Technological University Dublin

ARROW@TU Dublin

Books/Book Chapters

School of Computing

2013

Argumentation Theory for Decision Support in Health-Care: a Comparison with Machine Learning

Luca Longo
Technological University Dublin, luca.longo@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/scschcombk

Part of the Computational Engineering Commons

Recommended Citation

Argumentation Theory for Decision Support in Health-Care: A Comparison with Machine Learning

Luca Longo and Lucy Hederman

Centre for Next Generation Localization
Department of Computer Science and Statistics - Trinity College Dublin
{llongo,hederman}@scss.tcd.ie

Abstract. This study investigates role of defeasible reasoning and argumentation theory for decision-support in the health-care sector. The main objective is to support clinicians with a tool for taking plausible and rational medical decisions that can be better justified and explained. The basic principles of argumentation theory are described and demonstrated in a well known health scenario: the breast cancer recurrence problem. It is shown how to translate clinical evidence in the form of arguments, how to define defeat relations among them and how to create a formal argumentation framework. Acceptability semantics are then applied over this framework to compute arguments justification status. It is demonstrated how this process can enhance clinician decision-making. A well-known dataset has been used to evaluate our argument-based approach. An encouraging 74% predictive accuracy is compared against the accuracy of well-established machine-learning classifiers that performed equally or worse than our argument-based approach. This result is extremely promising because not only demonstrates how a knowledge-base paradigm can perform as well as state-of-the-art learning-based paradigms, but also because it appears to have a better explanatory capacity and a higher degree of intuitiveness that might be appealing to clinicians.

1 Introduction

The amount of evidence produced in clinical environments has been rapidly increasing thanks to the adoption of new technologies, such as Electronic Health Records, for assisting clinicians in their daily activities. Although this shift is good for the advance of science and knowledge, it introduces difficulties for practitioners/researchers in terms of degree of efficiency and accuracy in assimilating, acquiring and aggregating clinical evidence. In the *health-care* sector, knowledge and new evidence are often heterogeneous and complex, inconsistent and incomplete. These factors play an important role in many clinical decision-making processes, most of the time made under conditions of uncertainty and with partial information. Current *clinical decision support systems* have become more complex because plausible conclusions need to be extracted from a set of heterogeneous pieces of evidence, sometimes contradictory, and from different points of view and interpretations. They are mainly based on case-base or probability-based reasoning, and they adopt techniques borrowed from Artificial Intelligence such as machine learning or Fuzzy Logic. However, the majority of them require well structured evidence, not partial and are based on learning from previous data or cases. In addition, the amount of evidence required for the learning process must be high in order to

significantly infer recommendations for clinical decisions. These systems manipulate knowledge and evidence in a numerical, usually complex way, not using familiar terms, thus being not attractive to clinicians. Health-care practitioners tend to follow a *defeasible reasoning* process for taking plausible decisions. Defeasible reasoning is a kind of analysis and interpretation that is based on reasons that are defeasible: a conclusion can be retracted in the light of new evidence. Indeed decisions are based on evidence-based knowledge, but the aggregation of pieces of evidence tends to be close to the way humans reason. This kind of reasoning process can be formalised using *Argumentation Theory*, an emerging paradigm, based on arguments, aimed at investigating their consistency and reducing uncertainty. According to the limitations of current state-of-the-art approaches, clinicians and health practitioners, in general, prefer decision-making support systems that deliver more explanations than numerical aids. In other words, they would adopt qualitative systems rather than quantitative tools. Indeed, numerical outcomes are more accurate than linguistic outcomes, but most of the time they are difficult to interpret. Furthermore, the inference process that leads to a recommendation, can be hard to be understood by clinical experts, in the case it is only based on a numerical manipulation of evidence and knowledge.

This study follows another recent study [8] and it is aimed at investigating the *role of defeasible reasoning and argumentation theory for supporting decision-making processes under uncertainty in the health-care sector*. The objective is to support clinicians with a tool for taking plausible and rational medical decisions that can be better justified and explained. The remainder of this paper is organised as follows. Section 2 introduces *Argumentation theory* and the building blocks of abstract argumentation. This theory is extended in section 3 by incorporating the notion of degree of truth. This new argument-based approach is applied and evaluated in 4, comparing its predictive accuracy against the accuracy of few machine-learning classifier, using a well-known Breast Cancer dataset. Section 5 discusses findings, emphasising advantages and drawbacks of the new approach. Related similar work in the health-care sector follows. Section 6 highlights the contribution of this study individuating potential areas of improvements and future work.

2   Argumentation Theory

Argumentation theory has evolved from its original primary context as a sub-discipline in philosophical logic, to emerge, in the last decade, as an important area of logic-based AI [11]. The theory gained importance with the introduction of formal models, inspired by human-like reasoning. These extended classical reasoning models based on deductive logic that appeared increasingly inadequate for problems requiring non-monotonic reasoning [1], commonly used by humans, and explanatory reasoning, not available in standard non-monotonic logics such as default logic. In non-monotonic reasoning a conclusion can be retracted in the light of new evidence whereas in deductive reasoning the set of conclusions always grows. The modularity and intuitiveness of argumentation lends to explanatory reasoning, avoiding the monolithic approach of many traditional non-monotonic reasoning logics. The reasoning required in many practical decision-making processes, especially in health-care, is both non-monotonic and explanatory.

