lower in the case of *Delayed*) than the mean heart rate of the cluster. In the case of *Advanced*, by default it is assumed that 50% of *Regular* beats have a higher heart rate than the mean. The number of beats with higher heart rate exceeding this 50% is then added to the number of *Advanced* beats, and if this count exceeds the number of *Regular* beats, the cluster is tagged as *Advanced*. The same procedure is applied for *Delayed* beats considering the number of beats with lower heart rate than the mean. For example, for the interpretation shown in Fig. 4.3 and according to these rules, the 4th, 10th and 13th heartbeats have the *Advanced* value for the *Rhythm* feature, while the others have the *Regular* value.

Feature	Qualitative Values	
Heart Rate	-1:	Lower than 60 beats per minute (Bradycardia)
	0:	Between 60 and 100 bpm (Normal rhythm)
	1:	Between 100 and 150 bpm (Asymptomatic tachycardia)
	2:	Over 150 bpm (Symptomatic tachycardia)
QRS Duration	-1:	Less than 80 milliseconds (Narrow)
	0:	Between 80 ms and 100 ms (Normal)
	1:	Between 100 ms and 120 ms (Abnormal)
	2:	More than 120 ms (Wide)
Heart Axis	-1:	Angle lower than -45° (Inverted axis)
	0:	Angle between -45° and 45° (Balanced axis)
	1:	Angle higher than 45° (Normal axis)
P Wave	0:	Absent P wave, or mean amplitude lower than 50 µV
	1:	Mean amplitude over 50 µV
Rhythm	0:	Regular
	1:	Atrial fibrillation
	2:	Advanced beat
	3:	Delayed beat

Table 4.4: Feature values for single cluster classification

Once these qualitative features have been obtained for each cluster, a simple set of general rules is applied in order to decide the physiological origin of the beats in the cluster. To ensure robustness in this step, which determines the reliability of the comparative classification, only the clusters with a significant number of members are considered. In our case, we require as a rule of thumb a minimum of 30 QRS to ensure a stable mean value for the classification features. Table 4.5 shows the four classification rules considering only the individual features of each cluster. The rules are applied in order, so if the antecedent of one rule is met no further rules are evaluated.

$Rhythm = Regular$ and $Pwave$	→	`NORMAL`
$Rhythm = Atrial\ fib.$ and $Rate \geq 0$	→	`AFIB`
$Rhythm = Advanced$ and $Duration = Narrow$	→	`SVEB`
$Pwave$ and $Duration = Narrow$	→	`NORMAL`

Table 4.5: Single cluster classification rules

Clusters classified as `NORMAL` by rules in table 4.5 are further analyzed to identify possible intraventricular conduction abnormalities such as right bundle branch block (RBBB) and left bundle branch block (LBBB) [Wagner, 2008]. These conditions are detected from the representative QRS morphology, with RBBB requiring the QRS duration to be higher than 100 ms and the QRS morphology in lead V1 to finish with a positive wave with amplitude higher than 0.5 mV. LBBB requires the QRS duration to be higher than 120 ms, and the QRS morphology in lead V1 to be tagged as *QS* or *rS*, according to [Teijeiro et al., 2015].

In the MIT-BIH Arrhythmia Database used for the validation of the proposal, the single cluster classification is able to provide a label for more than 84% of the heartbeats.

Comparative classification

In some records, the variability of the cardiac conduction and rhythm makes it impossible to distinguish a predominant "normality" situation satisfying the rules described for single cluster classification. But even in such circumstances, establishing a baseline cardiac behavior will permit to discriminate the possible transient anomalies by comparison. In our case, if no clusters were classified by the single cluster classification rules, the reference cluster Q_N is selected as follows:

- If there are clusters with more than 30 QRS, with a P wave, and with a duration lower than 120 ms, then the largest one is selected as Q_N.
- If there are clusters with a predominant *Regular* or *Atrial fibrillation* rhythm label, the largest cluster in this set is selected.
- If no clusters match any of these two criteria, the cluster with more complexes interpreted as *Regular* or *Atrial fibrillation* is selected.

The comparative classification stage considers an extended set of features, calculated by comparison between the target unclassified cluster Q_i and the reference cluster Q_N, which are described below. The qualitative values for these features are detailed in table 4.6.

1. *Heart rate difference (Heart Rate$'$)*: Measures the heart rate difference between the target cluster Q_i and a reference cluster Q_N.
2. *Duration difference (QRS Duration$'$)*: Measures the duration difference between the representative QRS complexes of Q_i and Q_N.
3. *Axis difference (Heart Axis$'$)*: Measures the heart axis difference between the representative QRS complexes of Q_i and Q_N.
4. *Amplitude difference (QRS Amplitude$'$)*: This feature measures the relative amplitude difference between the representative QRS of Q_i and Q_N. For this, the lead in which Q_i has a higher amplitude is selected, and the ratio of both amplitudes is calculated. The amplitude is calculated as the difference between the maximum and minimum ECG values inside the region delineated as QRS complex.
5. *Morphological Similarity*: This feature provides a measure of the signal similarity between the QRS onset and offset in all available leads for the representative QRS of Q_i and Q_N. Similarity is obtained by cross-correlation (*xcorr*), and the result is the average value of the maximum for each lead.

On the basis of the attributes of the clusters labeled in the basic classification stage, we have identified two specific contexts that require specific rules for the classification of Q_i:

a) *Wide QRS context:* This context is set on the presence of an artificial pacemaker or if a bundle branch block (left or right) was identified in some cluster. The presence of an

Feature	Qualitative Values	
Heart Rate'	-1:	$Heart\ Rate_N - Heart\ Rate_i \geq 20bpm$ (Slower rate)
	0:	$-20bpm < Heart\ Rate_N - Heart\ Rate_i < 20bpm$ (Equal rate)
	1:	$Heart\ Rate_i - Heart\ Rate_N \geq 20bpm$ (Faster rate)
QRS Duration'	-1:	$QRS\ Duration_N - QRS\ Duration_i \geq 20ms$ (Narrower complex)
	0:	$-20ms < QRS\ Duration_i - QRS\ Duration_N < 20ms$ (Equal duration)
	1:	$20ms \leq QRS\ Duration_i - QRS\ Duration_N < 40ms$ (Wider complex)
	2:	$QRS\ Duration_i - QRS\ Duration_N \geq 40ms$ (Much wider complex)
Heart Axis'	0:	$\lvert Heart\ Axis_N - Heart\ Axis_i \rvert < 45°$ (Equal axis)
	1:	$45° \leq \lvert Heart\ Axis_N - Heart\ Axis_i \rvert < 90°$ (Deviated axis)
	2:	$90° \leq \lvert Heart\ Axis_N - Heart\ Axis_i \rvert < 135°$ (Far deviated axis)
	3:	$\lvert Heart\ Axis_N - Heart\ Axis_i \rvert \geq 135°$ (Opposite axis)
QRS Amplitude'	-1:	$Amplitude_i / Amplitude_N < 0.75$ (Lower amplitude)
	0:	$0.75 \leq Amplitude_i / Amplitude_N \leq 1.25$ (Equal amplitude)
	1:	$Amplitude_i / Amplitude_N > 1.25$ (Higher amplitude)
Morphological Similarity	0:	$xcorr(Q_N, Q_i) < 0.25$ (Very different)
	1:	$0.25 \leq xcorr(Q_N, Q_i) < 0.5$ (Different)
	2:	$0.5 \leq xcorr(Q_N, Q_i) < 0.75$ (Similar)
	3:	$0.75 \leq xcorr(Q_N, Q_i) < 0.9$ (Very similar)
	4:	$xcorr(Q_N, Q_i) \geq 0.9$ (Identical)

Table 4.6: Feature values for comparative classification

artificial pacemaker is set if for some cluster more than 20% of beats are classified as paced by the QRS delineation algorithm [Teijeiro et al., 2015].

b) *Atrial fibrillation context:* This context is set if some cluster was classified as *Atrial fibrillation*.

If none of these specific contexts are recognized, then a Normal Sinus Rhythm context is assumed, and the reference cluster Q_N is selected as the largest cluster classified as normal.

The comparative classification tries first to account for the physiological origin of the cluster, considering the morphological features, and then the temporal location using the rhythm features. If there is an already classified cluster Q_j with identical similarity and equal amplitude than Q_i, then the supraventricular/ventricular origin of Q_i is assumed to be the same of Q_j. If no such Q_j exists, then the origin is determined by considering typical values for the seven morphological features and checking which type is best matched with the feature values of Q_i through a simple majority vote. In the unusual case of tie vote, a supraventricular origin is assumed. Table 4.7 shows these characteristic values for each origin.

S	$Heart\ Axis \geq 0$	$QRS\ Duration < 2$	$Pwave \geq 0$	$Similarity > 2$
	$Heart\ Axis' = 0$	$QRS\ Duration' < 2$	$QRS\ Amplitude' \leq 0$	
V	$Heart\ Axis \leq 0$	$QRS\ Duration > 0$	$Pwave = 0$	$Similarity < 3$
	$Heart\ Axis' > 0$	$QRS\ Duration' > 0$	$QRS\ Amplitude' \neq 0$	

Table 4.7: Characteristic values for supraventricular/ventricular discrimination

Table 4.8 shows the classification rules used for each final resulting class, considering the physiological origin and rhythm-related features. As can be seen, each specific context modulates the rules according to the special characteristics of "normality" they represent. For example, in the *Atrial fibrillation* context classification rules avoid the use of *Rate* and *Rate'* features, due to the highly erratic behavior of the RR interval [Wagner, 2008]. On the other hand, in the *Wide QRS* context, clusters with regular rhythm and a QRS morphology similar to Q_N (high morphological similarity or low axis deviation) are classified as normal, and clusters with advanced rhythm and appreciable morphological differences with respect to Q_N (in duration or similarity) are classified as ectopic ventricular beats. Finally, the *Normal sinus rhythm* context includes a rule for the classification of fusion beats in clusters with an identified supraventricular origin. The reason not to include this rule in the *Atrial fibrillation* and *Wide QRS* contexts is its dependence with the *Duration* and *Rate* features, that are not representative in these contexts.

4.2.5 Results

In the bibliography, the commonly used methodology for heartbeat classification evaluation is the presented by de Chazal *et al.* [de Chazal et al., 2004], focused on the MIT-BIH Arrhythmia database. This methodology excludes the records of patients using pacemakers and divides the remaining into two datasets, DS1 (training) and DS2 (test). In our case, a training set is not necessary and the validation was performed with all records, including those with paced beats. In this manner, we can evaluate the interpretation and classification algorithms without making any a priori assumptions about the characteristics of the ECG signal. Like other

Wide QRS and Normal sinus rhythm contexts		
$Origin = S$		
$Heart\ Rate = 0$ and $QRS\ Duration = 0$ and $Similarity = 4$	→	NORMAL
$Rhythm = 2$ or $Heart\ Rate' = 1$ or $Heart\ Rate = 2$	→	SVEB
$Rhythm = 3$ and $Heart\ Rate' = -1$	→	SVESC
$Default$	→	NORMAL
$Origin = V$		
$Rhythm = 2$ or $Heart\ Rate' = 1$ or $Heart\ Rate = 2$	→	VEB
$Rhythm = 3$ or $Heart\ Rate \leq 0$	→	VESC
$Rhythm = 0$ and $QRS\ Duration = 1$ and $Heart\ Rate = 1$	→	FUSION
$Default$	→	VEB
Atrial fibrillation context		
$Origin = S$		
$Rhythm = 0$ and $QRS\ Duration = 0$ and $Similarity = 4$	→	NORMAL
$Rhythm = 2$	→	SVEB
$Rhythm = 3$	→	SVESC
$Default$	→	NORMAL
$Origin = V$		
$Rhythm = 2$	→	VEB
$Rhythm = 3$ or $Heart\ Rate \leq 0$	→	VESC
$Default$	→	VEB
Wide QRS context		
$Rhythm = 0$ and $QRS\ Duration' = 0$ and $Similarity > 2$	→	NORMAL
$Rhythm = 0$ and $QRS\ Duration' = 0$ and $Heart\ Axis' < 2$	→	NORMAL
$Rhythm = 2$ and $QRS\ Duration > 0$ and $QRS\ Duration' > 0$	→	VEB
$Rhythm = 2$ and $QRS\ Duration > 0$ and $Similarity < 4$	→	VEB
Normal sinus rhythm context		
$Origin = S$		
$Rhythm = 0$ and $QRS\ Duration = 1$ and $Heart\ Rate = 1$	→	FUSION

Table 4.8: Rules for comparative classification

works following this methodology, the beat annotations in the `.atr` files included in the database were used as the initial evidence for the interpretation stage, considering them as *R-Deflection* observations. It should be noted that these annotations contain perfect knowledge about QRS locations, but the interpretation algorithm treats them like any other initial evidence that might be modified during the hypothesize-and-test cycle. For example, Fig. 4.5 shows an ECG fragment that is interpreted as an *Atrial fibrillation* episode, and the 4th *R-Deflection* is incorrectly abstracted in a *T wave* instead of a *QRS complex*, leading to a false negative in the ’V’ class. Even if it is detrimental to the results, we believe this is the most fair validation strategy, since the main distinctive feature of the adopted reasoning paradigm is to not assume any previous conclusion as unfailing.

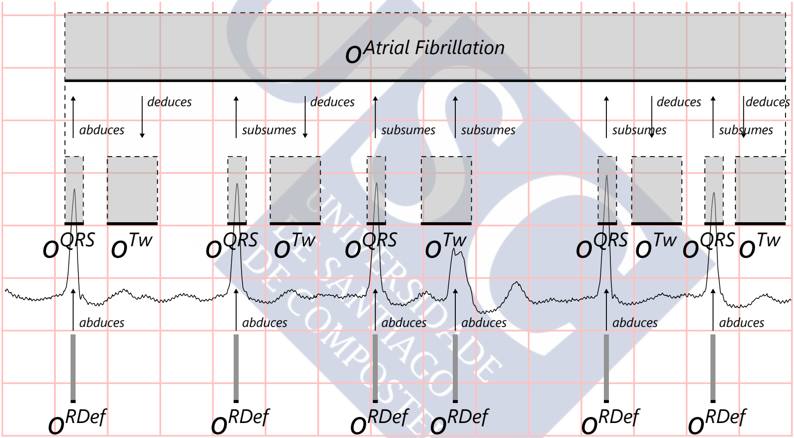


Figure 4.5: False negative VEB detection due to an oversimplified interpretation [Source: MIT-BIH arrhythmia DB, recording: 202, lead: MLII, between 27:58.600 and 28:02.100. *R-Deflections* are the reference `.atr` annotations]

Table 4.9 shows the confusion matrices of the classification results for all the 48 records in the database, for the 44 records in the DS1 and DS2 sets, and for the records in the DS2 dataset commonly used in comparison studies. These matrices were obtained as described in section 4.2.1, fixing the incorrect label association for MIT-BIH classes ’j’ and ’e’.

