Mining Multimodal Sequential Patterns: A Case Study on Affect Detection

Héctor P. Martínez
Center for Computer Games Research
IT University of Copenhagen
Rued Langgaards vej 7
2300, Denmark
hpma@itu.dk

Georgios N. Yannakakis
Center for Computer Games Research
IT University of Copenhagen
Rued Langgaards vej 7
2300, Denmark
yannakakis@itu.dk

ABSTRACT

Temporal data from multimodal interaction such as speech and bio-signals cannot be easily analysed without a preprocessing phase through which some key characteristics of the signals are extracted. Typically, standard statistical signal features such as average values are calculated prior to the analysis and, subsequently, are presented either to a multimodal fusion mechanism or a computational model of the interaction. This paper proposes a feature extraction methodology which is based on frequent sequence mining within and across multiple modalities of user input. The proposed method is applied for the fusion of physiological signals and gameplay information in a game survey dataset. The obtained sequences are analysed and used as predictors of user affect resulting in computational models of equal or higher accuracy compared to the models built on standard statistical features.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human factors*; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Games*

General Terms

Algorithms, Experimentation, Human Factors

Keywords

sequence pattern mining, sequence classification, preference learning, heart rate variability, skin conductance, game events

1. INTRODUCTION

Most of the modalities available for the analysis and modeling of interaction with computers such as speech, gestures, physiology and key presses (among others) are temporal by nature. However, multimodal signal fusion and modeling methods require the transformation of those streams of data into a fixed-sized vector of scalar features that describe different characteristics of the signals; such features are given in a format suitable for empirical analysis and computational modeling. This process, namely *feature extraction*, generally involves a domain expert who chooses a set of statistical features (e.g. average values) based on the available data in an attempt to capture all the relevant qualitative characteristics of the signals.

This paper applies *frequent sequence mining* [1] techniques to automatically extract *sequential features* which capture a range of signal characteristics and potentially reveal important interaction factors hindered by traditional feature extraction. Sequential features can be used to analyse the temporal trends within and across modalities but also provide appropriate input vectors for classification or prediction models which are built in multimodal input spaces. Frequent sequence mining has been applied before to *sequence classification* in different tasks including protein classification and plan monitoring [4, 8]; however, this is the first time — to the best of the authors' knowledge — that has been applied to mine data streams across multiple user input modalities. With this approach, the different sources of information are combined before building a model (data-level fusion [14]) achieving deeper fusion than feature-level or decision-level fusion that facilitates a low level interpretation of the interrelation between modalities and its effect on the target user output. Even though there exist machine learning approaches that stream input data (e.g. Recurrent Neural Networks, Hidden Markov Models and Dynamic Bayesian Networks), the extraction of sequential features makes time-related information available for and generic across any other non-temporal based technique. Furthermore, the analysis of dynamic models is often far from trivial due to recurrent connections or complex state transitions.

We evaluate the proposed method in a game survey dataset that includes gameplay logs and physiological recordings of players. The Generalized Sequential Patterns (GSP) [18] sequence mining algorithm is utilized to find sequential patterns among the gameplay metrics and two physiological signals — skin conductance (SC) and blood volume pulse (BVP) — from the players. The sequences are analysed and compared to a set of statistical features as inputs to Artificial Neural Network (ANN) models trained to predict players' affective states reported as pairwise preferences. The results of this initial study show that sequential features yield models of affect that are equally or significantly more accurate

ICMI'11, November 14–18, 2011, Alicante, Spain.

than the models built on simple statistical features for all affective states examined.

2. RELATED WORK

This work applies sequence mining to fuse physiological signals and gameplay information to generate an input vector for the prediction of player emotional preferences. This section reviews earlier work on physiology and game-context fusion and studies in which sequential patterns have been used for prediction and classification tasks.