Argumentation theory and in particular abstract argumentation frameworks have been proved to be fruitful in modelling and studying defeasible reasoning processes [1] [7] [12] [14] [10]. The argumentation process starts by formalising arguments considering a knowledge base. The second step requires the explication of defeat relations among arguments. Arguments and defeat relations are then organised in an argumentation framework for evaluation. The subsequent step is the application of acceptability semantics for computing arguments' justification status and rationally accepting some of them in an *extension* (set of arguments) [6]. Eventually, aggregation of these arguments is needed is a rational decision has to be taken. This step is the last of a chain of inference steps for non-monotonic entailment. The underlying idea behind argumentation theory is that, given a set of arguments, where some of them defeat (attack) others, a decision is to be taken to determine which arguments can ultimately be accepted.
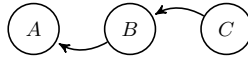
Fig. 1. Argument reinstatement

Merely looking at an argument's defeaters to determine the acceptability status of an argument is not enough: it is also important to determine whether the defeaters are defeated themselves. We say that an argument B *defeats* argument A iff B is a reason against A. If the internal structure of arguments and the reasons why they defeat each other, are not considered, what remains is called an *argumentation framework* (AF) [6]. An *abstract argumentation framework* AAF is a pair $< Arg, attacks >$ where:

- Arg is a finite set of element referred to as (abstract) *arguments*, and
- $attacks \subseteq Arg \times Arg$ is binary relation over Arg

Given sets $X, Y \subseteq Arg$ of arguments, X *attacks* Y if and only if there exists $x \in X$ and $y \in Y$ such that $(x, y) \in attacks$.

The question is which arguments should ultimately be accepted. In fig. 1, A is defeated by B, and apparently A should not be accepted since it has a counterargument. However, B is itself defeated by C that is not defeated by anything, thus C should be accepted. But if C is accepted, then B is ultimately rejected and does not form a reason against A anymore. Therefore A should be accepted as well. In this situation we say that C *reinstates* A. Due to this issue of reinstatement, a formal criterion that determines which of the arguments of an AAF can be accepted, is needed. In the literature, this criterion is known as *semantic*: given an AAF, it specifies zero or more sets of acceptable arguments, called *extensions*. Various argument-based semantics have been proposed but here we focus on the following semantics [6]: a set $X \subseteq Arg$ of argument is

- *admissible* if and only if X does not attack itself and X attacks every set of arguments Y such that Y attacks X;
- *complete* if and only if X is admissible and X contains all arguments it *defends*, where X defends x if and only if X attacks all attacks agains x;
- *grounded* if and only if X is minimally complete (with respect to $\subseteq$);
- *preferred* if and only if X is maximally admissible (with respect to $\subseteq$);

Example 1. *In the AF of fig. 1 there is just one complete extension, $\{A, C\}$, which is conflict-free and defends exactly itself. It can be seen as a subjective but internally coherent point of view: someone can disagree, but can not point out an internal inconsistency [13]. The grounded extension is $\{A, C\}$. The admissible sets are $\{C\}$, $\{A, C\}$. $\{B\}$ and $\{A\}$ are not admissible as they do not defend themselves respectively against C and B. One preferred extension exits: $\{A, C\}$.*

3 Design of a Defeasible Model

Abstract argumentation theory ([6]) can be extended for supporting decision-making processes in practical scenarios. The following formal definitions are clarified by illustrative examples concerning a health-care problem: breast cancer recurrence prediction. Predicting recurrence is important for assisting the identification of patients with critical prognosis and minimising unnecessary therapies. We have chosen this domain because real data from a public dataset [1] is available and repeatedly used in the machine learning literature from 1986 up to 2011 ([12] [4]). It includes 286 instances of real patients who went through a breast cancer operation (9 records contains incomplete values). Each instance is described by 9 possible predictive attributes and a binary outcome, 'recurrence' or 'no recurrence' (table 1 column 1). The dataset consists of attributes with value ranges as in table 1 (column 2). For 81 patients, the illness reappeared after 5 years while 196 did not have recurrence. In this context, the aims of our proposal are: a) to translate a knowledge-base into a set of structured defeasible arguments and defeat relationships among them; b) to run acceptability semantics for extracting sets of consistent arguments; c) to recommend a set for decision-making support. We propose to implement a) adopting membership functions and degree of truth; to execute b) using grounded and preferred semantics from abstract argumentation; to implement c) recommending the set that maximises the degree of truth of its arguments. This approach can handle two types of uncertainty: *vagueness* in defining attributes of a knowledge-base; *ambiguity* in defining arguments as defeasible inference rules. The rationale behind adopting membership functions is their usefulness for modelling vaguely defined sets and human reasoning that is approximate rather than fixed and exact. They map an attribute's value to the relative set with a degree of truth.