In addition to the five standard beat classes defined by the AAMI, we included the **O** class to represent false positives and false negatives in QRS detection due to the modifications introduced by the abductive interpretation process in the initial set of beat annotations. The **Q** class represents beats with unknown origin, but the standard tools also use it to represent paced

		Classification Result					
Ground Truth		O	N	S	V	F	Q
Full database	O	0	19	6	49	0	0
	N	115	89459	355	446	9	2
	S	15	330	2575	106	0	0
	V	85	275	116	6705	41	14
	F	1	273	5	80	444	0
	Q	2	713	0	143	1	7184

		Classification Result					
Ground Truth		O	N	S	V	F	Q
DS1+DS2	O	0	19	6	49	0	0
	N	115	88960	355	441	9	0
	S	15	330	2575	106	0	0
	V	73	273	116	6506	41	0
	F	1	273	5	80	444	0
	Q	0	13	0	1	1	0

		Classification Result					
Ground Truth		O	N	S	V	F	Q
DS2	O	0	2	0	0	0	0
	N	84	43634	277	47	4	0
	S	10	199	1787	54	0	0
	V	37	50	64	3048	22	0
	F	0	140	5	42	201	0
	Q	0	7	0	0	0	0

Table 4.9: Confusion Matrices of the Beat Classification Results for all 48 records in the MIT-BIH Arrhythmia Database (top), for the 44 records in the DS1 and DS2 sets (middle), and for the 22 records in the DS2 test set (bottom)

Dataset	Method	VEB		SVEB	
		Se	P^+	Se	P^+
DS1+DS2	This work - Automatic	92.82	92.23	85.10	84.51
	Llamedo *et al.* [Llamedo and Martinez, 2012] - Assisted	90±1	97±0	89±2	88±3
	Kiranyaz *et al.* [Kiranyaz et al., 2016] - Assisted	93.9	90.6	60.3	63.5
	Ince *et al.* [Ince et al., 2009] - Assisted	84.6	87.4	63.5	53.7
	Llamedo *et al.* [Llamedo and Martinez, 2012] - Automatic	80±2	82±3	76±2	43±2
DS2	This work - Automatic	94.63	96.79	87.17	83.98
	Llamedo *et al.* [Llamedo and Martinez, 2012] - Assisted	93±1	97±1	92±1	90±3
	Kiranyaz *et al.* [Kiranyaz et al., 2016] - Assisted	95.0	89.5	64.6	62.1
	Oster *et al.* [Oster et al., 2015] - Assisted	92.7	96.2	NA	NA
	Chazal *et al.* [de Chazal, 2014] - Assisted	93.4	97.0	94.0	62.5
	Chazal *et al.* [de Chazal and Reilly, 2006] - Assisted	94.3	96.2	87.7	47.0
	Zhang *et al.* [Zhang et al., 2014] - Automatic	85.48	92.75	79.06	35.98
	Llamedo *et al.* [Llamedo and Martinez, 2012] - Automatic	89±1	87±1	79±2	46±2
	Chazal *et al.* [de Chazal et al., 2004] - Automatic	77.7	81.9	75.9	38.5

Table 4.10: VEB and SVEB classification performance and comparison with the most relevant automatic and assisted methods of the state-of-the-art

beats and the fusion of paced and normal beats. This is the reason for the great proportional difference of this class in the first matrix with respect to the other.

Table 4.10 shows a comparison of the classification performance with the most relevant algorithms of the state-of-the-art, using sensitivity and positive predictive value of the ventricular and supraventricular ectopic beat classes. It is worth noting that most of the proposals are assisted, meaning that they require an expert to manually annotate some beats. In this sense, some of the assisted methods are closer to a *clustering* approach rather than a *classification* approach, since once the expert has labeled one of the beats in each group the remaining beats are assigned the same label. In contrast, the present proposal provides a method that autonomously assigns a label to each and every beat in a record.

Results show that the proposed method outperforms any other automatic method in the bibliography, and even most of the assisted methods. With respect to automatic approaches, results show an average improvement of around 10% in the sensitivity and positive predictive value of ventricular ectopic beats and in the sensitivity of supraventricular ectopic beats. This improvement is increased to around 40% in the positive predictive value of supraventricular ectopic beats. With respect to assisted approaches, the improvement in the sensitivity and positive predictive value of ventricular ectopic beats is around 1%. For supraventricular ectopic beats, the average improvement in sensitivity and positive predictive value is around 8% and 20%, respectively. The statistical significance of these differences could not be assessed due to the lack of detailed validation results for the other methods, but the magnitude of the differences, specially in the specificity of the supraventricular class, is notable beyond any doubt. In this sense, the only algorithm with better performance is the assisted version of [Llamedo and Martinez, 2012].

In general, the classification of supraventricular ectopic beats (S class) is the most remarkable improvement with respect to the state-of-the-art. The main difficulty to classify this type of beat is that the morphology is usually very similar to the normal morphology, and therefore class separation has to be made according to rhythm-related features. In this sense, the information provided by the rhythm level of the abductive interpretation, and in particular the *Rhythm* feature detailed in section 4.2.4, goes far beyond than the classical analysis of the RR intervals around the beat, which is severely affected by phenomena such as atrial fibrillation. Also, it should be noted that results of other approaches in table 4.10 are affected by the incorrect label association discussed in section 4.2.1, so the actual sensitivity in the classification of the S class is expected to be lower.

CHAPTER 5

DISCUSSION

This chapter discusses the main theoretical aspects of the proposed interpretation framework, with emphasis in those characteristics leading to overcoming some of the most important weaknesses of traditional classifiers solely based on a deductive perspective, and which are the basis for the practical results presented in Chapter 4.

5.1 Avoiding a casuistry-based interpretation

In the time domain, classification-based recognition of multiple processes occurring concurrently usually leads to a casuistry-based proliferation of classes, in which a new class is usually needed for each possible superposition of processes in order to properly identify all situations. It is common to use a representation in the transform domain, where certain regular processes are easily separable, although at the expense of a cumbersome representation of the temporal information [Mörchen, 2003]. In contrast, in the present framework, the hypothesize-and-test cycle aims to conjecture those hypotheses that best explain the available evidence, including simultaneous hypotheses in a natural way as long as these are not mutually exclusive.

ECG interpretation provides some interesting examples of this type of problem. Atrial fibrillation, a common heart arrhythmia caused by the independent and erratic contractions of the atrial muscle fibers, is characterized by an *irregularly irregular* heart rhythm [Wagner, 2008]. Most of the classification techniques for the identification of atrial fibrillation are based on the analysis of the time interval between consecutive beats, attempting to detect this irregularity [Petrenas et al., 2015]. These techniques offer good results in those situations in which atrial fibrillation is the only anomaly, but they fail to properly identify complex scenarios which go beyond the distinction between atrial fibrillation and normal rhythm. In the strip shown in Figure 5.1, obtained during a pilot study for the home follow-up of patients with cardiac diseases [Sacchi et al., 2015], such a classifier would wrongly identify this segment as an atrial fibrillation episode, since the observed rhythm variability is consistent with the description of this arrhythmia. In contrast, the present interpretation framework properly explains the first five beats as a sinus bradycardia, compatible with the presence of a premature ectopic beat in the second position, followed by a trigeminy pattern during six beats, and finally another ectopic beat with a morphology change. The reason to choose this interpretation, despite being more complex than the atrial fibrillation explanation, is that it is able to abstract some of the small P waves before the QRS complexes, increasing the interpretation coverage.

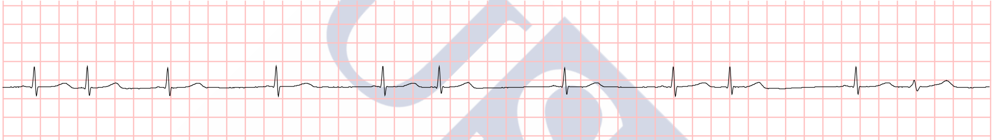


Figure 5.1: False atrial fibrillation episode. [Source: Mobiguide Project [Sacchi et al., 2015], private recording]

5.2 Coping with ignorance

Most of the classifiers solve a separability problem among classes, either by learning from a training set or by eliciting prior knowledge, and these are implicitly based on the closed-world assumption, i.e., every new instance to be classified is assigned to one of the predefined classes. Such classifiers may additionally include a 'reject' option for all those instances that could be misclassified since they appear too close to the classification boundaries [Chow, 1970; Fumera et al., 2000]. This reject option is added as another possible answer expressing doubt. However, such classifiers fail to classify new instances of unknown classes, since they cannot express ignorance. An approach to this problem can be found in novelty detection proposals [Pimentel et al., 2014], which can detect when a new instance does not fit any of the predefined classes as it substantially differs from those instances available during training. Still, these are limited to a common feature representation for every instance, hindering the identification of what is unintelligible from the available knowledge.

In the present framework, it is assumed that a set of abstraction patterns results from the compilation of the knowledge available on poorly or incompletely understood domains, covering only a portion of observable phenomena, many of which cannot be identified with certainty or even known a priori. In this sense, those initial observations that have not been abstracted yet, at some time in the interpretation process, can be considered 'unobserved'; and those initial observations that will not be finally abstracted by any abstraction pattern, can be considered 'unintelligible'. This provides an expression of ignorance as a common result of the interpretation problem. As long as the abstraction model is incomplete, the non-coverage of some piece of evidence by any interpretation is an expression of partial ignorance. In the extreme case, the trivial interpretation I_0 may be a proper solution for an interpretation problem, expressing total ignorance. Furthermore, abduction naturally includes the notion of ignorance in the reasoning process, since any single piece of evidence can be sufficient to guess an interpretation, and the hypothesize-and-test cycle can be understood as a process of incremental addition of evidence against an initial state of ignorance, while being able to provide an interpretation at any time based on the available evidence.

As an example, consider the interpretation problem illustrated in Figure 5.2. Let the initial evidence be the set of QRS annotations obtained by a state-of-the art detection algorithm [Zong et al., 2003b]. In this short strip, the eighth and ninth annotations correspond to false positives caused by the presence of noise. A classification-based strategy processes these two annotations as true QRS complexes, and the monotone nature of the reasoning prevents their possible refutation, probably leading to beat misclassification and false arrhythmia detection, with errors propagating onwards to the end of the processing. In contrast, the present framework provides a single normal rhythm as the best interpretation, which explains all but the two aforementioned annotations, which are ignored and considered unintelligible in the available model. It is also worth noting the ability of this framework to integrate the results of an available classifier (the QRS detection algorithm) as a type of constraint specification in the interpretation cycle.

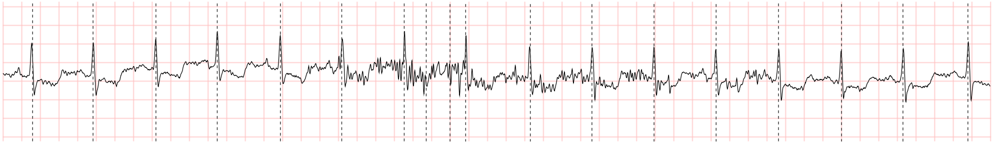


Figure 5.2: Unintelligible evidence due to noise. [Source: MIT-BIH arrhythmia DB, recording: 112, between 13:46.200 and 13:56.700]

5.3 Looking for missing evidence

The application of the classification paradigm to pattern detection also entails the potential risk of providing false negative results. In the worst case, a false negative result may be interpreted by a decision maker as evidence of absence, leading to interpretation errors with their subsequent costs, or in the best case as an absence of evidence caused by the lack of a proper detection instrument.

Even though abduction is fallible, and false negative results persist, the hypothesize-and-test cycle involves a prediction mechanism that points to missing evidence that is expected and, moreover, estimates when it should appear. Both the bottom-up and top-down processing performed in this cycle reinforces confidence in the interpretation, since the semantics of any conclusion is widened according to its explanatory power.

As an example, consider the interpretation problem illustrated in Figure 5.3. The initial evidence is again a set of QRS annotations obtained by a state-of-the-art detection algorithm [Zong et al., 2003b]. Note that the eighth beat has not been annotated, due to a sudden decrease in the signal amplitude. This error can be amended in the hypothesize-and-test cycle, since the normal rhythm hypothesis that abstracts the first seven QRS annotations predicts the following QRS to be in the position of the missing annotation, and the PREDICT() procedure can look for this (e.g., checking an alternative set of constraints).

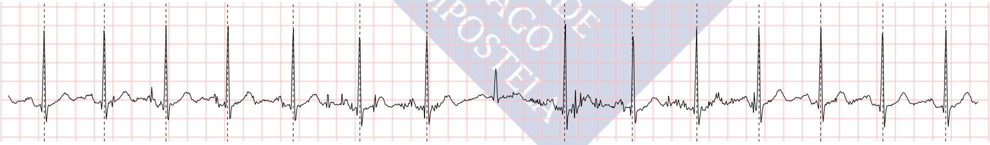


Figure 5.3: Missing evidence that may be discovered by prediction. [Source: MIT-BIH normal sinus rhythm DB, recording: 18184, between 09:12:45.000 and 09:12:55.500]

5.4 Interpretability of the reasoning process and the results

The interpretability of a reasoning formalism, defined as the ability to understand and evaluate its conclusions, is an essential feature for achieving an adequate confidence in decision making [Nauck and Kruse, 1999]. In this sense, there are a number of classification methods with good interpretability; however, the methods that typically offer the best performance belong to the so-called *black box* approaches.

The present interpretation framework is able to provide a justification of any result in relation to the available abstraction model. Beyond the truth concern of deductive inference, that is, in regards to state what is true, abduction concerns theory formation, that is, in regards to explain why is true [Poole, 1990]. Given any solution or partial solution of an interpretation problem, the searching path up to I_0 gives full details of all the reasoning steps taken to this end, and any abstraction hypothesis can be traced back to the information supporting it.

This interpretation framework is also able to answer the question of why a certain hypothesis has been rejected or neglected at any reasoning step. This is done by exploring the branches outside the path between I_0 and the solution. Since the K exploration parameter within the CONSTRUE() and CONSTRUE-ONLINE() algorithms has been chosen as the maximum number of hypotheses that may explain a given observable, it is possible to reproduce the reasoning steps taken in the conjecture of any abstraction hypothesis, and to check why this did not succeed (non-satisfaction of pattern constraints, lower coverage, etc.). This can be useful in building and refining the knowledge base.

5.5 Semantical considerations

An abstraction model, built on a set of abstraction patterns, establishes a causal responsibility for the behavior observed in a complex system [Josephson and Josephson, 1994]. This responsibility is expressed in the language of processes: a process is said to be observable if it is assumed that it causes a recognizable trace in the physical quantity to be interpreted. The successive proposal of more abstract hypotheses provides successive reinterpretations of the observations as more processes are involved, expanding its scale and scope by going up in the abstraction hierarchy. This notion of causality is behind perception, i.e., concerned with the explanation of sensory data, in contrast with the notion of causality in diagnosis, concerned with the explanation of abnormality [Console and Torasso, 1991]. In this way, an abstraction model allows to provide an explanation as a projection of elements in an accepted theory for the solution of a problem in a particular domain [Callaway, 2014].

The intuition behind an abstraction pattern can be described from two different perspectives, which we will call backward and forward logical meaning, following a reasoning similar to that of Brusoni et al. [1998]:

- **Backward meaning**. From the backward reading of an abstraction pattern P, a hypothesis h is a possible abstraction of $m_1, \dots, m_n$, provided that the constraints in C_P hold.

An abstraction pattern satisfies the *compositionality principle* of abductive reasoning, and hence an abstraction hypothesis can be conjectured from a single piece of evidence, and new pieces of evidence can be added later [Flach, 1996], so h *abstracts* $m_1\{C_P()\}$, h *abstracts* m_1,m_2 $\{C_P()\}$, h *abstracts* m_1,m_2,m_3 $\{C_P()\}$, and so on. On the other hand, if there are multiple ways of observing h by means of multiple patterns, and their respective constraints are inconsistent with the evidence, we do not conclude $\neg h$, interpreted as failure to prove h; we will only conclude $\neg h$ in all those interpretations conjecturing an observation of a different h', being h and h' mutually exclusive.