2.1 Fusion of Physiological Signals and Game Context Information

In the interaction between a user and a system, the user's inputs — both low level (e.g. moving the mouse) and high level (e.g. dragging an icon) — and their consequences on the system (e.g. cursor and icon moved) determine a key component of user experience. Fusing such information with other modalities can yield user experience models that are not only more accurate but also more informative since context is added to the model [20].

Within game-player interaction studies, McQuiggan et al. [13] created predictors of self-efficacy reports that rely on statistical features extracted from several heart rate and skin conductance signal intervals in conjunction with players' self-reports and information about the visited positions of the avatar and the cursor in a 3D learning environment. In that study, the sequential information from the bio-signals is transformed into a vector of features representing the average of the signals over different intervals. The gameplay data, in turn, is reduced to a set of spatio-temporal features such as the time spent on a task or a list of locations visited. On the same basis, Martinez et al. [12, 20] trained predictors of self-reported preferences in a 3D prey/predator game on statistical features derived from heart rate, skin conductance, blood volume pulse and different game events. In that study, the physiological statistical features (e.g. average, standard deviation and first and second absolute differences) are extracted from the full length of the game experience (i.e. 90 seconds) while the gameplay features consist of statistical information such as distance to enemies or pellets and performance measures such as the final score and the fraction of the map explored.

While there exist several studies in player experience research that use game context information and bio-signal information for the analysis and modeling of experience (e.g. [9, 10, 3] among others) none has fused the time-series data of these modalities by considering sequential patterns across modalities.

2.2 Sequence Classification using Frequent Patterns

Frequent sequential patterns are typically mined to detect interesting common trends in the data and discover association rules, correlations and other relationships [7] but they can be applied also to the classification and clustering of sequences.

Protein homology detection and classification is one of the most popular sequence mining tasks. In [2], a set of relevant subsequences (motifs) are predefined and each protein is represented as a vector of attributes; each attribute represents the number of occurrences of a motif in the represented protein (generally, each motif will either occur once or not occur).

In [4] a Naïve Bayes Classifier identifies the most probable family to which a protein belongs to using as inputs the number and average length of the frequent subsequences shared within each of the protein families.

One of the many differences between the task of classifying proteins and processing multimodal signals is the temporal nature of the signals: in a protein string every pair of consecutive elements has a distance of 1 unit while in a signal the time distance between two elements is derived by the sampling rate. This difference has an impact on the structure of sequences (e.g. in multiple time-series sequences two events might occur simultaneously), the matching conditions of sequential patterns and the procedure to match sequential patterns (e.g. in non-temporal sequences the number of gaps between elements can be constrained whereas in a temporal sequence the time between elements is constrained instead).

Lesh et al. [8] define an effective method to mine features for sequence classification. This method consists of mining all sequential frequent patterns and prune those that are either not distinctive of one of the target classes or correlated with a pattern already selected. This pruning stage is necessary since frequent mining can produce an enormous amount of features that cannot be efficiently handled by a classifier. In this paper, a similar approach to [8] is used but the set of frequent sequential patterns is reduced by automatic feature selection which searches for the combinations of sequences that are more relevant for predicting a target output (i.e. affective state in the case study presented).

3. FREQUENT SEQUENCE MINING

In this study we focus on the extraction of sequences that combine events across different user input modalities collected during the interaction with a system. We expect these sequences to be critical for the analysis of the interaction in terms of the interplay among different modalities of user input. We also anticipate that frequent sequences across multimodal signals can improve the detection accuracy of user affect over other approaches that involve the reduction of the signals through standard feature extraction. In this section we formalize the problem of mining frequent sequences and describe the algorithm applied to the case study.

Given a dataset in which each sample is a sequence of events, namely *data-sequence*, a sequential pattern defined as a subsequence of events is a *frequent sequence* if it occurs in the samples of the dataset regularly. In general, the events that form the sequences correspond to relevant changes on the observed signal or system and are associated with a time stamp (discrete moment on time when the event occurs) and an identification (type of event). An event could be for example an increase on heart rate, a key pressed or an action unit activation in a facial expression task [15].