Definition 1 (Membership function). *For any set X, a membership function on X is any function $f : X \rightarrow [0, 1] \in \Re$. Membership functions on X represent fuzzy subsets of X. For an element x of X, the value $f(x)$ is called the membership degree of x in the fuzzy set and quantifies the grade of membership of x to the fuzzy set X. We indicate $MF_X = \{f | f : X \rightarrow [0, 1] \in \Re \}$ as the set of membership functions defined over X.*

3.1 Translating Knowledge-Bases into Arguments and Attack Relations

Informally, an argument is a defeasible rule (open to defeats) composed by a set of premises and a claim. In other words, from the premises a claim can be inferred. This process is intrinsically uncertain, because the inference rules are defeasible in nature, and not strict and totally certain. This is coherent with human reasoning that is uncertain rather than exact. In our illustrative scenario, a claim is a possible conclusion

---

[1] University Medical Center, institute of Oncology, Ljubljana M. Zwitter and M. Soklic)

Table 1. Dataset attributes and a possible vague expert's knowledge-base description

Attribute	Dataset Range	Agent's Knowledge-base - Description
Age	10-19, 20-29, .. , 90-99	The strongest risk factor for breast cancer is age: the older the woman, the higher the risk of cancer (and presumably recurrence).
Menopausal	lt40, ge40, premeno	Pre-menopausal status is a reason to believe recurrence is not likely.
Tumor size	0-4, 5-9, ... , 55-59	The greatest diameter of the excised tumor, the greater chance of recurrence.
Node involvement	0-2, 3-5, ..., 36-39	Since the axillary lymph nodes act as a primary site of drainage for the breast, they represent common site of early metastasis. The more lymph nodes involved are, the more likely recurrence is. This is probably the most influential factor for recurrence.
Node capsular invasion	yes, no	If the cancer does metastasis to a lymph node, even if outside the original tumor site, it can remain 'contained' by the lymph node's capsule. However, the tumor may replace the lymph node penetrating the capsule, invading the surrounding tissues. If capsular invasion, recurrence is more likely.
Degree of malignancy	1, 2, 3	The tumor's histological grade affects recurrence. If 1 (tumors consist of cells that, while neoplastic, retain many of their usual characteristics), recurrence is less likely. If it is 2 or 3 (tumors consists of highly abnormal cells, with marked variation in cell size, or a high index of mitotic activity in the cells) recurrence is more likely.
Breast	left, right	Although cancer can occur in either breast, there is no difference in incidence between breasts. A slightly higher (but unexplained) recurrence risk, on left side, exists.
Breast quadrant	left-up/low, right-up/low, central	The breast may be divided in 4 quadrants, using the nipple as a central point. Breast cancer often occurs in the upper outer quadrant increasing the chance of recurrence.
Irradiation	yes, no	Radiotherapy for breast cancer reduces recurrence
Outcome (class)	Recurrence (R), no-Recurrence (NR)	Reappearance of cancer after 5 years / No Reappearance of cancer after 5 years

available to a clinician (Recurrence, no-Recurrence) to support decisions. In argumentation theory, arguments might be considered *forecast* when they are in favour or against a certain claim (but justification is not infallible), and *mitigating arguments*, when they defeat (undermine justification for) forecast or other mitigation arguments [9].

Definition 2 (Argument - Forecast). *A forecast argument β is defined over a membership function f_α for attribute α and a claim c. $ARG_F : MF_{ATTR} \times C$ and "$\beta : f_\alpha \to c$" can be read as 'there is a reason to believe c from f_α' or 'c is what reasonably follows from f_α.'*

Definition 3 (Argument - Mitigating). *A mitigation argument β is defined over a membership function f_α for the attribute α and another argument δ (either forecast or mitigation). $ARG_M : MF_{ATTR} \times ARG_F \cup ARG_M$ and '$\beta : f_\alpha \to \neg\delta$' can be read as 'there is a reason to believe $\neg\delta$ from f_α' or 'the justification of δ is undermined by f_α.'*

Example 2. · *A forecast argument: $(ar : \text{'old age'} \to R')$, with 'Old' the fuzzy subset of the attribute 'age' defined with the membership function f_{Age}^{Old}, R ('recurrence') the claim that reasonably follows the premise. · A mitigating argument: $[ar : high\ tumor\ size \to \neg(low\ age \to NR)]$ with 'high' and 'low' respectively the membership functions for attributes 'tumor size and 'age', NR ('no recurrence') is the claim that follows 'low age' undermined by 'high tumor size'. The fact that a low age is a reason to believe no recurrence is undermined by the high tumor size.*

A knowledge-base might contain contradicting and inconsistent evidence. In argumentation theory, this notion is expressed by defeating relations that might be *rebuttal* or *undercutting*. The former occurs between two *forecast arguments* contradicting each other because supporting mutually exclusive claims (bi-directional). The latter occurs when a mitigating argument challenges the inference that links premises to claim of a forecast or another mitigating argument (uni-directional).