– **Forward meaning**. An abductive observation is built upon an archetypical representation of a hypothesis h, creating an observation as an instance of h by estimating, from the available evidence, its attribute values $\mathbf{A}$ and its temporal location T^b and T^e by means of an observation procedure $\pi_P()$. From a forward reading, assuming h is true there is an observation for each observable of the set $m_1,\ldots,m_n$ such that the constraints in C_P hold. However, the estimated nature of abstraction does not usually allow us to infer, from the observation of h, the same observations of $m_1,\ldots,m_n$ that have been abstracted into h. We must presume instead that assuming h is true entails the occurrence of an observation for each observable of $m_1,\ldots,m_n$, without necessarily entailing its attribute values and its temporal location.

Both the forward and the backward meanings of an abstraction pattern support the incremental building of an interpretation in the present framework. Thus, what initially was defined as a set covering problem of a time series fragment —a completely intractable problem as it moves away from a toy example— can be feasibly solved if it is properly structured in a set of abstraction levels, on which four reasoning modes (abduction, deduction, subsumption and prediction) can make a more efficient search of the best explanation under a parsimony criterion. Moreover, this incremental reasoning must primarily follow the time direction, since the available knowledge is usually compiled in the form of a set of processes that can be expected to be found in a certain sequence, which underscores the anticipatory information contained in the evidence.

Representing and reasoning about context is a relevant issue both in model-based diagnosis [Brusoni et al., 1998; Console and Torasso, 1991; Peng and Reggia, 1990] and in temporal abstraction [Shahar, 1996, 1997; Combi et al., 2010]. A contextual observation is nothing more than another observation that is not supposed to be explained nor abstracted. In most of the bibliography, the distinction between these two roles must be defined beforehand. Several other works enable the same observation to play different roles in different causal patterns, thus providing some general operations for expressing common changes made by the context in a diagnostic pattern [Juárez et al., 2008; Palma et al., 2006]. In the proposed interpretation framework, an observation can either be part of the evidence to be explained by the process hypothesized in a certain abstraction pattern, or can be part of the environment in another abstraction pattern. The environment of an abstraction pattern is a set of evidence needed to support the hypothesis of the pattern but that is not explained by the process underlying that hypothesis, and therefore it captures the notion of context. For example, in order to conjecture a turbulence process from the observation of a rapid shaking of our seat it is necessary to be in a plane in flight, even though being in a plane is not explained by the turbulence process. Both types of observation play a part in the hypothesize-and-test cycle, with the only difference that observations of the environment of an abstraction pattern are not expected to be abstracted by this pattern. Hence, observations of the environment are naturally included in the deduction, subsumption and prediction reasoning modes.


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Conclusions

This work proposes a novel knowledge-based framework for time series interpretation, grounded on the initial hypothesis that abduction provides a proper reasoning paradigm to overcome the major limitations of traditional classification-based approaches, and inspired by how humans identify and characterize the patterns appearing in a time series. This framework relies on some basic assumptions: (i) interpretation of the behavior of a system from the set of available observations is a sort of conjecturing, and as such follows the logic of abduction; (ii) the interpretation task involves both bottom-up and top-down processing of information along a set of abstraction levels; (iii) at the lower levels of abstraction, the interpretation task is a form of precompiled knowledge-based pattern recognition; (iv) the interpretation task involves both the representation of time and reasoning about time and along time.

The framework provides a knowledge representation formalism based on the notion of *temporal abstraction pattern*. A temporal abstraction pattern defines an abstraction relation between observables and provides the knowledge and methods to conjecture new observations from previous ones, thus assigning a dual role to the observables: as a hypothesis on the observation of an underlying process, and as the pieces of evidence supporting that hypothesis. This relation between observables may be established in multiple abstraction levels by using the hypothesis observable of an abstraction pattern as the evidence for a pattern in a higher level, thus building a hierarchical language for the description and characterization of the processes that can be observed in a specific domain.

In addition to the knowledge representation formalism, a set of algorithms for the resolution of interpretation problems is provided, implementing a hypothesize-and-test cycle guided by an attentional mechanism that dynamically builds an interpretation according to four heuristic principles: 1) a *coverage principle*, which maximizes the explained evidence; 2) a *simplicity principle*, which minimizes the number of hypotheses in the final interpretation; 3) an *abstrac-*

tion principle, which prefers the use of explanatory hypotheses in higher abstraction levels; and 4) a *predictability principle*, which prioritizes interpretations that properly predict future evidence. The CONSTRUE() algorithm is proposed for the resolution of *off-line* interpretation problems, in which all the evidence is available before the beginning of the interpretation. This algorithm performs a complete search on the space of consistent interpretations by means of four reasoning modes (abduction, deduction, subsumption and prediction), exploiting the behavior of abstraction grammars as pattern generators and incrementally approaching the best explanation. An evolution of this algorithm, named CONSTRUE-ONLINE(), is proposed for the resolution of *on-line* interpretation problems, in which new evidence may be acquired continuously during the interpretation. In this version completeness is sacrificed for the sake of greater efficiency in the search, including a pruning strategy to limit resource consumption. Also, the knowledge structure of the abstraction model is exploited by identifying sets of operational necessary and sufficient conditions, which allow the hypothesize-and-test cycle to focus in those pieces of evidence with greater capacity to refute or confirm a given hypothesis. Finally, a temporal knowledge representation formalism based on the Simple Temporal Problem (STP) is proposed to accelerate temporal reasoning and constraints checking operations.

The proposed framework has been applied to some well-known problems in the electrocardiogram analysis domain. On the one hand, the non-monotonic nature of abductive reasoning allows to correct both false negative and false positive QRS detections in the state-of-the-art `WQRS` algorithm [Zong et al., 2003b], thus improving an essential stage in classical ECG processing algorithms. On the other hand, an interpretation in multiple abstraction levels has proven to be valuable for the construction of a set of high-level features describing the heart function in the same terms used by experts, enabling us to build a simple rule-based heartbeat classifier that outperforms state-of-the-art automatic classifiers, and even most classifiers requiring expert assistance to provide a result. These results have been published in [Teijeiro et al., 2014] and [Teijeiro et al., 2016], respectively. The key factor behind these results is the non-monotonic nature of the hypothesize-and-test cycle, making it possible to exploit the complementarity between bottom-up and top-down processing, in order to find the best explanation consistent with the evidence. As in perception, ECG interpretation is assumed to be mostly bottom up, although top-down processing has proven to be decisive to cope with noise, artifacts or ambiguities in the signal. As an additional result, a comprehensive knowledge base for the ECG domain has been formalized, describing processes at the three abstraction levels commonly used by experts in electrocardiogram interpretation: signal deviation, electrical activation/recovery of the heart chambers, and rhythm analysis. This abstraction model is described in detail in Appendix A, and with the aim of supporting reproducible research the full source code implementing both the abstraction model and the interpretation algorithms has been published under an Open Source License.

The present work takes the first steps in a new line of research for model-based interpretation and analysis of temporal data, and a number of issues should be addressed in the future work to make this proposal more general and suitable for a wider range of problems. The main research directions we think are worth exploring can be summarized as follows:

- The proposed framework formulates an interpretation problem as an *abduction problem with constraints*, where the purpose of the interpretation is to find a set of hypotheses covering all the findings in the initial evidence while satisfying a set of constraints on the attribute and temporal values of the observations. Thus, *consistency* is the only property to evaluate the plausibility of a hypothesis as a true/false value, and any evoked hypothesis (no matter how unusual it is) for which inconsistent evidence cannot be found is considered as plausible and explored in the interpretation cycle. Even though this simple approach has provided remarkable results, it can be expected that the inclusion of a hypothesis evaluation scheme, typically based on probability [Peng and Reggia, 1990; Poole, 2000] or possibility [Dubois and Prade, 1995; Palma et al., 2006] theories, will allow us to better discriminate between plausible and implausible hypotheses, leading to better explanations with fewer computational requirements. However, there are a number of considerations that must be taken into account. On the one hand, forcing the user to include probabilistic or possibilistic knowledge in the already demanding task of defining abstraction grammars may be unworkable for certain domains. For this reason, any evaluation scheme should work as an extension of the current consistency-based approach for hypothesis discrimination, and not require additional knowledge. In this sense, some authors claim that human abductive reasoning relies on a coarse notion of plausibility with rough qualitative confidence intervals, and even so it provides excellent results [Josephson and Josephson, 1994]. In addition, if the discrimination between two competing hypotheses depends on a precise calculation, then the difference between the best hypothesis and the second best hypothesis could be small, and thus a confident answer cannot be drawn.

- The expressiveness of abstraction grammars as a knowledge representation model should be enhanced to support the definition of a broader range of patterns, and particularly those describing the *absence* of some piece of evidence, thus requiring the notion of negated events. A potential starting point is the *exclusion relation* defined in page 31. For each possible observable $q \in \mathcal{Q}$, it should be possible to define an observable $\neg q$ representing the "absence of q", and include q and $\neg q$ in the exclusion relation of the domain. The grammar conjecturing $\neg q$ may define its own set of abstraction patterns, requiring different sets of present (or absent) evidence to support the hypothesis of "absence of q". Another possible solution not requiring the definition of specific knowledge is to consider $\neg q$ as the failure to find a consistent hypothesis supporting q. The drawbacks of this approach is that exploring a negation would be very costly, and that the semantics of a negated observation would be more restrictive than the semantics of a positive observation. For example, in a noisy interval it can be desirable that both q and $\neg q$ are treated as consistent hypotheses.

- The heuristics guiding the exploration in an on-line interpretation problem are easy to calculate and they intuitively capture the coverage and simplicity principles. In practice, they have proven to be useful for constructing valuable interpretations while restraining the exploration to acceptable levels. However, these heuristics should be improved to avoid some counterintuitive side effects. For example, consider an on-line interpretation problem that, eventually, requires the explanation of a set of temporally ordered observations $\mathcal{O} = \{o_1, \ldots, o_n\}$. Now suppose that the available knowledge supports the generation of two competing interpretations: I_1, which explains all the observations but o_1, and I_2, which explains all the observations but o_n. Even though both interpretations have the same final coverage, the on-line interpretation algorithm will prioritize the exploration leading to I_2 due to a higher coverage at the beginning, postponing the advancement towards I_1 until all other interpretations explaining o_1 have been explored. Also, if the search space is large enough, a pruning operation will discard the nodes leading to I_1 before they can be explored.

- This work demonstrates that there is still room for improvement in automatic ECG analysis, and particularly in heartbeat classification. This is essential to achieve an effective transfer of these techniques to the clinical routine, integrated in the bedside instrumentation or in the emergent home monitoring. To this end, the evolution of the present work will focus on three main objectives: On the one hand, beat labeling should be more strongly integrated in the hypothesize-and-test cycle. In the proposal discussed in Section 4.2, classification is performed at the end, after the interpretation makes the necessary features available. A better result could be expected if every class label for each QRS observation is considered as a conjecture, enabling to correct it on the basis of posterior evidence, at the expense of a greater complexity of reasoning. In addition, in a dynamic classification approach notions like "normality context" for comparative classification are no longer static, and they have to be updated during the interpretation process. Also, we aim to meet real-time constraints in the execution of the full interpretation cycle.

- The requirement to manually define abstraction models is an important limitation for the adoption of the proposed framework. Thus, it is worth exploring different possibilities for the inclusion of machine learning mechanisms, both for the definition of the knowledge base and within the hypothesize-and-test cycle. A first approach may address the automatic calibration of the temporal and value constraints among recurrent findings in abstraction grammars generating an infinite number of patterns. In this manner, for example, temporal constraints between consecutive heartbeats in a normal rhythm abstraction grammar could be adapted to the characteristics of the subject whose ECG is being interpreted, allowing the identification of possible deviations from normality with greater sensitivity. On the other hand, the automatic learning of abstraction grammars would entail a great advance in the practical capabilities of the framework, making it able to address a greater range of problems, and not limited to the exploitation of expert knowledge but also to the generation of new knowledge, which is the greatest quality of abductive reasoning. In this sense, there are several unsupervised machine learning methods for discovering temporal patterns, and particularly Simple Temporal Networks [Alvarez et al., 2013] that may serve as a starting point for such an approach.

- Finally, the present framework has to be tested in different domains and with other types of signals beyond the ECG. In the field of biosignal interpretation, it is of great interest to combine simultaneous signal types to provide more robust algorithms [Silva et al., 2015]. From our perspective, the inclusion of multiple time series in the hypothesize-and-test cycle is completely straightforward, as long as it provides more evidence for the description and characterization of the processes underlying the full set of signals, and higher abstraction levels can naturally integrate information of various kinds. For example, it is easy to extend the *Heartbeat* abstraction described in this work to include the cycles observed in a blood pressure signal [Zong et al., 2003a]. With respect to the application in different domains, an example of a problem that fits the proposed approach is *chronicle recognition* [Barták et al., 2014]. In this case, the initial evidence is a potentially large collection of time-stamped data, and the objective is to summarize it in a set of chronicles, describing more abstract events and relations. Notions like the focus of attention or the use of abstraction levels have already been used by other authors [Dousson and Le Maigat, 2007], albeit from a different, non abductive perspective.

APPENDIX A

ECG ABSTRACTION MODEL

This appendix describes the knowledge base for the interpretation of multi-lead ECG signals. First, the set of observables $\mathcal{Q}$ is described, relating them with the underlying processes they represent. Then, the grammars designed to generate the abstraction patterns supporting the hypothesize-and-test cycle are explained, organized in abstraction levels. A couple of special symbols, '#' and '@' are used to describe the membership of a finding to the E_P or A_P set, respectively. The full source code implementing this abstraction model is available in the project public repository [1].

A.1 Observables alphabet

The observables used to describe the abstraction patterns for the ECG interpretation problem are organized into the following three abstraction levels:

A.1.1 Signal deviation

Observables in this abstraction level correspond to signal deviation phenomena, and include a set of intervals or time points labeled as *deflections*, consistent with the presence of electrical activity from the cardiac muscle. Observations from this level constitute the initial evidence for the abductive interpretation process. There are two observables in this category:

---

[1] `https://github.com/citiususc/construe`

- *Deflection*: Signal deviation consistent with the electrical activity of the cardiac muscle fibers.
- *R-Deflection*: Signal deviation consistent with the electrical activity generated in the ventricular activation.

As can be seen in the description, *R-Deflection* is a particular case of a *Deflection*, so observables in this abstraction level satisfy the generalization relation *R-Deflection* `is a` *Deflection*.

A.1.2 Electrical activation/recovery

This abstraction level includes observables representing intracardiac conduction phenomena. These can be conjectured taking as evidence observations from the first abstraction level, and provide a description of the ECG as a sequence of waves corresponding to the atrial activation, ventricular activation and ventricular recovery processes. It is the lowest level of physiological phenomena that can be identified in an ECG. There are three observables in this level:

- *P wave*: Represents the atrial electrical activation.
- *QRS complex*: Represents the ventricular electrical activation.
- *T wave*: Represents the ventricular electrical recovery.

A.1.3 Cardiac rhythms

This abstraction level collects the observables characterizing the *cardiac rhythm* under different physiological conditions, including normal rhythm and arrhythmias. An interesting property of this level is that it guarantees a full coverage of the interpreted time interval, this is, there are no gaps between the observations of this level. For this, all patterns are left recursive, requiring the observation of a *Cardiac Rhythm* as the first finding, and setting the beginning of the new rhythm to the end of the previous one. Observables in this level are:

- *Rhythm Start*: Represents the beginning of a sequence of cardiac rhythms, and it is included it in the knowledge base to allow to finish left recursion defined by the other rhythm patterns. There can be only one *Rhythm Start* observation in a given interpretation.