More formally, a sequential pattern is an ordered list of *elements* — denoted as $\langle e_0 e_1 \dots e_n \rangle$; e_i is the i^{th} element of the sequence — each containing a non-empty (unordered) set of m simultaneous *events* — denoted as $(x_0, x_1, \dots, x_m)$; x_i is an event. For example, an element could be two keys pressed simultaneously, several action units executed at the same time or an increase on heart rate (an element with only one event). A frequent sequence can be defined as a sequential pattern that is *supported* by, at least, a minimum

Algorithm 1 Generalized Sequential Patterns.

Input: a dataset of data-sequences, S_{min}, G_{max} and W_{max}
Output: the set of sequence patterns that are supported by more than S_{min} data-sequences.

procedure GSP(data, S_{min}, G_{max}, W_{max})
Count the number of data-sequences in which each different event is contained (support count).
Insert the events with a support count greater than S_{min} into the set of frequent 1-sequences (L_1).
k = 1
while L_k is not empty
    Generate the set of candidate $(k+1)$-sequences (C_{k+1}). See Algorithm 2 for more details.
    Determine the support count of the sequences contained in C_{k+1}.
    Create the set of frequent $(k+1)$-sequences (L_{k+1}) with the sequences in C_{k+1} that present a support count greater than S_{min}.
    k = k +1
return $L_1 \cup L_2 \cup \cdots \cup L_{k-1}$

Algorithm 2 Candidate generation in GSP.

Input: a set of $(k-1)$-sequences L_{k-1}.
Output: the set of candidate k-sequences C_k.

procedure generateCandidates(L_{k-1})
for each pair of sequences $s_x, s_y \in L_{k-1}$ with $s_x = \langle e_1^x e_2^x \ldots e_n^x \rangle$ and $s_y = \langle e_1^y e_2^y \ldots e_n^y \rangle$
    if the two sequences obtained by dropping the first event of s_x and the last event of s_y are identical
        if e_n^y has only one event $e_n^y = (y_1)$
            Generate the candidate sequence s_{xy} by inserting y_1 as last event of e_n^x: $s_{xy} = \langle e_1^x e_2^x \ldots e_{n-1}^x (e_n^x, y_1) \rangle$
        else
            Generate the candidate sequence s_{xy} by replacing e_n^x with e_n^y: $s_{xy} = \langle e_1^x e_2^x \ldots e_{n-1}^x e_n^y \rangle$.
        if all contiguous subsequences of s_{xy} are contained in L_{k-1}
            Insert s_{xy} into C_k.
return C_k

amount of data-sequences as determined by the *minimum support* (S_{min}) value. A data-sequence supports a sequential pattern if and only if it contains all the events present in the pattern in the same order. Note that this definition does not restrict that events in consecutive positions within the pattern must be strictly consecutive in the data-sequence. For example, the data-sequence $\langle x_0 x_1 x_2 x_3 x_4 x_5 \rangle$ supports the pattern $\langle e_0 e_5 \rangle$ with $e_0 = (x_0)$ and $e_5 = (x_5)$ if further constrains are not specified. The amount of data-sequences that support a sequential pattern is referred as the *support count*.

In this paper we extent the basic definition of frequent sequence with two of the generalizations proposed in [18] which are as follows:

- Sliding window: given an element e_i containing two or more simultaneous events $(x_0, x_1 ... x_m)$, a data-sequence contains e_i if and only if all events occur in the data-sequence within a given time window W_{max}. In other words, two or more events that occur in less than W_{max} may be considered to occur simultaneously.
- Time constrain: given two consecutive elements in a pattern, $e_i e_{i+1}$, a data-sequence may support the pattern only if both elements occur in the specified order and the time difference between their occurrences is lower than a specified time threshold, *maximum gap* (G_{max}).