Definition 4 (Attack - rebutting). *Given two forecast arguments a, b with $a : f_\alpha \to c_1$, $b : f_\beta \to c_2$ we say that a attacks b and we indicate (a, b), iff $c_1 \neq c_2$ and c_1 and c_2 are*

mutually exclusive. $ATT_R : ARG_F \times ARG_F$. *As a rebuttal attack is symmetrical it holds that iff* (a, b) *then also* (b, a).

Definition 5 (Attack - undercutting). *A mitigating argument A attacks a forecast or another mitigating argument B, (indicated as (a,b)), if in the agent's knowledge base there is evidence suggesting that B is no longer justified because of A.* $ATT_U : ARG_M \times ARG_F \cup ARG_M$

From the translation of a knowledge-base a *contextual argumentation framework* CAF containing the pool of designed arguments (forecast and mitigation) and the pool of designed attack relations (rebutting and undercutting) emerges:

- $ARG_{pool} : \{a \mid a \in ARG_F \cup ARG_M\}$.
- $ATT_{pool} = \{(a, b) \mid (a, b) \in ATT_R \cup ATT_U \text{ and } a, b \in ARG_{pool}\}$.

We have interviewed an expert in the domain of breast cancer and the CAF that resulted from the translation of her knowledge-base includes the attributes (fuzzy sets) of table 1, the corresponding membership functions (fuzzy subset) of figure 2 and arguments of table 2 (left side). Bibliography references of designed arguments are skipped and only those attributes accounted in the Ljubljana dataset (dated 1986) were used. Indeed, nowadays, new arguments can be designed considering recent discoveries.

Example 3. *Arguments and undercutting attacks of table 2 form the following CAF:*

- $ARG_F = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p\}$
- $ARG_M = \{q, r, s, t, u, v, w, x, y, z\}$
- $ARG_{pool} = ARG_F \cup ARG_M$
- $ATT_R = \{(\alpha, \beta) \mid (\alpha, \beta) \; \forall \alpha, \beta \in ARG_F \text{ and } f_\alpha \rightarrow c_1{'}, \; `\beta : f_\beta \rightarrow c_2{'}, \; c_1 \neq c_2\}$
- $ATT_U = \{(q, d), (r, o), (s, p), (t, a), (u, a), (v, o), (w, p), (x, o), (y, p), (z, p)\}$
- $ATT_{pool} = ATT_R \cup ATT_U$
- $CAF = (ARG_{pool}, ATT_{pool})$
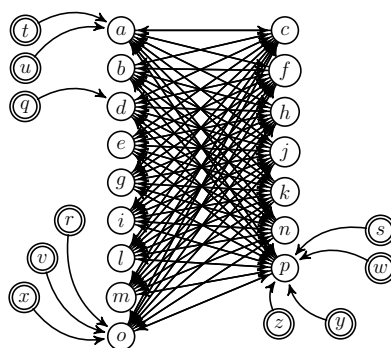
Plain white nodes are forecast arguments, with symmetrical straight black arrows indicating rebuttal attack. Double-circled nodes are mitigating arguments, with asymmetrical curvy black arrows representing undercutting attack. Arguments t, u, q, r, v, x (left side) challenge forecast arguments; s, w, y, z (right side) indicate a preference over forecast arguments.

Once a knowledge-base has been fully translated into arguments and attacks relations (CAF), the next step is to individuate which arguments and attacks are objectively activated in a given practical scenario. A forecast argument is activated if the membership function contained in its premises returns a non-zero value. Similarly, not all the mitigating arguments are activated. A mitigating argument a is activated if and only if the membership function contained in its premises returns a non-zero value *and* if and only if the argument defeated by a is in the set of the activated forecast arguments. In other words, if the mitigating attacker (premise) is not activated, or if the defeated argument is not activated (claim of the mitigating argument), the existence of the whole mitigating argument does not make sense, thus it is not activated.

Table 2. Pool of arguments and undercutting attacks for the illustrative scenario

Attribute: Age, Evid.: 1			Attribute: Menopause, Evid: 2		
Arg	MF_{Age}	**Claim**	**Arg**	$MF_{Menopause}$	**Claim**
a	low	no rec.	d	pre	no rec.
b	medium	no rec.	e	post-lt40	no rec.
c	high	rec.	f	post-gt40	rec.
Attribute: Tumor size, Evid: 3			Attribute: Node invas., Evid: 4		
Arg	MF_{TumorS}	**Claim**	**Arg**	$MF_{NodeInv}$	**Claim**
g	low	no rec.	i	low	no rec.
h	high	rec.	j	high	rec.
Attribute: Node Caps, Evid: 5			Attribute: Deg. Malig., Evid: 6		
Arg	MF_{NodeC}	**Claim**	**Arg**	$MF_{DegMalig}$	**Claim**
k	true	rec.	m	low	no rec.
l	false	no rec.	n	high	rec.
Attribute: Breast, Evid: 7			Attribute: Breast quad, Evid: 8		
Available evidence suggests that the attribute breast is not influential thus no argument is built			**Arg**	$MF_{BreastQ}$	**Claim**
			o	lower	no rec.
			p	upper	rec.