- *Sinus Rhythm*: Normal rhythm originated in the sinoatrial node.
- *Bradycardia*: Regular rhythm with a low frequency heart rate (under 60 beats per minute).
- *Tachycardia*: Regular rhythm with a high frequency heart rate (over 100 beats per minute).
- *Extrasystole*: Premature contraction of the heart.
- *Bigeminy*: Rhythm pattern in which every sinus beat is followed by a premature beat.
- *Trigeminy*: Rhythm pattern in which every second sinus beat is followed by a premature beat.
- *Couplet*: Concurrence of two consecutive extrasystoles.
- *Rhythm Block*: One-time elongation of the cardiac rhythm.
- *Atrial Fibrillation*: Arrhythmia caused by the independent and erratic contractions of the atrial muscle fibers, characterized by an irregularly irregular heart rhythm.
- *Ventricular Fibrillation*: Rapid ventricular activity without discernible QRS complexes or T waves in the ECG.
- *Asystole*: Interval of absence of ventricular activity.
- *Cardiac Rhythm*: Observable that generalizes all the other observables in this level.

A.2 Segmentation grammars

Segmentation grammars support the generation of hypotheses in the abstraction level of electrical activation/recovery phenomena.

A.2.1 QRS complex grammar

$$H \rightarrow \textit{R-Deflection}@\ \{tc_1, vc_1\}$$

with the following constraints (all units are milliseconds):

$$\begin{aligned} tc_1 &= \{\tau c(T_{RDef}, T_H^b) = [-80,80], \tau c(T_{RDef}, T_H^e) = [15,400]\} \\ vc_1 &= \ qrsdel() \end{aligned}$$

The grammar for the abstraction of a QRS complex hypothesis is extremely simple, generating a single pattern abstracting an *R-Deflection* observation. Temporal constraints tc_1 delimit the temporal interval for the delineation of the complex, and the $qrsdel()$ algorithm performs the delineation and checks some additional value constraints. This algorithm is fully described in Appendix B.

A.2.2 P wave grammar

$$H \rightarrow Deflection@A\ \{tc_1\}$$

$$A \rightarrow QRS\#\ \{tc_2, vc_2\}$$

$$\begin{aligned} tc_1 = \{&\tau c(T_H^b, T_H^e) = [40,250], && \tau c(T_{Def}^b, T_{Def}^e) = [0,200], \\ &\tau c(T_H^b, T_{Def}^b) = [-350,350], && \tau c(T_H^e, T_{Def}^e) = [-350,350], \\ &\tau c(T_{Def}^b, T_H^e) = [0,\infty), && \tau c(T_H^b, T_{Def}^e) = [0,\infty), \\ &\tau c(T_{Def}^b, T_{QRS}^b) = [20,400], && \tau c(T_{Def}^e, T_{QRS}^b) = [20,400]\} \\ tc_2 = \{&\tau c(T_H^b, T_{QRS}^b) = [80,400], && \tau c(T_H^e, T_{QRS}^b) = [20,200]\} \\ vc_2 = \ &delineate\_pwave() \end{aligned}$$

The grammar builds a single pattern on the basis of two findings: a *Deflection* that is abstracted in the resultant *P wave*, and an environment *QRS* that must be located short after the end of the *P wave* hypothesis. The value constraints vc_2 define the method to delineate the hypothesis, which is based on the use of a one-class Support Vector Machine [Schölkopf et al., 2001] trained with delineated P wave examples from the QT database [Laguna et al., 1997].

It is worth mentioning that from a physiological point of view, it is not necessary the presence of a *QRS* complex for the observation of a *P wave* (for example in patients with atrial flutter it is common to observe a rapid sequence of *P waves* without associated *QRS* complex [Wagner, 2008]). However, from a practical perspective this definition simplifies the observation procedure and avoids many false positives.

A.2.3 T wave grammar

$$H \rightarrow QRS\#A \; \{tc_1\}$$
$$A \rightarrow Deflection@ \; \{tc_2, vc_2\}$$

$$\begin{aligned}
tc_1 = \{&\tau c(T^e_{QRS}, T^b_H) = [0, 250], \quad \tau c(T^b_{QRS}, T^e_H) = [250, 900],\\
&\tau c(T^e_{QRS}, T^e_H) = [150, 750], \; \tau c(T^b_H, T^e_H) = [80, 450]\}\\
tc_2 = \{&\tau c(T^b_{Def}, T^e_{Def}) = [0, 450], \quad \tau c(T^b_H, T^b_{Def}) = [-450, 300],\\
&\tau c(T^e_H, T^e_{Def}) = [-300, 80], \; \tau c(T^b_{Def}, T^e_H) = [0, \infty),\\
&\tau c(T^b_H, T^e_{Def}) = [0, \infty), \qquad \tau c(T^b_H, T^e_H) = [T^e_{QRS} - T^b_{QRS} - 40, \infty),\\
&\tau c(T^b_{QRS}, T^e_{Def}) = [0, 900], \quad \tau c(T^e_{QRS}, T^b_{Def}) = [0, \infty)\}\\
vc_2 = \; & delineate\_twave()
\end{aligned}$$

The grammar for the T wave pattern is formally very similar to the P wave grammar, but altering the order of the findings. It requires the observation of a context *QRS complex* and a posterior abstracted *Deflection*. Temporal constraints tc_1 require the *QRS* to be located before but near the beginning of the *T wave*, and introduce the common constraints in the QT interval and ST segment duration. On the other hand, constraints tc_2 define the temporal relation between the observed *Deflection* and the *T wave* hypothesis, and the value constraints vc_2 define the delineation method for the *T wave*, based on the work by Zhang et al. [2006].

A.3 Rhythm grammars

Rhythm grammars allow to build the description of the ECG in the highest abstraction level, as a sequence of consecutive rhythm patterns that fully explain any temporal interval. As explained in the description of the rhythm observables in section A.1.3, all grammars in this level are left recursive. Recursion is broken with the rhythm start pattern, described below. To make easier to understand the grammars in this section, the equivalent finite automaton representation is included, and the specific description of all the constraints is avoided, giving a qualitative summary of the meaning of each constraint.

A.3.1 Rhythm start grammar

$$H \rightarrow P\ wave@A\ \{tc_1\} \mid QRS@D\ \{tc_3,tc_4\} \mid QRS@\ \{tc_3,tc_4,tc_5\}$$

$$A \rightarrow QRS@B\ \{tc_2,tc_4\}$$

$$B \rightarrow T\ wave@\ \{tc_6,tc_7\}$$

$$D \rightarrow T\ wave@\ \{tc_7\}$$

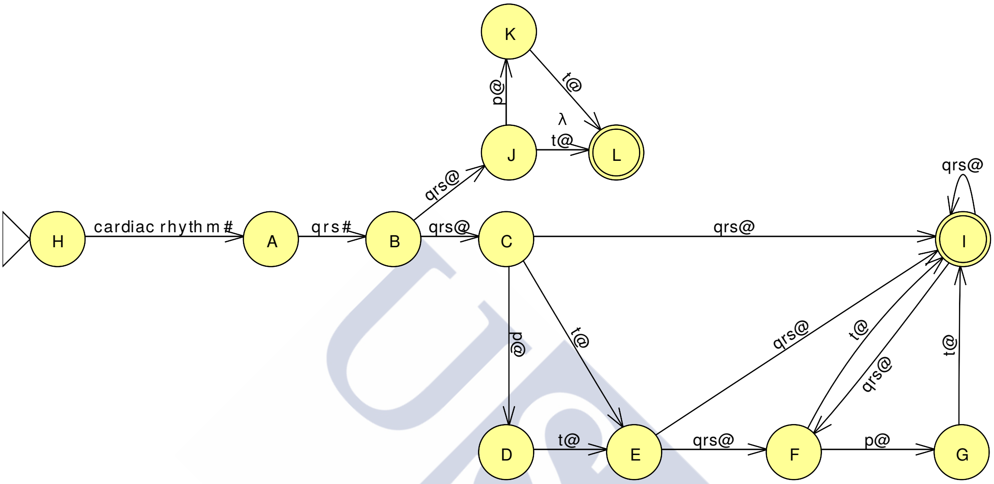
$$\begin{aligned}
tc_1 &= \{\tau c(T^b_{Pw},T^e_{Pw}) = [40,250], \quad \tau c(T^b_{Pw},T^b_H) = [0,0]\}\\
tc_2 &= \{\tau c(T^b_{Pw},T^b_{QRS}) = [100,300],\ \tau c(T^e_{Pw},T^b_{QRS}) = [1,\infty)\}\\
tc_3 &= \{\tau c(T^b_H,T^b_{QRS}) = [0,0]\}\\
tc_4 &= \{\tau c(T^b_{QRS},T^e_{QRS}) = [15,400]\}\\
tc_5 &= \{\tau c(T^e_H,T^e_{QRS}) = [0,0]\}\\
tc_6 &= \{\tau c(T^e_{Pw},T^b_{Tw}) = [0,600]\}\\
tc_7 &= \{\tau c(T^e_{QRS},T^b_{Tw}) = [0,250], \quad \tau c(T^b_{QRS},T^e_{Tw}) = [0,520],\\
&\quad\ \tau c(T^b_{Tw},T^e_{Tw}) = [1,\infty), \qquad \tau c(T^e_{Tw},T^e_H) = [0,0]\}
\end{aligned}$$

This grammar supports the generation of three abstraction patterns: The first one assembles the common observables of a normal heartbeat, i.e. a *P wave*, a subsequent *QRS complex*, and a final *T wave*, completing the full atrial and ventricular activation/recovery cycle. The second pattern avoids the observation of a *P wave*, which may occur if the first heartbeat has an activation focus outside the sinoatrial node. Finally, the third pattern only requires the observation of a *QRS complex*, which might be useful if the quality of the signal makes it difficult to properly observe the subsequent *T wave*.

In the set of constraints, tc_1 represents the *P wave* duration limits, and sets the beginning of the heartbeat as the beginning of the *P wave*; tc_2 represents the PR interval limits; tc_3 sets the beginning of the heartbeat as the beginning of the *QRS complex* in absence of a preceding *P wave*; tc_4 represents the *QRS complex* duration constraints; tc_5 sets the end of the heartbeat as the end of the *QRS complex* in absence of a posterior *T wave*; tc_6 limits the combined duration of the PR segment, the *QRS complex*, and the ST segment; tc_7 sets the limits of the ST segment and the QT interval, and sets the end of the heartbeat as the end of the abstracted *T wave*.

With the structure defined for all the remaining rhythm grammars, the first *Cardiac Rhythm* observation in any interpretation is always a *Rhythm Start* observation, abstracting the first heartbeat and supporting the generation of other rhythm patterns.

A.3.2 Regular rhythm grammars (Bradycardia, Tachycardia and Sinus Rhythm)



$$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1, vc_1\}$$

$$A \rightarrow QRS\#B\ \{tc_2\}$$

$$B \rightarrow QRS@C\ \{tc_2\} \mid QRS@J\ \{tc_2, vc_2\}$$

$$C \rightarrow P\ wave@D\ \{tc_3\} \mid T\ wave@E\ \{tc_4\} \mid QRS@I\ \{tc_2, vc_3\}$$

$$D \rightarrow T\ wave@E\ \{tc_4\}$$

$$E \rightarrow QRS@F\ \{tc_2\} \mid QRS@I\ \{tc_2, vc_3\}$$

$$F \rightarrow P\ wave@G\ \{tc_3\} \mid T\ wave@I\ \{tc_4, vc_3\}$$

$$G \rightarrow T\ wave@I\ \{tc_4, vc_3\}$$

$$I \rightarrow QRS@F\ \{tc_2\} \mid QRS@I\ \{tc_2, vc_3\} \mid \lambda$$

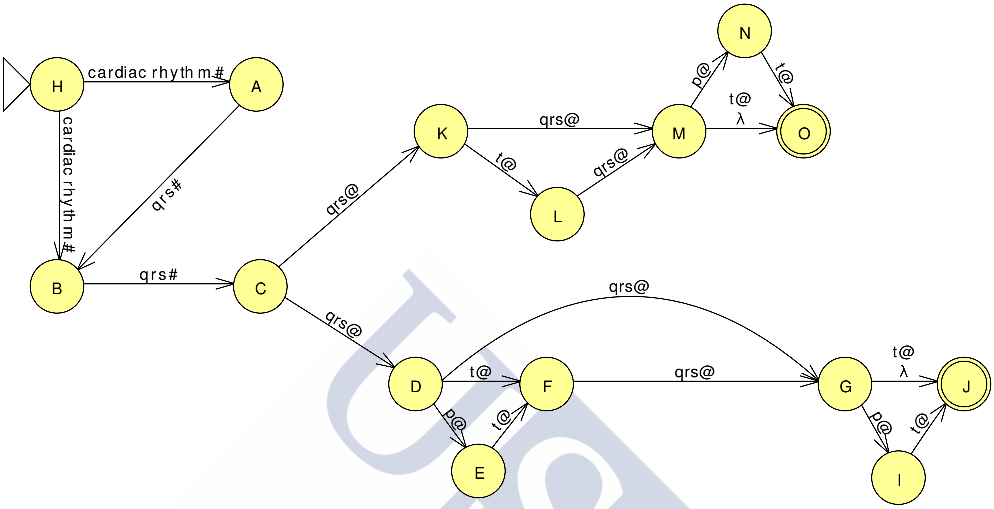
$$J \rightarrow P\ wave@K\ \{tc_3\} \mid T\ wave@\ \{tc_4\} \mid \lambda$$

$$K \rightarrow T\ wave@\ \{tc_4\}$$

Grammars enabling the conjecture of the observables *Sinus Rhythm*, *Bradycardia* and *Tachycardia* have the same form, differing only in the intervals of the temporal constraints between consecutive *QRS* observations.

The grammar for these patterns starts, as every rhythm grammar hereafter, with two environment observations: the preceding *Cardiac Rhythm* and the *QRS complex* that points out the end of the preceding rhythm and the beginning of the current one. Then, there are two possible ways to accept a sequence of heartbeats. The main one, from B to I defines a sequence of *QRS* observations, each one that may be surrounded by its corresponding *P wave* and *T wave*. The minimum number of abstracted *QRS* observations is two, and the maximum is unbounded. The other way, from state B to L, enables the conjecture of a regular rhythm with a single abstracted *QRS* observation and the preceding environment *QRS*, but imposes more constraints that in the main way, related with the properties of the preceding rhythm and specified in vc_2. Time intervals between consecutive *QRS* observations are $[1000, 2000]$ms for *Bradycardias*, $[475, 1200]$ms for *Sinus Rhythms* and $[200, 600]$ms for *Tachycardias*. The overlap between intervals makes the hypothesis more robust in case a rhythm is in the frontier of two classes. Temporal constraint tc_1 sets the beginning of the hypothesis as the ending of the preceding rhythm, while vc_1 checks that the hypothesized rhythm is different from the preceding one (for example, the concatenation of two consecutive *Bradycardias* makes no sense). The constraint tc_2 sets the allowed temporal distance of a *QRS complex* with respect to the previous one, and tc_3 and tc_4 define the temporal constraints of a *P wave* and a *T wave* with the corresponding *QRS complex*. Value constraints vc_3 check some properties of regular rhythms beyond the temporal distance between consecutive beats, like for example the rhythm variability in the full explained interval.