3.1 GSP algorithm

The Generalized Sequential Patterns algorithm [18] is used for mining the frequent sequences in this study. GSP is a *candidate generation and test* algorithm which supports the constrains mentioned above. It first finds the frequent sequences with one single event, namely 1-sequences. That set of sequences is self-joined to generate all 2-sequence candidates for which we calculate their support count. Those sequences that are frequent (i.e. their support count is greater than a threshold value S_{min}) are self-joined to generate the set of 3-sequence candidates. The algorithm is iterational increasing the length of the sequences in each algorithmic step, until the next set of candidates is empty. The basic principle of the algorithm is that if a sequential pattern is frequent, then its *contiguous* subsequences are also frequent. Given two sequences s_x and s_y, s_y is a contiguous subsequence of s_x if either: 1) s_y is obtained by dropping the first or last event of s_x; or 2) s_y is obtained by dropping an event from an element of s_x with two or more events; or 3) there exists a sequence s_z such that s_z is a contiguous subsequence of s_x and s_y is a contiguous subsequence of s_z.

By self-joining a set of frequent sequences of length k, the algorithm obtains only the $(k+1)$-sequences whose contiguous subsequences are frequent, thereby, reducing the number of sequential patterns for which support counts have to be determined. The reader is referred to Algorithm 1 and Algorithm 2 for a more detailed presentation of the basic steps of the GSP algorithm.

4. USER EXPERIENCE MODELING

The dataset used in this paper contains sequences labeled with pairwise preferences of affect (see Section 5). Neuroevolutionary preference learning is applied to model those preferences relying on a subset of the available features selected automatically through Genetic Feature Selection. The two algorithms are briefly described in the following subsections.

4.1 Genetic Feature Selection

Feature selection (FS) is essential in scenarios where the available features do not have a clear relationship and, thus, impact to the prediction of a target output (i.e. it is not easy to decide *a priori* which features are useful and which are irrelevant for the prediction). Moreover the computational cost of testing all available feature sets is combinatorial and exhaustive search might not be computationally feasible in large feature sets. Under these conditions, FS is critical for finding an appropriate set of model input features that can yield highly accurate predictors [20].

Genetic feature selection (GFS) [12] is a global search FS algorithm guided by a genetic search. The fitness function is calculated as the average cross validation performance on unseen folds of classification data. The search starts by evaluating the fitness of several subsets with one feature; in subsequent iterations combinations of the fittest subsets from the previous iterations are evaluated. The algorithm stops after a fixed number of iterations or when highly fit feature subsets are found. More details about GSF can be found in [12].

4.2 Neuroevolutionary Preference Learning

We apply preference learning [5] to build affective models that predict users' self-reported emotional preferences based on the subsets of features selected by the GFS algorithm. In this study, the models are implemented as single layer perceptrons (SLPs) that are trained via neuroevolutionary preference learning (as in [19, 20]) to map the selected features to a predictor of the reported pairwise emotional preferences. The expressivity of SLPs allows us to analyse the impact of each one of the selected features to the reported affective preferences. For instance, when a feature with a corresponding high connection weight value increases from one game to another, the magnitude of the predicted preference is increased or decreased depending on the sign of the weight value. On the other hand, weight connections with low values have a small impact on the prediction of preferences.

Note that the pairwise preference relationship of the training data is known (e.g. game A is preferred to game B) but the value of the target output is not (i.e. the magnitude of the preference is unknown). Thus, any gradient-based optimization algorithm is inapplicable to the training problem since the error function under optimization is not differentiable.

5. DATA COLLECTION

The dataset used in this paper was gathered during an experimental game survey in which 36 participants played four pairs of different variants of the same video-game. The test-bed game named *Maze-Ball* is a 3D prey/predator game that features a ball inside a maze controlled by the arrow keys. The goal of the player is to maximize her score in 90 seconds by collecting the pellets scattered in the maze while avoiding the red enemies that wander around. The eight available variants of the game differ only on the virtual camera profile used which determines how the virtual world is presented on screen.