q: High Age $\rightarrow \neg$ (Menop Pre$\rightarrow$NR)

r: High Age $\rightarrow \neg$ (Lower BreastQ$\rightarrow$NR)

s: High Age $\rightarrow \neg$ (Upper BreastQ$\rightarrow$R)

t: High NodeInv $\rightarrow \neg$ (Low age$\rightarrow$NR)

u: High TumorSize $\rightarrow \neg$ (Low age$\rightarrow$NR)

v: High NodeInv $\rightarrow \neg$ (Lower BreastQ$\rightarrow$NR)

w: High NodeInv $\rightarrow \neg$ (Upper BreastQ$\rightarrow$R)

x: high Tumorsize $\rightarrow \neg$ (Lower BreastQ$\rightarrow$NR)

y: High Age $\rightarrow \neg$ (Upper BreastQ$\rightarrow$R)

z: High TumorSize $\rightarrow \neg$ (Upper BreastQ$\rightarrow$R)
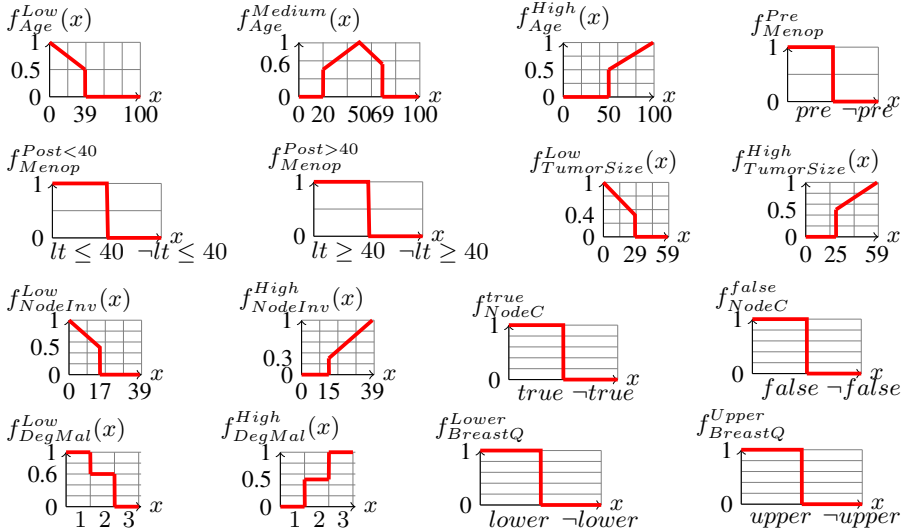
Fig. 2. Membership functions of attributes for the illustrative scenario

Definition 6 (Activated arguments). *The set of activated forecast arguments is a subset of the pool of arguments: $ARG_F^{Act} \subseteq ARG_{pool}$. For a given argument A defined over the attribute α, a membership function f_α, and a objective value α^{val}, $A \in ARG_F^{Act}$ iff $f_\alpha(\alpha^{val}) > 0$. The set of activated mitigating arguments is a subset of the pool of arguments: $ARG_M^{Act} \subseteq ARG_{pool}$. For a given mitigating argument $B : f_\alpha \rightarrow \neg\delta$, defined over the attribute α, a membership function f_α, and a objective value α^{val}, $B \in ARG_M^{Act}$ iff $\delta \in ARG_F^{Act}$ and iff $f_\alpha(\alpha^{val}) > 0$*

The same principle is applied to rebutting and undercutting attacks. A rebutting attack is activated if and only if both the attacker and the attacked are in the set of

activated forecast arguments. An undercutting attack (a, b) is activated if and only if the undercutting argument a is in the set of the activated mitigating arguments.

Definition 7 (Activated attacks). *The set of activated rebutting attacks is a subset of the pool of attacks:* $ATT_R^{ACT} \subseteq ATT_{Pool}$, $(a, b) \in ATT_R^{ACT}$ iff $a, b \in ARG_F^{Act}$. *The set of activated mitigating attacks is a subset of the pool of attacks:* $ATT_U^{ACT} \subseteq ATT_{Pool}$ *and* $(a, b) \in ATT_U^{ACT}$ iff $a \in ARG_M^{Act}$.

At this stage, the *instantiated argumentation framework* (IAF) emerges, which is a sub-CAF. $IAF = (ARG_M^{Act} \cup ARG_F^{Act}, ATT_R^{Act} \cup ATT_U^{Act})$

Example 4. *Let us consider a record of the Ljubljana dataset related to a patient as follows:* age *(40-49),* menopause *(premeno),* Tumor-size *(30-34),* Inv-nodes *(0-2),* Node-caps *(no),* Deg-malign *(2),* breast *(right),* Breast-quad *(right_low),* Irradiation *(no). For* age, Tumor-size, Inv-nodes, *we take respectively the centre of each interval, thus 44.5, 32, 1. The membership functions that return degrees of truth greater than zero are:* f_{age}^{medium}, f_{Menop}^{Pre}, $f_{TumorSize}^{High}$, $f_{NodeInv}^{Low}$, f_{NodeC}^{False}, $f_{DegMalig}^{Low}$, $f_{DegMalig}^{High}$, $f_{BreastQ}^{Lower}$. *The activated arguments and attack relationships, and the final AAF are:*