A.3.3 Extrasystole grammar



$$H \rightarrow \mathit{Cardiac\ Rhythm\#A}\ \{tc_1, vc_1\} \mid \mathit{Cardiac\ Rhythm\#B}\ \{tc_1, vc_2\}$$

$$A \rightarrow \mathit{QRS\#B}\ \{tc_2\}$$

$$B \rightarrow \mathit{QRS\#C}\ \{tc_3\}$$

$$C \rightarrow \mathit{QRS@D}\ \{tc_4, vc_3\} \mid \mathit{QRS@K}\ \{tc_5, vc_4\}$$

$$D \rightarrow \mathit{P\ Wave@E}\ \{tc_6\} \mid \mathit{T\ Wave@F}\ \{tc_7\} \mid \mathit{QRS@G}\ \{tc_8, vc_5\} \mid \mathit{QRS@}\{tc_8, vc_5\}$$

$$E \rightarrow \mathit{T\ Wave@F}\ \{tc_7\}$$

$$F \rightarrow \mathit{QRS@G}\ \{tc_8, vc_5\} \mid \mathit{QRS@}\ \{tc_8, vc_5\}$$

$$G \rightarrow \mathit{P\ Wave@I}\ \{tc_6\} \mid \mathit{T\ Wave@J}\ \{tc_7\}$$

$$I \rightarrow \mathit{T\ Wave@}\ \{tc_7, vc_6\}$$

$$K \rightarrow \mathit{T\ Wave@L}\ \{tc_7, vc_6\} \mid \mathit{QRS@M}\ \{tc_9, vc_7\} \mid \mathit{QRS@}\ \{tc_9, vc_7\}$$

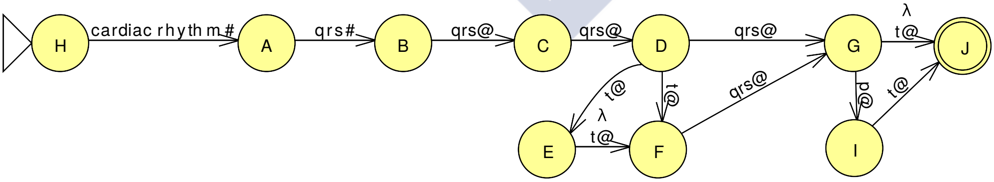
$$L \rightarrow \mathit{QRS@M}\ \{tc_9, vc_7\} \mid \mathit{QRS@}\ \{tc_9, vc_7\}$$

$$M \rightarrow \mathit{P\ Wave@N}\ \{tc_6\} \mid \mathit{T\ Wave@}\ \{tc_7, vc_6\}$$

$$N \rightarrow \mathit{T\ Wave@}\ \{tc_7, vc_6\}$$

This grammar generates a finite number of patterns abstracting two complete cardiac cycles, but considering different situations. An extrasystole is an ectopic beat with an origin outside the sinus node, and as a consequence is advanced with respect to the expected timing. This advancement is measured with the environment observations, supporting two options: The first one, following the sequence $H \to A \to B \to C$, obtains the reference heart rate as the distance between the two preceding *QRS complexes*, and it is selected if the preceding *Cardiac Rhythm* is regular. This condition is checked in the vc_1 constraints. The second option, following the sequence $H \to B \to C$, obtains the reference rate as the value of the `RR` attribute of the preceding *Cardiac Rhythm*, and is selected if the preceding rhythm is not regular, which is checked in the vc_2 constraints. Temporal constraints tc_1 and tc_3 set the end time of the preceding *Cardiac Rhythm* and the time of the last *QRS complex* to be equal to the beginning of the *Extrasystole*, while tc_2 check the temporal distance between the last and the second last *QRS complexes* in the environment, from which the reference rate is taken. From this point, basically two different patterns are recognized, with minor variations regarding the presence/absence of *P* and *T waves*. The first pattern, from C to J, recognizes classical extrasystoles with a compensatory pause after the ectopic beat, so the duration of the extrasystole is approximately two times the reference rate. On the other hand, the second pattern, from C to O recognizes rapid extrasystoles without a compensatory pause, so the duration of the *Extrasystole* is similar to the reference rate, and it is assumed the ectopic *QRS complex* does not have a preceding *P wave*. As in the regular rhythm grammars, the *P* and *T waves* are observed after the corresponding *QRS complex*. For example, the *P wave* in the $D \to E$ transition corresponds to the *QRS complex* observed in the $C \to D$ transition.

A.3.4 Couplet grammar



$$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1\}$$

$$A \rightarrow QRS\#B\ \{tc_2\}$$

$$B \rightarrow QRS@C\ \{tc_3, vc_1\}$$

$$C \rightarrow QRS@D\ \{tc_4\}$$

$$D \rightarrow T\,wave@E\ \{tc_5\} \mid T\,wave@F\ \{tc_5\} \mid T\,wave@F\ \{tc_6\} \mid$$

$$QRS@G\ \{tc_7, vc_2\} \mid QRS@\ \{tc_7, vc_2\}$$

$$E \rightarrow T\,wave@F\ \{tc_6\}$$

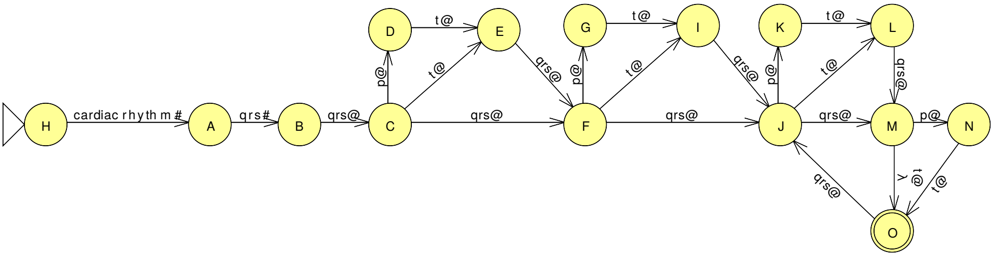
$$F \rightarrow QRS@G\ \{tc_7, vc_2\} \mid QRS@\ \{tc_7, vc_2\}$$

$$G \rightarrow P\,wave@I\ \{tc_8\} \mid T\,wave@\ \{tc_6\}$$

$$I \rightarrow T\,wave@\ \{tc_6\}$$

This grammar also generates a finite number of patterns modeling a *Couplet*, defined as the occurrence of two consecutive extrasystoles with a subsequent return to the preceding rhythm. The pattern defines as an environment evidence the previous *Cardiac Rhythm* and a *QRS complex*, which are required by tc_1 and tc_2 to finish at the beginning of the *Couplet*. Then, three additional *QRS complex* observations are required, with the first two corresponding to the ectopic beats and the third being the return to the preceding rhythm. A *P wave* is assumed to be absent in the two ectopic beats (this is not necessarily true, but strongly likely), and the *T wave* corresponding to the ectopic beats are observed after the two *QRS complexes* have been already observed (*D* to *F* transitions). This is due to the usual short distance between the second and third *QRS* in this pattern, that is overlapped with the normal distance between a *QRS* and its associated *T wave*.

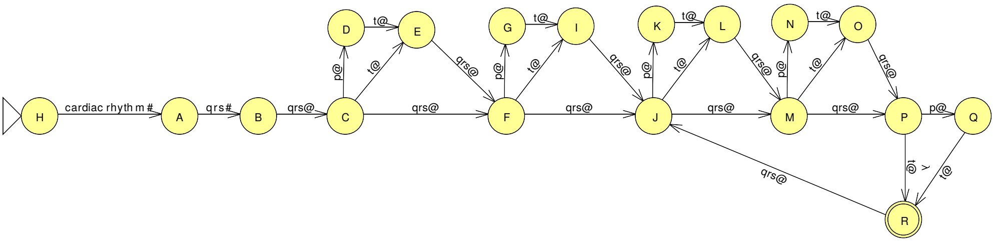
A.3.5 Bigeminy grammar



$$\begin{aligned}
H &\rightarrow \mathit{Cardiac\ Rhythm\#A}\ \{tc_1, vc_1\}\\
A &\rightarrow \mathit{QRS\#B}\ \{tc_2\}\\
B &\rightarrow \mathit{QRS@C}\ \{tc_3, vc_3\}\\
C &\rightarrow \mathit{P\ wave@D}\ \{tc_4\} \mid \mathit{T\ wave@E}\ \{tc_5\} \mid \mathit{QRS@F}\ \{tc_2\}\\
D &\rightarrow \mathit{T\ wave@E}\ \{tc_5\}\\
E &\rightarrow \mathit{QRS@F}\ \{tc_2, vc_2\}\\
F &\rightarrow \mathit{P\ wave@G}\ \{tc_4\} \mid \mathit{T\ wave@I}\ \{tc_5\} \mid \mathit{QRS@J}\ \{tc_3\}\\
G &\rightarrow \mathit{T\ wave@I}\ \{tc_5\}\\
I &\rightarrow \mathit{QRS@J}\ \{tc_3\}\\
J &\rightarrow \mathit{P\ wave@K}\ \{tc_4\} \mid \mathit{T\ wave@L}\ \{tc_5\} \mid \mathit{QRS@M}\ \{tc_2, vc_2\} \mid\\
&\quad\ \mathit{QRS@O}\ \{tc_2, vc_2\} \mid \mathit{QRS@}\ \{tc_2, vc_2\}\\
K &\rightarrow \mathit{T\ wave@L}\ \{tc_5\}\\
L &\rightarrow \mathit{QRS@M}\ \{tc_2, vc_2\} \mid \mathit{QRS@O}\ \{tc_2, vc_2\} \mid \mathit{QRS@}\ \{tc_2, vc_2\}\\
M &\rightarrow \mathit{P\ wave@N}\ \{tc_4\} \mid \mathit{T\ wave@O}\ \{tc_5\} \mid \mathit{T\ wave@}\ \{tc_5\}\\
N &\rightarrow \mathit{T\ wave@O}\ \{tc_5\} \mid \mathit{T\ wave@}\ \{tc_5\}\\
O &\rightarrow \mathit{QRS@J}\ \{tc_3\}
\end{aligned}$$

This grammar generates an infinite number of patterns consistent with the description of a *Bigeminy*, a heart arrhythmia in which every normal heartbeat is followed by an ectopic heartbeat. As usual, the first two findings are a *Cardiac Rhythm* and a *QRS complex* setting the beginning of the pattern. Then, an even number of *QRS complexes* with their corresponding *P* and *T waves* can be recognized, with constraints tc_2 forcing the *QRS complex* to keep a regular rhythm, and tc_3 checking the *QRS complex* to be advanced with respect to the regular rhythm. tc_4 and tc_5 define the constraints for *P* and *T waves*, respectively. Value constraints vc_2 check the mean and the standard deviation of the `RR` interval in the full duration of the *Bigeminy*, and vc_3 check the first ectopic beat to be significantly advanced with respect to the mean heart rate of the preceding *Cardiac Rhythm*.

A.3.6 Trigeminy grammar



$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1\}$

$A \rightarrow QRS\#B\ \{tc_2\}$

$B \rightarrow QRS@C\ \{tc_3, vc_1\}$

$C \rightarrow P\ wave@D\ \{tc_4\} \mid T\ wave@E\ \{tc_5\} \mid QRS@F\ \{tc_6, vc_2\}$

$D \rightarrow T\ wave@E\ \{tc_5\}$

$E \rightarrow QRS@F\ \{tc_6, vc_2\}$

$F \rightarrow P\ wave@G\ \{tc_4\} \mid T\ wave@I\ \{tc_5\} \mid QRS@J\ \{tc_7\}$

$G \rightarrow T\ wave@I\ \{tc_5\}$

$I \rightarrow QRS@J\ \{tc_7\}$

$J \rightarrow P\ wave@K\ \{tc_4\} \mid T\ wave@L\ \{tc_5\} \mid QRS@M\ \{tc_3\}$

$K \rightarrow T\ wave@L\ \{tc_5\}$

$L \rightarrow QRS@M\ \{tc_3\}$

$M \rightarrow P\ wave@N\ \{tc_4\} \mid T\ wave@O\ \{tc_5\} \mid QRS@P\ \{tc_6, vc_2\} \mid$
$QRS@R\ \{tc_6, vc_2\} \mid QRS@\ \{tc_6, vc_2\}$

$N \rightarrow T\ wave@O\ \{tc_5\}$

$O \rightarrow QRS@P\ \{tc_6, vc_2\} \mid QRS@R\ \{tc_6, vc_2\} \mid QRS@\ \{tc_6, vc_2\}$

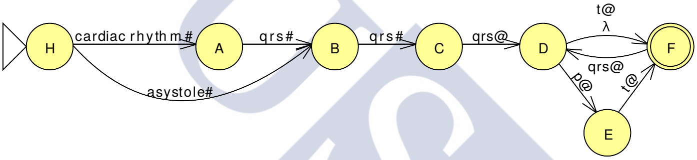
$P \rightarrow P\ wave@Q\ \{tc_4\} \mid T\ wave@R\ \{tc_5\} \mid T\ wave@\ \{tc_5\}$

$Q \rightarrow T\ wave@R\ \{tc_5\} \mid T\ wave@\ \{tc_5\}$

$R \rightarrow QRS@J\ \{tc_7\}$

This grammar is very similar to the previous one, with the only difference that the *QRS complex* observations are organized in groups with three complexes, instead of pairs as in the previous case. If we represent by **N** the regular cardiac cycles and by **V** the ectopic cardiac cycles, the equivalent regular expression generated by this grammar is **NVNN(VNN)*VN**. Temporal constraints between consecutive *QRS* are different based on the position inside each triple. tc_3 represent the constraints for ectopic *QRS complexes*, tc_6 represent the constraints for regular *QRS complexes* after an ectopic beat, defining the compensatory pause, and tc_7 represent the constraints for regular *QRS complexes* before the next ectopic beat.

A.3.7 Rhythm block grammar



$$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1, vc_1\} \mid Asystole\#B\ \{tc_1, vc_2\}$$

$$A \rightarrow QRS\#B\ \{tc_3\}$$

$$B \rightarrow QRS\#C\ \{tc_4\}$$

$$C \rightarrow QRS@D\ \{tc_5, vc_3\} \mid QRS@F\ \{tc_5, vc_3\} \mid QRS@\ \{tc_5, vc_3\}$$

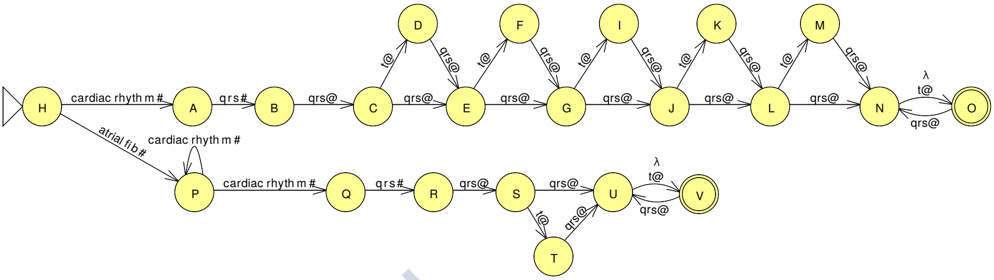
$$D \rightarrow P\ wave@E\ \{tc_6\} \mid T\ wave@F\ \{tc_7\}$$

$$E \rightarrow T\ wave@F\ \{tc_7\}$$

$$F \rightarrow QRS@D\ \{tc_8, vc_3\} \mid QRS@F\ \{tc_8, vc_3\} \mid QRS@\ \{tc_8, vc_3\}$$

This grammar allows the abstraction of *QRS* observations (and optionally the associated P and T waves) in which the interval between consecutive *QRS complexes* suffers an instantaneous elongation not compatible with the variability of a regular rhythm. Like the *Extrasystole* grammar, the delay is in general measured with respect to the two previous *QRS complexes*, considered as environment observation. However, if the previous rhythm is an *Asystole*, the reference heart rate is calculated from the `RR` attribute.