Blood volume pulse and skin conductance were recorded at 32Hz during the session; heart rate (HR) is inferred from the BVP signal every 5 seconds. Moreover, both gameplay information and user keystrokes are logged. The players filled in a 4-alternative forced choice questionnaire after completing a pair of variants reporting whether the first or the second game of the pair felt more anxious, challenging, exciting, frustrating, fun and relaxing, or whether both felt equally, or none of them did. The details of the Maze-Ball game design and the experimental protocol followed can be found in [20, 12].

5.1 Sequential Features

A data-sequence is created from the logs of each game, 224 in total, by concatenating all the events logged in temporal order. The following list describes the events we have chosen for this initial study:

- Performance Events
  - Player collects a pellet ($): 10 identical pellets are placed in different areas of the maze enforcing a difference of at least few seconds between two pellets. This event is picked as it is expected to have an impact on reported challenge and fun (among other reported user states).
  - Enemy hits the player (E): 14 enemies follow predefined paths guarding a pellet causing this event to occur very close in time with the $ events frequently. Enemy hits are selected as events since enemies are critical to player experience in a prey/predator game.
  - Countdown starts (t^{10}): when entering the last 10 seconds of the game the timer changes its color *rushing up* the player. This event occurs exactly once in each game, thus it does not provide sufficient information about the experience per se. However, sequences combining this event with physiological events or other gameplay events are expected to have a direct impact to reported anxiety and excitement.
- Navigation Events
  - Moving to a new area of the maze ($m^0, \ldots, m^7$): although there are not explicit boundaries between areas of the maze, 8 different sectors can be distinguished based on the different wall layout, placement of the pellets and movement of the enemies which, in turn, represent different degrees of difficulty. These events are expected to have a direct impact on the challenge reports.
  - Press an arrow key (▲, ▼, ◄, ►): pressing the right and left arrows make the ball turn if it is located in a corner; the down key forces the ball to turn 180^o and the up arrow has no effect. Each single one of these events most likely holds a tiny piece of information about user experience; however, sequences combining many of these events may point to more complex navigation patterns with a potential impact on experience.
  - Inactivity for more than 1 second (*Stop*): the player avatar is moving forward at any time unless it hits a wall. In that case, the ball will only continue moving if the player turns. When the ball is stopped for 1 second, the event is logged. It could indicate that the player is planning a strategy or waiting for an enemy to move away from a pellet. Thus, this event is relevant for the identification of a player's behavioral patterns and, indirectly, for affect detection.
- Physiological Events
  - Difference between two inter-beat intervals (*RR intervals*) is greater than 50 ms (r^{+50}, r^{-50}): the heart beats are detected from the BVP signal and when two consecutive inter-beat intervals differ for more than 50 ms, an event is logged. The threshold of 50 ms is commonly used in affective and medical studies [6, 20] as an indicator of arousal which in turn is one of potential identifiers of the affective states examined.
  - SC increase/decrease ($s^{\Uparrow}, s^{\Downarrow}$): sudden changes in the SC signal are detected and logged as events.

They are normally detected in pairs: after $s^{\Uparrow}$ the SC will increase for a while and $s^{\Downarrow}$ will be logged when it starts decreasing. These SC signal events are picked because they suggest changes in sympathetic activity and, thereby, may relate to reported user experience.

This list of events is far from being inclusive of all relevant events available such as those derived from heart rate variability indexes. We consider the study of more sequential features out of the scope of this initial study which focuses on the efficacy of frequent sequence mining across modalities of user input for affect detection.

After the frequent sequences have been found in the full dataset, each data-sequence is transformed into a vector of sequential features. Each feature corresponds to the number of occurrences of a frequent sequence within the data-sequence that is been transformed. An alternative representation transforms each feature into a boolean value which is 0 when there are no occurrences of the associated sequential pattern in the data-sequence and 1 otherwise.