- $ARG_F^{Act} = \{b, d, h, i, l, m, n, o\}$
- $ARG_M^{Act} = \{x\}$
- ATT_R^{Act} = $\{(b,h),(h,b),(b,n),(n,b),(d,h),$ $(h,d),(d,n),(n,d),(i,h),(h,i),(i,n),(n,i),(l,h),$ $(h,l),(l,n),(n,l)$ $(m,h),(h,m),(m,n),(n,m),$ $(o,h),(h,o),$ $(o,n),$ $(n,o)\}$

- $ATT_U^{Act} = \{(x,o)\}$
- $IAF = (ARG_F^{Act} \cup ARG_M^{Act},$ $ATT_R^{Act} \cup ATT_U^{Act})$
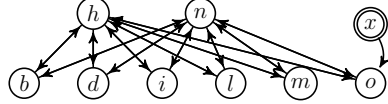
Note 1. *Arguments* m *and* n, *despite dealing with the same attribute (Degree of malignancy), are both activated but with different degrees of truth.* u *is not activated: although its premise* $f_{TumorSize}^{High}$ *is activated, its claim (the attacked argument) is not* $a : LowAge \to NR$ *as not present in* ARG_F. *Similarly,* q *is not activated because, even if the attacked argument* d *is in* ARG_F, *its premise (High Age) is not activated.*

3.2   Running Acceptability Semantics and Recommending an Extension

Abstract Argumentation semantics can now be executed on the IAF to obtain *extensions* of arguments and to decide which of them can be accepted. Grounded and the preferred semantics (as described in section 2) are applied. The former always returns one extension of arguments: if not empty, it contains all the arguments in the IAF that support the same claim. The latter may return multiple extensions thus an heuristic for selecting the *winning extension* and extracting the *winning claim*, is needed. We argue that the cardinality of an extension is important: it indicates how many arguments support the same claim. However, this might be reductive when one bigger extension has an overall degree of truth lower than a smaller extension. Thus we also propose to consider the degree of truth of an entire extension. This can be intuitively computed averaging the degrees of truth of the premise of each argument, in the same extension. However, we propose to use the fuzzy algebraic product (fuzzy intersection) because we want to follow a pessimistic approach, giving more importance to a single low degree of truth of an argument in an extension rather than arguments with high degrees of truth. This approach produces a decreasing affect, where the final output is always smaller, or equal,

than the smallest contributing degree of truth. Such an approach is then applied to every extension computed by the selected semantic. Eventually, the extension that maximises both the number of arguments and the fuzzy algebraic product is declared to be the *winning extension*. This is a consistent conflict-free set of arguments that support the same claim, the *winning claim*, that can be used to support decision-making.

Definition 8 (Strength of extension). *The strength deg of an extension E is the product between its cardinality (compared to the cardinality of the pool of arguments), and the algebraic product of the degrees of truth of each forecast argument in E.*

$$E_{deg} = \frac{card(E)}{card(ARG_{pool})} \times \prod_{i=1}^{n} f_{\alpha_i}(x_i)$$

with f_{α_i} the membership function associated to the premise of the forecast argument $(a_i : f_\alpha \to c) \in ARG_F^{Act}$ in the extension E, x_i the input value for the attribute α that activated the argument a_i and $card(ARG_{pool})$ the cardinality of the pool of arguments.

Definition 9 (Winning extension and claim). *The winning extension WE of a set E of n preferred extensions, is the strongest extension.*
- $WE = \{A \mid A \in E, \text{ and } A_{deg} = max(E_{deg}^1, E_{deg}^2, ..., E_{deg}^n)\}$

The winning claim c is the claim supported by all the arguments of WE.
- $\forall a \in WE$ *with* $a : (f_\alpha \to c) \in ARG_F^{Act}$, *$c$ is the winning claim.*

Note 2. *Only forecast arguments computing the algebraic product because mitigating arguments do not carry a claim.*

Example 5. *In example 4, the grounded extension is empty, the 2 preferred extensions are: $p_1 = \{b, d, i, l, m, x\}$, $p_2 = \{h, n, x\}$. The degrees of truth of arguments are:*
b: $f_{age}^{medium} = 0.9$, d: $f_{Menop}^{Pre} = 1$, i: $f_{NodeInv}^{Low} = 0.97$, l: $f_{NodeC}^{False} = 1$, m: $f_{DegMalig}^{Low} = 0.6$
h: $f_{TumorSize}^{High} = 0.6$ n: $f_{DegMalig}^{High} = 0.5$
Argument x is in both the extensions according to preferred semantics, but as it is a mitigating argument, it does not have an associated claim, thus it does not contribute to the computation of the strength of each extension. According to definitions 8 and 9:
- Ex. 1*: (b, d, i, l, m); algebraic product: 0.52; % of args: (5/26=0.19) so $E_d = 0.1$*
- Ex. 2*: (h, n); algebraic product: 0.3; % of args.:(2/26=0.08) so $E_d = 0.024$.*
- *The* winning extension *is Ex.1 thus the* winning claim *is 'no recurrence (NR)'.*