A.3.8 Atrial fibrillation grammar



$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1, vc_1\} \mid Atrial\ Fibrillation\#P\ \{tc_2, vc_2\}$

$A \rightarrow QRS\#B\ \{tc_3\}$

$B \rightarrow QRS@C\ \{tc_4, vc_3\}$

$C \rightarrow T\ wave@D\ \{tc_5\} \mid QRS@E\ \{tc_4, vc_3\}$

$D \rightarrow QRS@E\ \{tc_4, vc_3\}$

$E \rightarrow T\ wave@F\ \{tc_5\} \mid QRS@G\ \{tc_4, vc_3\}$

$F \rightarrow QRS@G\ \{tc_4, vc_3\}$

$G \rightarrow T\ wave@I\ \{tc_5\} \mid QRS@J\ \{tc_4, vc_3\}$

$I \rightarrow QRS@J\ \{tc_4, vc_3\}$

$J \rightarrow T\ wave@K\ \{tc_5\} \mid QRS@L\ \{tc_4, vc_3\}$

$K \rightarrow QRS@L\ \{tc_4, vc_3\}$

$L \rightarrow T\ wave@M\ \{tc_5\} \mid QRS@N\ \{tc_4, vc_3\} \mid QRS@O\ \{tc_4, vc_3\} \mid QRS@\ \{tc_4, vc_3\}$

$M \rightarrow QRS@N\ \{tc_4, vc_3\} \mid QRS@O\ \{tc_4, vc_3\} \mid QRS@\ \{tc_4, vc_3\}$

$N \rightarrow T\ wave@O\ \{tc_5\}$

$O \rightarrow QRS@N\ \{tc_4, vc_3\} \mid QRS@O\ \{tc_4, vc_3\} \mid QRS@\ \{tc_4, vc_3\}$

$P \rightarrow Cardiac\ Rhythm\#P\ \{tc_6, vc_4\} \mid Cardiac\ Rhythm\#Q\ \{tc_1, vc_4\}$

$Q \rightarrow QRS\#R\ \{tc_3\}$

$R \rightarrow QRS@S\ \{tc_7, vc_5\}$

$S \rightarrow T\ wave@T\ \{tc_5\} \mid QRS@U\ \{tc_7, vc_5\} \mid QRS@V\ \{tc_7, vc_5\} \mid QRS@\ \{tc_7, vc_5\}$

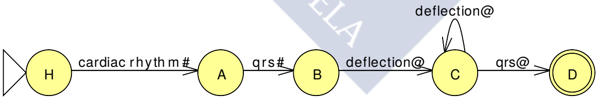
$T \rightarrow QRS@U\ \{tc_7, vc_5\} \mid QRS@V\ \{tc_7, vc_5\} \mid QRS@\ \{tc_7, vc_5\}$

$$U \rightarrow T\,wave@V\ \{tc_5\}$$

$$V \rightarrow QRS@U\ \{tc_7, vc_5\} \mid QRS@V\ \{tc_7, vc_5\} \mid QRS@\ \{tc_7, vc_5\}$$

This grammar represents the atrial fibrillation rhythm, characterized by an irregularly irregular rhythm between consecutive *QRS* observations and by the absence of a recognizable atrial activity (note the absence of *P wave* observations). The main challenge posed by this rhythm is the difficulty to support the hypothesis with a reduced number of *QRS complexes*. For this reason, the grammar defines two types of patterns: The first one, depicted at the top of the figure, requires the presence of at least seven consecutive *QRS complexes* (with the possible observation of the associated *T waves*). This minimum number is assumed to be enough to detect the characteristic rhythm irregularity. The lower path in the figure is characterized by the observation of a previous *Atrial fibrillation* environment episode finishing no longer than 10 seconds before the beginning of the hypothesized episode. Having this previous episode, the hypothesis is supported with only three *QRS complexes* showing an irregular rhythm. It is worth mentioning that even if in the *P* state there is a potentially infinite recursive rule, temporal constraints between the end of the environment *Atrial fibrillation* observation and the beginning of the hypothesis, as well as the constraints on the minimum duration of the environment *Cardiac rhythm* observations makes it impossible to have more than five environment *Cardiac rhythm* observations.

A.3.9 Ventricular fibrillation grammar



$$H \rightarrow Cardiac\ Rhythm\#A\ \{tc_1\}$$

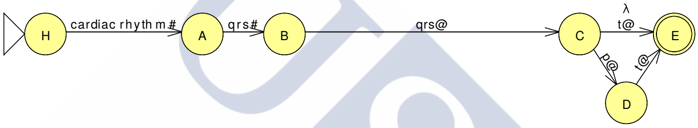
$$A \rightarrow QRS\#B\ \{tc_2\}$$

$$B \rightarrow Deflection@C\ \{tc_3, vc_1\}$$

$$C \rightarrow Deflection@C\ \{tc_4, vc_1\} \mid QRS@\ \{tc_5, vc_1\}$$

This grammar allows the abstraction of *Ventricular fibrillation* episodes, in which a rapid ventricular activity is observed without discernible QRS complexes or T waves. For this reason, the rules in the grammar do not rely on observations in the electrical activation/recovery abstraction level, but directly on *Deflection* observations. The most important piece of knowledge in this grammar is the constraint definition vc_1, which applies a state-of-the-art algorithm for the detection of ventricular fibrillation [Amann et al., 2007], checking if the full ECG fragment between the environment *QRS complex* and the last abstracted *Deflection* satisfies the criterion of the algorithm. It is worth mentioning that, even if this grammar works relatively well in practice and is correct from a knowledge description point of view, it is the only one that includes recursion before reaching a final state. This situation makes it possible to reject a hypothesis after abstracting an arbitrarily large number of observations, which may cause problems in the performance of the interpretation algorithm.

A.3.10 Asystole grammar



$$
\begin{aligned}
H &\rightarrow \mathit{Cardiac\ Rhythm\#A}\ \{tc_1\}\\
A &\rightarrow \mathit{QRS\#B}\ \{tc_2\}\\
B &\rightarrow \mathit{QRS@C}\ \{tc_3\} \mid \mathit{QRS@}\ \{tc_3, tc_4\}\\
C &\rightarrow \mathit{P\ wave@D}\ \{tc_5\} \mid \mathit{T\ wave@}\ \{tc_6, tc_7\}\\
D &\rightarrow \mathit{T\ wave@}\ \{tc_6, tc_7\}
\end{aligned}
$$

$$
\begin{aligned}
tc_1 &= \{\tau c(T^b_H, T^e_{Rhythm}) = [0,0]\}\\
tc_2 &= \{\tau c(T^b_H, T^e_{QRS}) = [0,0]\}\\
tc_3 &= \{\tau c(T^e_{QRS_1}, T^b_{QRS_2}) = [2000, 30000]\}\\
tc_4 &= \{\tau c(T^e_H, T^e_{QRS_2}) = [0,0]\}\\
tc_5 &= \{\tau c(T^b_{Pw}, T^b_{QRS}) = [100,300], \qquad \tau c(T^e_{Pw}, T^b_{QRS}) = [1,\infty)\}\\
tc_6 &= \{\tau c(T^e_{Pw}, T^b_{Tw}) = [0,600], \qquad \tau c(T^e_{QRS}, T^b_{Tw}) = [0,250],\\
&\qquad \tau c(T^b_{QRS}, T^e_{Tw}) = [0,520], \qquad \tau c(T^b_{Tw}, T^e_{Tw}) = [1,\infty)\}\\
tc_7 &= \{\tau c(T^e_{Tw}, T^e_H) = [0,0]\}
\end{aligned}
$$

This last grammar describes three possible abstraction patterns representing the absence of cardiac rhythm by simply defining a long temporal distance (between 2 and 30 seconds) between two consecutive *QRS complexes*. The three possibilities include the optional observation of the *P* and *T waves* associated to the abstracted *QRS complex*.


USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APPENDIX B

QRS DELINEATION ALGORITHM

This appendix describes a novel algorithm for QRS delineation, that is, the identification of the onset and end boundaries of an abstracted or predicted QRS complex. A proper QRS delineation makes easier the identification of heartbeats origin, and therefore empowers arrhythmia characterization in the highest abstraction level. It also may improve the specificity of QRS detection by enforcing additional constraints on the QRS waveform. This algorithm is formalized as the value constraints of the *QRS abstraction pattern*, described in Appendix A.

The QRS delineation problem has been addressed by diverse studies with quite remarkable results. For example, a number of relevant proposals are reviewed in [Ghaffari et al., 2009], showing a great precision with errors in the order of a few milliseconds, even exceeding the capacity of human experts. However, if we try to apply these methods in scenarios presenting a low signal quality compared with the reference databases, performance is greatly degraded. Indeed, the exploration of methods for biosignal processing in poorly controlled scenarios is of great interest nowadays, as it demonstrates the topic selected for recent Physionet challenges in 2014 and 2015 [Moody et al., 2014; Silva, 2015]. Thus, this approach poses as a main objective to improve the stability against high noise levels, as long as keeping the delineation error within acceptable limits.

In this appendix, section B.1 details the proposed algorithm and the intuition behind it, while section B.2 describes the evaluation of the method, including the generation of a test database and comparing the results with a well known reference algorithm.

B.1 Algorithm description

The proposed delineation algorithm, called *qrsdel*, is a multi-lead procedure decomposed in various stages depicted in figure B.1.

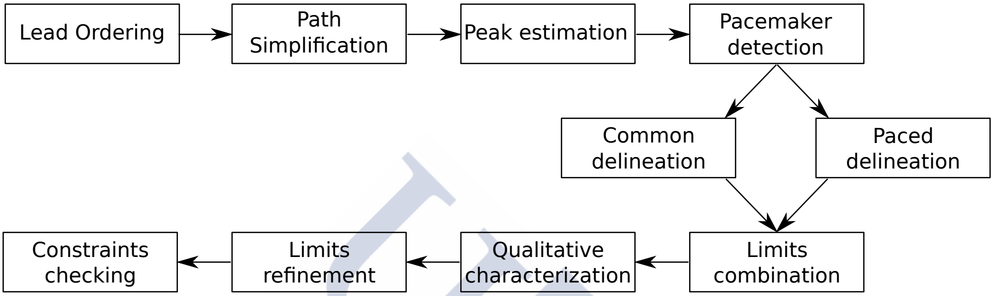


Figure B.1: Delineation algorithm steps

As a preliminary step, a temporal window is defined from a punctual beat annotation, that may be obtained by any QRS detection procedure. This window starts 80 ms before the annotation and finishes 200 ms after it, determining the signal fragment used by the subsequent steps. Note that we assume the annotation to be located near the beginning of the QRS complex, as is usual in most of the state-of-the-art QRS detection algorithms, such as those included in the WFDB toolbox [Goldberger et al., 2000].

B.1.1 Lead ordering

The first step of the algorithm is to take all signal fragments from every available ECG lead and sort them using kurtosis as a signal quality indicator [Li et al., 2008]. This ordering will determine the priority in the selection of the delineation points in later stages.

B.1.2 Path simplification

Path simplification [Douglas and Peucker, 1973] is the underpinning technique of *qrsdel*. The basic idea is to reduce the signal fragment containing the QRS complex to a small number of straight line segments, while preserving the basic waveform of the complex. The assumption this model makes is that, despite how noisy the signal is, the QRS complex is the most prominent element in the region surrounding it.

Figure B.2 illustrates this idea. We can see the original signal in light gray color, the s_{k_i} points selected by the path simplification algorithm, and the signal reconstructed using only that points. The solid and dashed segments correspond with the delineation of the QRS complex.

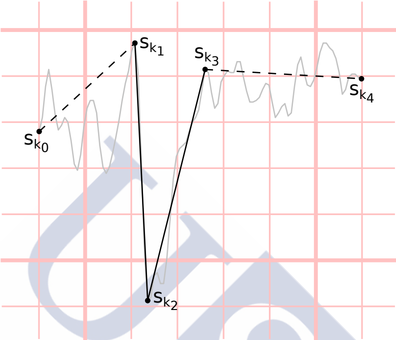


Figure B.2: QRS signal and its simplification

The algorithm we selected to perform signal simplification is a minor variation of the Douglas-Peucker method [Douglas and Peucker, 1973], considering only amplitude differences for point selection. The input for this algorithm is a signal fragment s as a sequence of n samples $(s_0, \ldots, s_{n-1})$; an integer p, with $2 \leq p \leq n$, representing the maximum number of points that can be included in the simplification; and a value $d > 0$ representing the minimum amplitude difference that can be considered to include a point in the simplification. As output, the algorithm returns a sequence $rp = (k_0, \ldots, k_{m-1})$, with $2 \leq m \leq p$, representing the indices of the samples selected for the simplification. The operation is detailed in Algorithm 9.

In this algorithm, the simplification always contains the endpoints 0 and $n-1$, and it is extended by adding at each step the point with maximum vertical distance to the segment defined by the linear interpolation (INTERP function) between any consecutive pair of points already included in the simplification. The procedure finishes when the maximum number of points is reached, or if none of the points exceeds the minimum amplitude difference.

In the *qrsdel* procedure, the DOUGLAS-PEUCKER algorithm is executed on each lead using as parameters $p = 9$ and $d = 50\mu V$. The value of p is determined by the maximum complexity of the waveform that can be recognized (see for example the *qrSRS* pattern in figure B.3), while d has been selected as the minimum peak-to-peak amplitude distance for a wave to have diagnostic implications in routine electrocardiography [CSE, 1985].

Algorithm 9 Douglas-Peucker simplification algorithm

```
function DOUGLAS-PEUCKER($s, p, d$)
    let $rp = (0, n-1)$
    let $s' =$ INTERP($s_0, s_{n-1}, n$)
    let $M, k = max(|s_i - s'_i|), i \in [0, n-1]$
    while $|rp| < p \wedge M > d$ do
        $rp = rp \cup \{k\}$
        $M = 0$
        for all $j \in [0, \ldots, |rp|-2]$ do
            $s' =$ INTERP($s_{k_j}, s_{k_{j+1}}, k_{j+1} - k_j$)
            $M', k' = max(|s_i - s'_{i-k_j}|), i \in [k_j, k_{j+1}]$
            if $M' > M$ then
                $M, k = M', k'$
            end if
        end for
    end while
    return $rp$
end function
```

B.1.3 Peak estimation

The next step is to estimate the temporal location of the peak of the QRS complex. In each lead, the peak is selected according to the following equation:

$$max_r \left(\frac{|s_r - b|}{1 + \frac{2 \cdot |r-t|}{150}} \right), r \in [0, \ldots, n-1]$$

where t is the time point of the annotation in ms, and b is the baseline value, estimated as the mode of the signal in a temporal window of 1 second around the time point of the annotation. This equation selects the point with highest deviation from the baseline, but it applies a distance function with respect to the beat annotation. Initially, the peak is selected from the best quality lead, but if the peaks in other leads are within a 40 ms margin, the earliest one is taken as reference.

B.1.4 Delineation and pacemaker detection

At this point, QRS delineation itself is performed. To this aim, the segments resulting from path simplification are clustered using a *k-means* strategy, using the following two features extracted for each segment $(s_{k_i}, s_{k_{i+1}})$:

$$f_1 = ln\left(1 + \left|\frac{(s_{k_{i+1}} - s_{k_i})}{(k_{i+1} - k_i)}\right|\right),\ f_2 = \begin{cases} r - k_i, & k_i < r \\ k_{i+1} - r, k_i \geq r \end{cases}$$

The f_1 feature represents the slope of each segment, while f_2 represents the distance to the peak estimation r. The *k-means* procedure is executed taking as centroids $c_1 = (max(f_1), min(f_2))$ and $c_2 = (min(f_1), max(f_2))$, and the limits of the QRS are set as the limits of the longest sequence of segments in the c_1 cluster, requiring the presence of at least one slope change. Figure B.2 shows a delineation example, with segments $\{(s_{k_1}, s_{k_2}), (s_{k_2}, s_{k_3})\} \in c_1$, and $\{(s_{k_0}, s_{k_1}), (s_{k_3}, s_{k_4})\} \in c_2$.