5.2 Statistical Features

A total of 42 statistical physiological features are extracted from the three physiological signals (BVP, SC and HR) including, for example, average and standard deviation of the three signals and heart rate variability measures such as standard deviation of the inter-beat time intervals. In addition, a total of 41 statistical gameplay features are extracted from the game metrics logged such as average and standard deviation of the distance to the closest pellet and enemy, and final score. For a full description of the statistical features extracted from both modalities of input, the reader is referred to [20, 11].

6. EXPERIMENTS

The first part of this section focuses on the process of mining and the analysis of multimodal frequent patterns obtained in the dataset examined. In the second part, sequence patterns are tested as inputs of computational models of affect.

6.1 Sequence Mining

In this section we first discuss the effects of the different GSP parameters in the task of mining sequences from multimodal datasets and later analyse the frequent sequences found in the Maze-Ball dataset.

6.1.1 Parameter Tuning

We set up the S_{min} to 100 sequences which forces a sequence pattern to occur in at least 44.64% of the samples (224 samples in total) to be considered frequent. We choose this high threshold because of the low *specificity* of the events that yields a high frequency of each event within the data-sequences. For example, consider the event of picking any of the ten available pellets ($) and ten different events associated to a specific pellet in the game $\{\$_0, \ldots, \$_9\}$. The support count for each $\$_x$ is expected to be lower than the support count of $ as different players will most likely pick different pellets. Consequently a lower minimum support would be required to consider sequence patterns containing the 10 specific $\$_x$ events frequent.

The W_{max} is chosen as a trade off between the frequency of the events within and across signals. On one hand, we require a low window to not consider simultaneous events that within the same signal are clearly not occurring at the same time (e.g. in a 90 second-long game, two gameplay events of more than a few seconds apart should not be processed as two simultaneous events). On the other hand, we need to allow a certain window to consider events from different asynchronous signals to occur at the same moment in time. For example, a variation on the RR signal is linked to the time interval between two heart beats; a gameplay event happening during this interval should be considered simultaneous to the RR variation even though the time stamps of both events are not identical. In all experiments reported in this paper W_{max} is 1 second.

The G_{max} parameter has a direct effect on the number of events that can be concatenated in a sequence and the support count of those sequences. Note that the definition of a sequence pattern does not require two consecutive events to occur immediately one after the other in the data-sequences. Consequently when G_{max} is much larger than the frequency of events in one modality, a high number of events can be skipped when matching a pattern with a data-sequence. For example, if we select a G_{max} of 5 seconds and we consider the sequence (►)(►), the number of data-sequences containing those two events in a 5 second interval is very high (arrow keys are pressed every 1-3 seconds) even though none of the players pressed the right arrow key twice in a row. This effect is enhanced by the fact that we are searching for sequences from different modalities. For instance, if we find the pattern (►)($s^{\Uparrow}$)(►) with a 5 second G_{max}, this sequence is supported by any data-sequence in which a player pressed the right key in a 5 second interval before and after his SC raised — independently of other keys being pressed before and after $s^{\Uparrow}$ or other events such as $ and E occurring. This has not only an effect on the informative value of the sequences but also increases greatly the number of frequent patterns (see Table 1). On the other hand, a large G_{max} may allow higher level sequences — i.e. sequences of high level events that do not occur very frequently — to emerge such as ($)($)($). Thus, when setting up this parameter, it is necessary to take into account the frequency of each event and the desired target sequences. In this initial study we decided to focus on the relationship between gameplay events and physiological responses and set up a G_{max} of 1.5 and 3 seconds which are both close to the frequency of the key presses (the most frequent event in the dataset). The two selected G_{max} values also approximate the physiological time responses to game events reported in several studies [16, 17].