4   Evaluation

We evaluated our approach using the Ljubljana Breast Cancer Dataset, adopted in many machine learning studies [12] [4] [2]. This includes 286 instances of real patients who went through a breast cancer operation. Each record is described by a set of features (table 1) whose value was recorded at the time of the operation, when the breast cancer was removed. 9 of these records are incomplete, due to missing values. For 81 patients, the breast cancer reappeared within 5 years of surgery (R), and for the remaining 205 cases, the cancer did not occur (NR). The goal was to compare the capacities of predicting the right prognosis (R or NR) by of the designed argument-based model and a selection of machine-learning classifiers. The attributes 'irradiation' and 'breast' were removed from the dataset because not accounted in the expert knowledge-base, thus

not used in the design of the contextual argumentation framework (table 2). The attribute 'irradiation' has been removed also because irradiation occurred after and not at the time of surgery. We have used WEKA machine learning software that implements state-of-the-art machine-learning classifiers. Six experiments were conducted varying fold cross-validation [2] and percentage of split [3].

To test the designed argument-based model predictive capacity, the winning claim, as per definition 9 was compared against the outcome class (R or NR) of the Ljubljana dataset, as in table 1, for each patient (record). Grounded and preferred extension, as described in 2, have been used as argument acceptability semantics. Results are in table 3. For 8 patients (out of 286) a non-empty grounded extension was computed: this is a clear coherent unique position. In 7 of these 8 cases, the winning claim coincides with the observed patient's recurrence status. In the remaining case, AT failed: the designed contextual argumentation framework was not enough for predicting the recurrence status: further knowledge/evidence is needed. With preferred semantics, 210 recurrence status (out of 286) were successfully predicted. The winning claim of the strongest preferred extension can be used for enhancing decision-making.

Table 3. Prediction rates: machine-learning vs. Argumentation theory

Classifier	10-folds	28-folds	40-fold	70& split	50% split	30% split
decision tables	73.42	75.52	73.42	73.25	74.12	74.00
bayesian network	72.37	73.07	73.07	68.60	72.70	73.00
best-first decision tree	66.78	70.62	73.07	62.79	74.12	72.00
regression	70.62	73.07	71.67	66.26	72.72	72.00
multilayer perceptron	65.38	68.88	65.73	58.13	65.03	65.00
alternating decision tree	74.47	75.17	74.82	65.11	69.93	72.50
Preferred semantic (AT)	**73.42**	**73.42**	**73.42**	**73.42**	**73.42**	**73.42**

5   Discussion and Related Work

Table 3 clearly emphasises the the high prediction rate of our model against machine-learning (ML) classifiers. Our approach does not require any training/learning and the output is always the same (unlike ML classifiers)[4]. Each case is evaluated independently by this CAF and the size of the dataset is negligible. An interesting property is that the 9 incomplete records of the Ljubljana dataset can still be considered using the CAF. With ML, an explicit algorithm to fill in the missing features of a record is required, if that record is accounted by the learning algorithm. Another interesting property of AT is its explanatory power. Firstly, the translation of knowledge-bases into a set of interactive argument is more intuitive, following a modular process based upon natural language

---

[2] *x-fold cross-validation*: the dataset was randomly reordered and split into x folds of equal size. For each iteration (total x iterations), one fold is used for testing the model while the remaining (x-1) folds are used for training the classifier. Results are then averaged over all folds giving the cross-validation estimate of the accuracy.

[3] *x% split*: x% of the records of the dataset is used for training the classifier and the remaining 100-x% is used to test the model and check its predictive capacity.

[4] We recall that in the experiments we have only evaluated just one expert's knowledge, which is not trained to fit the data, but it is used to build the CAF

terms familiar to clinicians. Secondly, the outcomes of an acceptability semantic, are conflict-free sets of the same input arguments, and not just numbers as produced by ML classifiers. In other words, the clinician not only can consider the strongest extension and the winning claim, but also s/he can individually take a look at each argument within this extension, being able to better justify final decisions. Aforementioned advantages are summarised as follows:

- *Inconsistency and Incompleteness: AT* provides a methodology for reasoning on available evidence, even if partial and inconsistent; missing data is simply discarded and even if an argument cannot be elicited, the argumentative process can still be executed with remaining data. This is powerful when a dataset is corrupted;
- *Expertise and Uncertainty: AT* captures expertise in an organised fashion, handling uncertainty and the vagueness associated to the clinical evidence, usually expressed with natural language propositions/statements;
- *Intuitiveness: AT* is not based on statistics/probability being close to the way humans reason. If the designer is anyway inclined to use statistical evidence, this can be modelled as an argument included in an argumentation framework; vague knowledge-bases can be structured as arguments built with familiar linguistic terms;
- *Explainability: AT* leads to explanatory reasoning thanks to the incremental, modular way of reasoning with evidence. *AT* provides semantics for computing arguments' justification status, letting the final decision be better explained/interpreted;
- *Dataset independency*: *AT* does not require a complete dataset and it may be useful for emerging knowledge where quantity evidence has not yet been gathered;
- *Extensibility and Updatability: AT* is an a open and extensible paradigm that allows to retract a decision in the light of new evidence: an argumentation framework can be updated with new arguments and evidence;
- *Knowledge-bases comparability*: *AT* allows comparisons of different subjective knowledge-bases. Two clinicians might build their own argumentation framework and identify differences in the definition of their arguments.