The selection of the slope as a key feature causes that the presence of pacemaker spikes, which usually show extreme slopes, strongly affects the clustering results. For this reason, a prior stage is included to determine the presence of a paced QRS complex, by looking for a signal wave satisfying the following constraints: (1) a duration smaller than 30 ms, (2) ascent and descent angles higher than 75º in the standard ECG scale, (3) an amplitude higher than 0.2mV, and (4) a rising edge of higher amplitude than the falling edge. If all these conditions are satisfied, we assume the presence of a paced beat, and the clustering results are only used to determine the end of the QRS complex, setting as its onset the beginning of the spike.

B.1.5 Limits combination and waveform characterization

At this point, the limits for the QRS complex are estimated in every available lead. The next step is to combine them, using the same strategy followed to set the peak estimation. Initially, the endpoints are considered only from the highest quality lead, and then are expanded using the endpoints in other leads only if the difference with the current estimation is lower than 40 ms. If a pacemaker spike has been detected in any lead, then the beginning of the QRS is always the beginning of the spike.

The next step consists of performing a qualitative characterization of the waveform delineated in each lead. For this, the line segments are considered as a sequence of waves, and each wave is labeled with the common names *q, Q, r, R, S*, etc. based on its position and positive/negative polarity [CSE, 1985]. Then, the longest sequence with a recognizable tag is selected for each lead, and the delineation limits are refined to fit these limits. Figure B.3 shows the full set of 27 recognized waveforms.

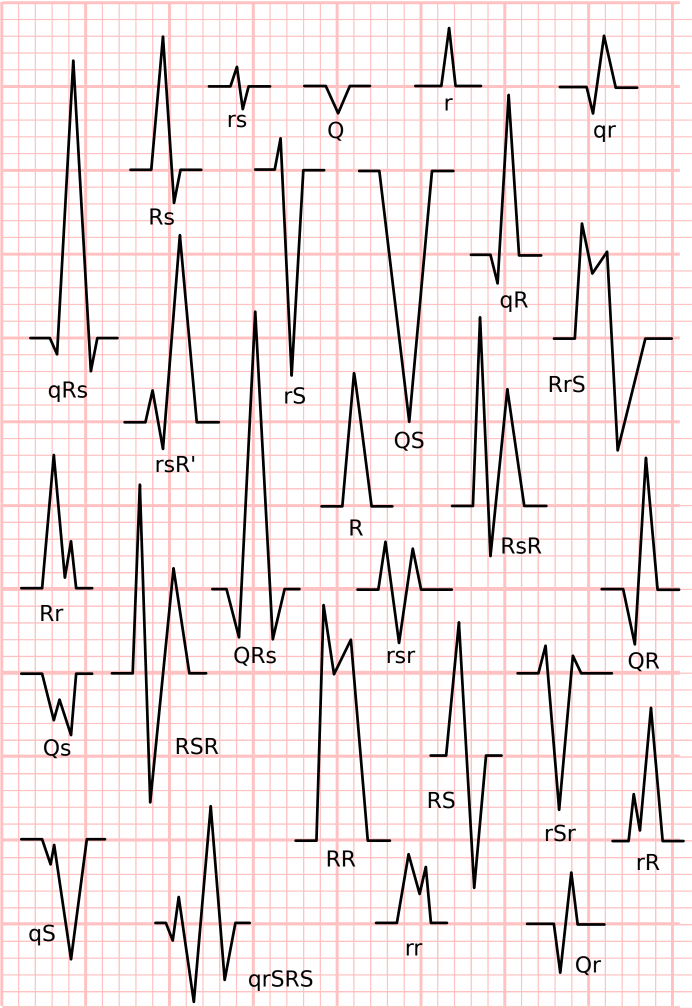


Figure B.3: Recognized QRS qualitative tags

Finally, to accept the delineation as valid, we check some common electrocardiographic constraints. Specifically, we require that at least in one lead we have identified a recognizable QRS pattern with an amplitude between 0.5 mV and 6 mV. If this requirement cannot be satisfied, then we consider sufficient to have an identified pattern in most leads. If none of this conditions can be met, then we discard the annotation, considering it as a false positive.

B.2 Evaluation

The gold standard for the evaluation of ECG delineation algorithms is the QT database [Laguna et al., 1997]. However, the set of records included in this database are more focused on the evaluation of the precision and the capability of the algorithms to deal with a wide variety of waveforms, and the signal quality is usually high. As long as the main objective is to assess the robustness of *qrsdel*, a new database was generated by adding different noise levels to the original records of the QT database, and then the behavior of *qrsdel* was compared with the results of the *ecgpuwave* application [Goldberger et al., 2000], as it is the only reference algorithm we were able to test with the new generated records.

To generate the test database the `nst` application was used [Goldberger et al., 2000], adopting the noise model recommendations from [Moody et al., 1984]. For each record in the QT database, 5 new records were created by adding electrode motion noise at different Signal-to-Noise ratios. Selected SNRs were 24 dB, 12 dB, 6 dB, 3 dB and 0 dB.

Table B.1 shows the sensitivity and measurement errors for the *qrsdel* and *ecgpuwave* algorithms at different noise levels. Since *ecgpuwave* is a single-lead algorithm, we calculated the errors using all the measurements that can be obtained in the available leads. Figure B.4 illustrates these results, allowing to appreciate the different behavior of both algorithms as noise level increases. Light-grey shaded regions represent the standard deviation of *ecgpuwave*, while dark regions represent the standard deviation of *qrsdel*. Even if *ecgpuwave* shows lower errors at reduced noise levels, it can clearly be seen that the error behavior is far more stable for the *qrsdel* algorithm, both for the mean and for the standard deviation. It is also worth noting the evolution of the sensitivity, showing a significant degradation in the case of *ecgpuwave*, while in *qrsdel* it is barely reduced from 100% even at the 0 dB SNR noise level.

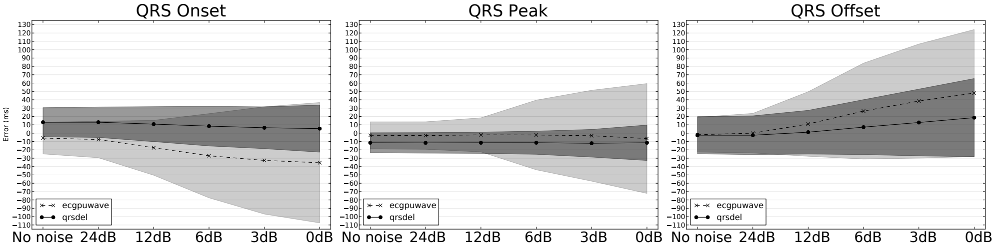


Figure B.4: Error comparison for the *qrsdel* and *ecgpuwave* algorithms

SNR		Se	QRS Onset (ms)	QRS Peak (ms)	QRS Offset (ms)
Original	*qrsdel*	100%	13.09±17.48	-11.47±12.01	-2.41±22.34
	ecgpuwave	96%	-5.87±19.03	-2.59±16.16	-1.83±20.80
24 dB	*qrsdel*	100%	13.12±18.24	-11.64±12.30	-2.47±23.21
	ecgpuwave	96%	-7.60±21.85	-2.84±16.47	-0.01±23.70
12 dB	*qrsdel*	100%	10.82±20.96	-11.48±12.75	1.17±26.16
	ecgpuwave	95%	-17.51±32.89	-2.04±20.41	10.96±38.66
6 dB	*qrsdel*	100%	8.42±23.76	-11.44±13.89	7.22±32.89
	ecgpuwave	91%	-27.05±50.44	-2.17±41.73	26.46±57.45
3 dB	*qrsdel*	100%	6.50±25.07	-12.09±16.56	12.78±39.98
	ecgpuwave	88%	-32.64±64.31	-3.05±54.44	38.45±68.36
0 dB	*qrsdel*	99%	5.47±28.19	-11.48±21.24	18.60±46.85
	ecgpuwave	85%	-35.41±72.14	-6.40±65.80	48.04±76.14

Table B.1: Sensitivity and delineation errors for the *qrsdel* and *ecgpuwave* algorithms

Regarding the accuracy of the results, while not being a primary goal of this algorithm it is necessary to ensure that the errors remain within acceptable limits. In this case, the majority of errors and standard deviations stay within 0.5 mm in the standard electrocardiographic scale (20 ms), and only for the QRS offset and with a SNR of 0 dB the standard deviation slightly exceeds the range of 1 mm (40 ms). We consider these margins admissible to provide a coarse approximation of the QRS waveform in noisy conditions.

Bibliography

[AAMI, 2007] AAMI (2007). American National Standard for Ambulatory Electrocardiographs, publication ANSI/AAMI EC38-2007.

[Ahmed et al., 1975] Ahmed, N., Milne, P. J., and Harris, S. G. (1975). Electrocardiographic data compression via orthogonal transforms. *IEEE Transactions on Biomedical Engineering*, 6(22):484–487.

[Aho et al., 2006] Aho, A., Lam, M., Sethi, R., and Ullman, J. (2006). *Compilers: Principles, Techniques and Tools*. Pearson Education, Inc.

[Alirezaie and Loutfi, 2014] Alirezaie, M. and Loutfi, A. (2014). Automated reasoning using abduction for interpretation of medical signals. *Journal of Biomedical Semantics*, 5(1):35.

[Alvarez et al., 2013] Alvarez, M., Félix, P., and Cariñena, P. (2013). Discovering metric temporal constraint networks on temporal databases. *Artificial Intelligence in Medicine*, (58):139–154.

[Amann et al., 2007] Amann, A., Tratnig, R., and Unterkofler, K. (2007). Detecting ventricular fibrillation by time-delay methods. *IEEE Transactions on Biomedical Engineering*, 54(1):174–177.

[Bardossy et al., 1995] Bardossy, A., Duckstein, L., and Bogardi, I. (1995). Fuzzy rule-based classification of atmospheric circulation patterns. *International Journal of Climatology*, 15(10):1087–1097.

[Barro et al., 1994] Barro, S., Marín, R., Mira, J., and Patón, A. (1994). A model and a language for the fuzzy representation and handling of time. *Fuzzy Sets and Systems*, 61:153–175.

[Barták et al., 2014] Barták, R., Morris, R. A., and Venable, K. B. (2014). An Introduction to Constraint-Based Temporal Reasoning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(1):1–121.

[Bortolan and Degani, 1983] Bortolan, G. and Degani, R. (1983). Comparative evaluation of three models of fuzzy decisionmaking in electrocardiography. *Proc. MEDINFO*, 83:537–540.

[Brusoni et al., 1997] Brusoni, V., Console, L., Terenziani, P., and Dupré, D. T. (1997). An efficient algorithm for temporal abduction. *AI* IA 97: Advances in Artificial Intelligence*, pages 195–206.

[Brusoni et al., 1998] Brusoni, V., Console, L., Terenziani, P., and Dupré, D. T. (1998). A spectrum of definitions for temporal model-based diagnosis. *Artificial Intelligence*, 102(1):39–79.

[Bylander et al., 1991] Bylander, T., Allemang, D., Tanner, M., and Josephson, J. (1991). The computational complexity of abduction. *Artificial intelligence*, (49):25–60.

[Callaway, 2014] Callaway, H. G. (2014). Abduction, Competing Models and the Virtues of Hypotheses. In *Model-Based Reasoning in Science and Technology. Theoretical and Cognitive Issues*, pages 263–280. Springer Berlin Heidelberg.

[Campos et al., 2010] Campos, M., Juárez, J., Palma, M., and Marín, R. (2010). Using temporal constraints for temporal abstraction. *Journal of Intelligent Information Systems*, (34):57–92.

[Castro et al., 2015] Castro, D., Félix, P., and Presedo, J. (2015). A Method for Context-Based Adaptive QRS Clustering in Real Time. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1660–1671.

[Chaovalitwongse et al., 2007] Chaovalitwongse, W. A., Fan, Y. J., and Sachdeo, R. C. (2007). On the Time Series K-Nearest Neighbor Classification of Abnormal Brain Activity. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 37(6):1005–1016.

[Charniak, 1989] Charniak, E. (1989). Motivation analysis, abductive unification and nonmonotonic equality. *Artificial Intelligence*, 34(3):275–295.

[Chow, 1970] Chow, C. (1970). On optimum recognition error and reject tradeoff. *IEEE Transaction on Information Theory*, 16(1):41–46.

[Clifford et al., 2006] Clifford, G., Azuaje, F., and McSharry, P. (2006). *Advanced methods and tools for ECG data analysis*. Artech House Publishing.

[Combi et al., 2010] Combi, C., Keravnou-Papailiou, E., and Shahar, Y. (2010). *Temporal information systems in medicine*. Springer Science & Business Media.

[Console et al., 1991] Console, L., Dupré, D., and Torasso, P. (1991). On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1(5):661–690.

[Console et al., 1996] Console, L., Portinale, L., and Dupré, D. T. (1996). Using compiled knowledge to guide and focus abductive diagnosis. *IEEE Transactions on Knowledge and Data Engineering*, 8(5):690–706.

[Console and Torasso, 1991] Console, L. and Torasso, P. (1991). A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 3(7):133–141.

[Corduas and Piccolo, 2008] Corduas, M. and Piccolo, D. (2008). Time series clustering and classification by the autoregressive metric. *Computational Statistics & Data Analysis*, 52(4):1860–1872.

[Cox and Pietrzykowski, 1986] Cox, P. and Pietrzykowski, T. (1986). Causes for events: their computation and application. In *Eighth Conference on Automated Abudction, LNCS 230*, pages 608–621. Springer-Verlag.

[CSE, 1985] CSE, W. P. (1985). Recommendations for measurement standards in quantitative electrocardiography. *European Heart Journal*, 6(10):815–825.

[Davis, 1984] Davis, R. (1984). Diagnostic reasoning based on structure and behavior. *Artificial Intelligence*, 24:347–410.

[de Chazal, 2014] de Chazal, P. (2014). An adapting system for heartbeat classification minimising user input. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 82–85.

[de Chazal et al., 2004] de Chazal, P., O'Dwyer, M., and Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*.

[de Chazal and Reilly, 2006] de Chazal, P. and Reilly, R. B. (2006). A Patient-Adapting Heartbeat Classifier Using ECG Morphology and Heartbeat Interval Features. *IEEE Transactions on Biomedical Engineering*, 53(12):2535–2543.

[de Kleer, 1986] de Kleer, J. (1986). An assumption-based TMS. *Artificial Intelligence*, 28:127–162.

[de Kleer et al., 1992] de Kleer, J., Mackworth, A., and Reiter, R. (1992). Characterizing diagnoses and systems. *Artificial Intelligence*, 56:197–222.

[de Kleer and Williams, 1984] de Kleer, J. and Williams, B. (1984). Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130.

[Dechter et al., 1991] Dechter, M., Meiri, J., and Pearl, J. (1991). Temporal constraint networks. *Artificial Intelligence*, 49:61–95.

[Douglas and Peucker, 1973] Douglas, D. and Peucker, T. (1973). Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *The International Journal for Geographic Information and Geovisualization*, 10(2):112–122.

[Dousson and Le Maigat, 2007] Dousson, C. and Le Maigat, P. (2007). Chronicle recognition improvement using temporal focusing and hierarchization. In *The International Joint Conference on Artificial Intelligence*, pages 324–329.

[Dubois and Prade, 1995] Dubois, D. and Prade, H. (1995). Fuzzy relation equations and causal reasoning. *Special Issue on "Equations and Relations on Ordered Structures : Mathematical Aspects and Applications" (A. Di Nola, W. Pedrycz, S. Sessa, eds.), Fuzzy Sets and Systems*, 75:119–134.