6.1.2 Sequence Analysis

A large amount of frequent sequences were found in the dataset; specifically, looking at the 3 and 5 second G_{max} values (see Table 1) one may observe the exponential increase of frequent sequential patterns with regards to the number of events considered. The large amount of patterns is caused, in part, by the right and left key press events which are so frequent in the data-sequences that result in almost *wild card* events (i.e. between any two events is very likely that either the right or the left key are pressed).

The more restrictive G_{max} of 1.5 seconds did not produce any frequent sequence combining the sector events with

Table 1: Amount of frequent sequential patterns for different values of G_{max}.

# events	G_{max}		
	1.5	3	5
1	19	19	19
2	161	188	232
3	785	1650	2113
4	505	8387	18259
5	28	18985	118192
6	0	21704	545667
7	0	11251	NaN
8	0	1954	NaN
9	0	36	NaN

Table 2: Support counts of a subset of frequent sequences containing key board events.

Sequences	G_{max}	
	1.5	3
$(s^{\Uparrow})(s^{\Downarrow})(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)$	< 100	108
$(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)$	< 100	108
$(\blacktriangleleft)(s^{\Uparrow})(s^{\Downarrow})(s^{\Uparrow})(s^{\Downarrow})(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)$	< 100	101
$(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)(\$)$	< 100	104
$(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)(\$)(\blacktriangleleft)$	< 100	104
$(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\$)$	< 100	149
$(\blacktriangleright)(\blacktriangleleft)(m^7)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)$	< 100	105
$(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(E)$	< 100	104
$(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(E)(\blacktriangleright)(\blacktriangleleft)$	< 100	102
$(\blacktriangleleft)(s^{\Uparrow})(s^{\Downarrow})(\blacktriangleright)(E)(\blacktriangleleft)$	< 100	101
$(\blacktriangleleft)(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)$	115	202
$(\blacktriangleright)(s^{\Uparrow})(\blacktriangleright)(s^{\Downarrow})(\blacktriangleleft)$	106	192
$(s^{\Uparrow})(\blacktriangleright)(s^{\Downarrow})(\blacktriangleright)(\blacktriangleleft)$	101	187

the key presses making more difficult, if not impossible, to map a sequence of key presses to an area in the maze (e.g. $(\blacktriangleleft)(\blacktriangleright)(\blacktriangleright)(\blacktriangleleft)(\blacktriangleleft)$). Even though specific paths can not be inferred from the key presses, some interesting results can be observed. For 1.5 and 3 second G_{max} values the sequences of maximum length found consist only of the events $\{\blacktriangleleft, \blacktriangleright, s^{\Uparrow}, s^{\Downarrow}\}$; neither other gameplay events nor RR variations are included in the most frequent sequences (see Table 2). These sequences suggest that the combination of key press patterns in the Maze-Ball game and SC variation are frequent and should be considered in the analysis of player experience. Furthermore, in the longest sequences that combine key presses with the events $\{E, \$\}$ (see Table 2), the latter always appear in the last position of the sequences suggesting that players follow similar strategies to approach pellets and enemies but more dissimilar behaviours are presented after picking a pellet or being hit by an enemy.

Table 3 shows some of the most frequent 2-sequences and 3-sequences that combine the main performance events, \$ and E, with physiological events. With the more restrictive value for G_{max} (1.5 second), all the frequent 3-sequences contain the subsequence ($s^{\Uparrow}$)($s^{\Downarrow}$) with the event \$ or E in any position, being the most frequent sequence in which the gameplay event occurs simultaneously with an increase of SC, followed by a decrease of SC. The 3 second G_{max} on the other hand, produced almost any combination of two physiological events with one of the gameplay events. This might indicate that the threshold is too large and does not capture a meaningful fusion of the modalities.

Table 3: Support counts of a subset of the most frequent sequences including physiological and performance events. Events enclosed in the same parentheses occur simultaneously (in any order within an interval of 1 second).