The above properties are not shared by ML classifiers, automatic procedures, that learn from previous examples. However some weaknesses of *AT* are:

- *knowledge-base translation*: the initial translation of a knowledge-base into interactive arguments may require effort, particularly with several pieces of evidence. In ML this translation is not required;
- *lack of learning*: *AT* is not a learning-based paradigm, thus rules/patterns cannot be automatically detected as in ML. However, ML relies on big datasets of evidence sometimes requiring not-negligible time to complete the learning process.

Although argumentation theory (AT) is gaining momentum in AI , the area of related works within health-care is sparse. In [3] AT was applied for group decision support among oncologists for discussing treatment therapies for larynx cancer cases. This approach showed how AT was promising in supporting decision justification. Hunter et al. investigated the role of *AT* as a mean for reasoning and comparing treatments in a typical health-care decision-making situation [12]. In particular, one of their work has been proved useful for breast cancer prognosis when applied in conjunction with Bayesian nets [7]. The work of Fox et al. illustrated how AT can be applied in different

practical medical scenarios. In [5] the application of assumption-based argumentation is described and applied to decision-making with early-stage breast cancer clinical data. Our study differs from the aforementioned works because it compares AT against ML.

6   Conclusions and Future Work

New technologies are undoubtedly useful for the advance of knowledge, especially in health-care and medicine. These facilitate clinicians' daily activities providing them with a wider range of tools for managing patients' information. However, despite the increasing amount of available information, decision-making is getting more complex because this new information, often incomplete and not coherent needs to be aggregated. Argumentation theory AT is an new paradigm, recently being considered in health-care for aggregating clinical evidence intuitively and modularly. It is not learning-based nor probability-based, but it is a knowledge-based paradigm that can work when evidence is limited, partial, incoherent and subject to uncertainty. It is built upon the representation of a knowledge-base into interactive arguments and it is capable of handling contradictions. These properties seem to be appealing for creating decision-support tools that follow a qualitative rather than a quantitative aggregation of evidence. The main contribution of this study was to show how AT can be practically applied in a real-world scenario: the breast cancer recurrence prediction. Results of experiments demonstrated how this knowledge-base approach can perform as well as state-of-the-art machine learning classifiers that however, are poor in explanatory power. Promising findings suggest further research can be carried towards a qualitative approach for enhancing decision-making. Future works will be on how to render the translation of knowledge-bases an intuitive/easy process with evaluations in different health-care settings.

References

1. Baroni, P., Guida, G., Mussi, S.: Full non-monotonicity: a new perspective in defeasible reasoning. In: European Symposium on Intelligent Techniques, pp. 58–62 (1997)
2. Cestnik, G., Konenenko, I., Bratko, I.: Assistant-86: A knowledge-elicitation tool for sophisticated users. In: Progress in Machine Learning, pp. 31–45 (1987)
3. Chang, C.F., Ghose, A., Miller, A.: Mixed-initiative argumentation: A framework for justification management in clinical group decision support. In: AAAI (November 2009)
4. Clark, P., Niblett, T.: Induction in noisy domains. In: Progress in Machine Learning (from Proceedings of the 2nd European Working Session on Learning), pp. 11–30 (1987)
5. Craven, R., Toni, F., Cadar, C., Hadad, A., Williams, M.: Efficient argumentation for medical decision-making. In: KR (2012)
6. Dung, P.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77, 321–357 (1995)
7. Hunter, A., Williams, M.: Argumentation for aggregating clinical evidence. In: 22nd International Conference on Tools with Artificial Intelligence, vol. 1, pp. 361–368 (2010)

8. Longo, L., Kane, B., Hederman, L.: Argumentation theory in health care. In: 25th International Symposium on Computer-Based Medical Systems (2012)
9. Matt, P., Morgem, M., Toni, F.: Combining statistics and arguments to compute trust. In: 9th International Conference on Autonomous Agents and Multiagent Systems, vol. 1 (2010)
10. Prakken, H.: An abstract framework for argumentation with structured arguments. Arguments and Computations (1), 93–124 (2010)
11. Toni, F.: Argumentative agents. In: Multiconference on Computer Science and Information Technology, pp. 223–229 (2010)
12. Williams, M., Williamson, J.: Combining argumentation and bayesian nets for breast cancer prognosis. Journal of Logic, Language and Information 15(1-2), 155–178 (2006)
13. Wu, Y., Caminada, M., Podlaszewski, M.: A labelling based justification status of arguments. Workshop on Non- Monotonic Reasoning, Studies in Logic 3(4), 12–29 (2010)
14. Wyner, A., Bench-Capon, T., Dunne, P.: Instantiating knowledge bases in abstract argumentation frameworks. Artificial Intelligence 1 (1995)