[Edelkamp and Schrödl, 2011] Edelkamp, S. and Schrödl, S. (2011). *Heuristic Search: Theory and Applications*. Morgan Kaufmann.

[Einthoven, 1893] Einthoven, W. (1893). Nieuwe methoden voor clinisch onderzoek [New methods for clinical investigation]. *Ned T Geneesk*, 29(II):263–286.

[Félix et al., 1999] Félix, P., Fraga, S., Marín, R., and Barro, S. (1999). Linguistic representation of fuzzy temporal profiles. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 7:243–256.

[Felner et al., 2003] Felner, A., Kraus, S., and Korf, R. (2003). KBFS: K-best-first search. *Annals of Mathematics and Artificial Intelligence*, 39(1-2):19–39.

[Ferrucci et al., 2012] Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., and Mueller, E. (2012). Watson: Beyond Jeopardy. *Artificial Intelligence*, 199–200:93–105.

[Flach, 1994] Flach, P. (1994). *Simply Logical. Intelligent Reasoning by Example*. Wiley Professional Computing.

[Flach, 1996] Flach, P. (1996). Abduction and induction: Syllogistic and inferential perspectives. In *Abductive and Inductive Reasoning Workshop Notes*, pages 31–35. University of Bristol.

[Flach and Kakas, 2000a] Flach, P. and Kakas, A. (2000a). *Abduction and Induction: Essays on their relation and integration*, volume 18. Kluwer Academic Publishers.

[Flach and Kakas, 2000b] Flach, P. and Kakas, A. (2000b). Abductive and inductive reasoning: background and issues. In *Abduction and Induction: Essays on their relation and integration*, pages 1–27. Springer Netherlands.

[Fumera et al., 2000] Fumera, G., Roli, F., and Giacinto, G. (2000). Reject option with multiple thresholds. *Pattern Recognition*, (33):2099–2101.

[Gamper and Nejdl, 1997] Gamper, J. and Nejdl, W. (1997). Abstract temporal diagnosis in medical domains. *Artificial Intelligence in Medicine*, 10(3):209–234.

[Genesereth, 1984] Genesereth, M. (1984). The use of design descriptions in automated diagnosis. *Artificial Intelligence*, 24:411–436.

[Geurts, 2001] Geurts, P. (2001). Pattern extraction for time series classification. *Principles of Data Mining and Knowledge Discovery*, 2168:115–127.

[Ghaffari et al., 2009] Ghaffari, A., Homaeinezhad, M., Akraminia, M., Atarod, M., and Daevaeiha, M. (2009). A robust wavelet-based multi-lead Electrocardiogram delineation algorithm. *Medical Engineering & Physics*, 31(10):1219–1227.

[Goldberger et al., 2000] Goldberger, A. L. et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):215–220.

[Hamscher, 1991] Hamscher, W. (1991). Modeling digital circuits for troubleshooting. *Artificial Intelligence*, 51(1-3):223–271.

[Hartshorn et al., 1931] Hartshorn, C. et al. (1931). *Collected papers of Charles Sanders Peirce*. Harvard University Press.

[Hastie et al., 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer New York, New York, NY, 2nd edition.

[Henson et al., 2012] Henson, C., Sheth, A., and Thirunarayan, K. (2012). Semantic Perception: Converting Sensory Observations to Abstractions. *IEEE Internet Computing*, 16(2):26–34.

[Hobbs et al., 1993] Hobbs, J., Stickel, M., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

[Hopcroft et al., 2001] Hopcroft, J., Motwani, R., and Ullman, J. (2001). *Introduction to automata theory, languages and computation*. Addison-Wesley.

[Hu et al., 1997] Hu, H., Palreddy, S., and Tompkins, W. (1997). A Patient-Adaptable ECG Beat Classifier Using a Mixture of Experts Approach. *IEEE Transactions on Biomedical Engineering*, 44(9):891–900.

[Ince et al., 2009] Ince, T., Kiranyaz, S., and Gabbouj, M. (2009). A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals. *IEEE Transactions on Biomedical Engineering*, 56(5):1415–1426.

[Josephson and Josephson, 1994] Josephson, J. and Josephson, S. (1994). *Abductive inference. Computation, philosophy, technology*. Cambridge University Press.

[Juárez et al., 2008] Juárez, J. M., Campos, M., Palma, J., and Marín, R. (2008). Computing context-dependent temporal diagnosis in complex domains. *Expert Systems with Applications*, 35(3):991–1010.

[Kampouraki et al., 2009] Kampouraki, A., Manis, G., and Nikou, C. (2009). Heartbeat Time Series Classification With Support Vector Machines. *IEEE Transactions on Information Technology in Biomedicine*, 13(4):512–518.

[Kanwisher, 1987] Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, 27(2):117–143.

[Kautz and Allen, 1986] Kautz, H. and Allen, J. (1986). Generalized Plan Recognition. In *Proc. of National Conference on Artificial Intelligence, AAAI*, pages 32–37.

[Kehagias and Petridis, 1997] Kehagias, A. and Petridis, V. (1997). Predictive Modular Neural Networks for Time Series Classification. *Neural Networks*, 10(1):31–49.

[Kiranyaz et al., 2016] Kiranyaz, S., Ince, T., and Gabbouj, M. (2016). Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675.

[Kodratoff, 1994] Kodratoff, Y. (1994). Induction and the Organization of Knowledge. *Machine Learning: A Multistrategy Approach*, 4:85–106.

[Kors and Bemmel, 1990] Kors, J. and Bemmel, J. V. (1990). Classification methods for computerized interpretation of the electrocardiogram. *Methods of Information in Medicine*.

[Laguna et al., 1994] Laguna, P., Jané, R., and Caminal, P. (1994). Automatic detection of wave boundaries in multilead ECG signals: validation with the CSE database. *Computers and Biomedical Research*, 27:45–60.

[Laguna et al., 1996] Laguna, P., Jané, R., Olmos, S., Thakor, N. V., Rix, H., and Caminal, P. (1996). Adaptive estimation of QRS complex wave features of ECG signal by the Hermite model. *Medical and Biological Engineering and Computing*, 34(1):58–68.

[Laguna et al., 1997] Laguna, P., Mark, R., Goldberg, A., and Moody, G. (1997). A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In *Computers in Cardiology*, pages 673–676.

[Lavrac et al., 1985] Lavrac, N., Bratko, I., Mozetic, I., Cercek, B., Grad, A., and Horvat, M. (1985). KARDIO-E: An expert system for electrocardiographic diagnosis of cardiac arrhythmias. *Expert Systems*, pages 46–50.

[Li et al., 2008] Li, Q., Mark, R., and Clifford, G. (2008). Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, 29(1):15–32.

[Lipton, 2004] Lipton, P. (2004). *Inference to the Best Explanation*. Routledge/Taylor and Francis Group.

[Litman and Allen, 1987] Litman, D. and Allen, J. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

[Llamedo and Martinez, 2012] Llamedo, M. and Martinez, J. P. (2012). An Automatic Patient-Adapted ECG Heartbeat Classifier Allowing Expert Assistance. *IEEE Transactions on Biomedical Engineering*, 59(8):2312–2320.

[Long, 1996] Long, W. (1996). Temporal reasoning for diagnosis in causal probabilistic knowledge base. *Artificial Intelligence in Medicine*, 8:193–215.

[Luz et al., 2016] Luz, E. J. S., Schwartz, W. R., Cámara-Chávez, G., and Menotti, D. (2016). ECG-based Heartbeat Classification for Arrhythmia Detection: A Survey. *Computer Methods and Programs in Biomedicine*.

[MacDonald and Zucchini, 1997] MacDonald, I. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*, volume 110. CRC Press.

[Magnani, 2015] Magnani, L. (2015). The eco-cognitive model of abduction: Naturalizing the logic of abduction. *Journal of Applied Logic*, 13(3):285–315.

[Magnani et al., 2010] Magnani, L., Carnielli, W., and Pizzi, C. (2010). *Model-Based Reasoning in Science and Technology. Abduction, Logic, and Computational Discovery*. Springer-Verlag Berlin Heidelberg, 1 edition.

[Moody et al., 1984] Moody, G. et al. (1984). A noise stress test for arrhythmia detectors. *Computers in cardiology*, 11:381–384.

[Moody et al., 2014] Moody, G., Moody, B., and Silva, I. (2014). Robust Detection of Heart Beats in Multimodal Data: the PhysioNet / Computing in Cardiology Challenge.

[Moody and Mark, 2001] Moody, G. B. and Mark, R. G. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50.

[Mörchen, 2003] Mörchen, F. (2003). Time series feature extraction for data mining using DWT and DFT. Technical Report no. 33, Department of Mathematics and Computer Science, University of Marburg.

[Moskovitch and Shahar, 2015] Moskovitch, R. and Shahar, Y. (2015). Classification of multivariate time series via temporal abstraction and time intervals mining. *Knowledge and Information Systems*, 45(1):35–74.

[Nauck and Kruse, 1999] Nauck, D. and Kruse, R. (1999). Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine*, 16(2):149–169.

[Oster et al., 2015] Oster, J., Behar, J., Sayadi, O., Nemati, S., Johnson, A. E. W., and Clifford, G. D. (2015). Semisupervised ECG Ventricular Beat Classification With Novelty Detection Based on Switching Kalman Filters. *IEEE Transactions on Biomedical Engineering*, 62(9):2125–2134.

[Palma et al., 2006] Palma, J., Juarez, J. M., Campos, M., and Marin, R. (2006). Fuzzy theory approach for temporal model-based diagnosis: An application to medical domains. *Artificial Intelligence in Medicine*, 38(2):197.

[Pavlovic et al., 1999] Pavlovic, V., Frey, B. J., and Huang, T. S. (1999). Time-series classification using mixed-state dynamic Bayesian networks. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, page 615.

[Pearl, 1988] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann.

[Peng and Reggia, 1990] Peng, Y. and Reggia, J. (1990). *Abductive inference models for diagnostic problem-solving*. Springer-Verlag.

[Petrenas et al., 2015] Petrenas, A., Marozas, V., and Sörnmo, L. (2015). Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Computers in biology and medicine*, 65:184–91.

[Pierce, 2002] Pierce, B. C. (2002). *Types and Programming Languages*. The MIT Press.

[Pimentel et al., 2014] Pimentel, M., Clifton, D., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, (99):215–249.

[Pipberger et al., 1960] Pipberger, H. V., Freis, E. D., Taback, L., and Mason, H. L. (1960). Preparation of Electrocardiographic Data for Analysis by Digital Electronic Computer. *Circulation*, 21(3):413–418.

[Poole, 1989] Poole, D. (1989). Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence*, (5):97–110.

[Poole, 1990] Poole, D. (1990). A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems*, 5(5):521–548.

[Poole, 2000] Poole, D. (2000). Learning, Bayesian Probability, Graphical Models, and Abduction. In *Abduction and Induction: Essays on their Relation and Integration*, pages 153–168. Springer Netherlands.

[Poole et al., 1987] Poole, D., Goebel, R., and Aleliunas, R. (1987). *The Knowledge Frontier*, chapter Theorist: a logical reasoning system for default and diagnosis, pages 331–352. Springer-Verlag.

[Reiter, 1987] Reiter, R. (1987). A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–96.

[Sacchi et al., 2015] Sacchi, L., Parimbelli, E., Panzarasa, S., Viani, N., Rizzo, E., Napolitano, C., Budasu, R. I., and Quaglini, S. (2015). Combining Decision Support System-Generated Recommendations with Interactive Guideline Visualization for Better Informed Decisions. In *Artificial Intelligence in Medicine*, pages 337–341. Springer International Publishing.

[Saxena et al., 2002] Saxena, S., Kumar, V., and Hamde, S. (2002). Feature extraction from ECG signals using wavelet transforms for disease diagnostics. *International Journal of Systems Science*, 33(13):1073–1085.

[Schölkopf et al., 2001] Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., and Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution. *Neural Computation*, 13(7):1443–1471.

[Shahar, 1996] Shahar, Y. (1996). Dynamic temporal interpretation contexts for temporal abstraction. In *Third International Workshop on Temporal Representation and Reasoning (TIME '96)*, pages 64–71.

[Shahar, 1997] Shahar, Y. (1997). A framework for knowledge-based temporal abstraction. *Artificial intelligence*, 90(1–2):79–133.

[Silva, 2015] Silva, I. (2015). Reducing False Arrhythmia Alarms in the ICU: The PhysioNet/Computing in Cardiology Challenge 2015.

[Silva et al., 2015] Silva, I., Moody, B., Behar, J., Johnson, A., Oster, J., Clifford, G., and Moody, G. B. (2015). Robust detection of heart beats in multimodal data. *Physiological Measurement*, 36(8):1629.

[Smith and Kosslyn, 2013] Smith, E. E. and Kosslyn, S. M. (2013). *Cognitive Psychology: Pearson New International Edition: Mind and Brain*. Pearson Higher Ed.

[Sörnmo and Laguna, 2005] Sörnmo, L. and Laguna, P. (2005). *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press.

[Teijeiro et al., 2014] Teijeiro, T., Félix, P., and Presedo, J. (2014). Using Temporal Abduction for Biosignal Interpretation: A Case Study on QRS Detection. In *2014 IEEE International Conference on Healthcare Informatics*, pages 334–339. IEEE.

[Teijeiro et al., 2015] Teijeiro, T., Félix, P., and Presedo, J. (2015). A noise robust QRS delineation method based on path simplification. In *2015 Computing in Cardiology Conference (CinC)*, pages 209–212.

[Teijeiro et al., 2016] Teijeiro, T., Félix, P., Presedo, J., and Castro, D. (2016). Heartbeat classification using abstract features from the abductive interpretation of the ECG. *IEEE Journal of Biomedical and Health Informatics*.

[Thagard, 2010] Thagard, P. (2010). How brains make mental models. In *Studies in Computational Intelligence*, volume 314, pages 447–461. Springer Berlin Heidelberg.

[Trahanias and Skordalakis, 1990] Trahanias, P. and Skordalakis, E. (1990). Syntactic pattern recognition of the ECG. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 12(7):648–657.

[Wagner, 2008] Wagner, G. (2008). *Marriott's Practical Electrocardiography*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 11 edition.

[Wainer and de Melo, 1997] Wainer, J. and de Melo, A. (1997). A temporal extension to the parsimonious covering theory. *Artificial Intelligence in Medicine*.

[Young and Huggins, 1963] Young, T. Y. and Huggins, W. H. (1963). On the Representation of Electrocardiograms. *IEEE Transactions on Bio-medical Electronics*, 10(3):86–95.

[Zhang et al., 2006] Zhang, Q., Illanes, A., Médigue, C., Papelier, Y., and Sorine, M. (2006). An algorithm for robust and efficient location of T-wave ends in electrocardiograms. *IEEE transactions on bio-medical engineering*, 53(12 Pt 1):2544–52.

[Zhang et al., 2014] Zhang, Z., Dong, J., Luo, X., Choi, K., and Wu, X. (2014). Heartbeat classification using disease-specific feature selection. *Computers in Biology and Medicine*, 46:79–89.

[Zong et al., 2003a] Zong, W., Heldt, T., Moody, G. B., and Mark, R. G. (2003a). An open-source algorithm to detect onset of arterial blood pressure pulses. In *Computers in Cardiology*, pages 259–262.

[Zong et al., 2003b] Zong, W., Moody, G., and Jiang, D. (2003b). A robust open-source algorithm to detect onset and duration of QRS complexes. In *Computers in Cardiology*, pages 737–740.