3-sequences	G_{max}		2-sequences	G_{max}	
	1.5	3		1.5	3
$(\$\ s^{\Uparrow})(s^{\Downarrow})$	141	168	$(\$\ s^{\Uparrow})$	184	184
$(E\ s^{\Uparrow})(s^{\Downarrow})$	131	163	$(\$\ r^{-50})$	178	178
$(s^{\Uparrow})(\$)(s^{\Downarrow})$	123	164	$(E\ s^{\Uparrow})$	175	175
$(s^{\Uparrow})(E)(s^{\Downarrow})$	116	164	$(E\ s^{\Downarrow})$	174	174
$(s^{\Uparrow})(s^{\Downarrow})(\$)$	112	181	$(E\ r^{+50})$	174	174
$(E)(s^{\Uparrow})(s^{\Downarrow})$	109	175	$(\$\ s^{\Downarrow})$	170	170
$(s^{\Uparrow})(s^{\Downarrow})(E)$	106	180	$(s^{\Uparrow})(\$)$	166	194
$(\$)(s^{\Uparrow})(s^{\Downarrow})$	105	186			
$(s^{\Uparrow})(\$\ s^{\Downarrow})$	105	158			
$(s^{\Uparrow}\ s^{\Downarrow})(\$)$	102	139			

The frequent 2-sequences correspond to all possible combinations of one of the gameplay events with one physiological event showing more occurrences when combined in the same element. Note that the count support indicates only the number of data-sequences in which the sequence pattern appears and not the number of occurrences within each data-sequence. Opposite to long sequences, these short patterns are expected to occur more than once in each sequence. Therefore, the number of occurrences of the patterns within each data-sequence is required for a full analysis of the physiological responses to game events. This study, however, is out of the scope of this paper and will be examined in future work.

It is worth mentioning that none of the other high level game play events occur frequently with the physiological responses. While this is not entirely surprising for the sector events — given that there is no visual or audio feedback when changing sectors in the maze — one would expect that the count down event would trigger a change on the player that would have been reflected on her physiology. However, such a relationship is not observed frequently via event sequences.

6.2 Preference Learning

For the preference learning experiments reported here we chose to use only the frequent sequences found using $G_{max} = 1.5$, $S_{min} = 100$ and $W_{max} = 1$ that did not contain the key board and *Stop* events. This last constrain reduces substantially the number of frequent patterns from 1498 to 140 in an effort to lower the dimensionality of space and the computational cost of training. Preliminary studies showed that using the full set of frequent patterns did not improve the prediction accuracy of the user preference models.

In this section we first test two approaches to transform the sequence patterns into feature vectors as those were mentioned in Section 5 and we then built and compare computational models of affect which are based on either sequential or statistical features, or both. In all tests GFS is applied to select the inputs of the computational models. The accuracy of the models is calculated as the average 3-fold cross-validation accuracy on predicting unseen pairwise preference data.

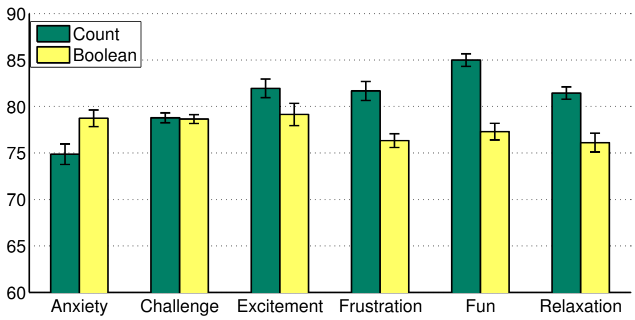

Figure 1: Average accuracy (3-fold cross-validation) and standard error of 10 runs of GFS in sequential multimodal features using the count of occurrences (count) or the existence to a data-sequence (boolean) as representation.

6.2.1 *Sequences as User Preference Model Input*

Figure 1 shows the average accuracy of 10 user preference ANN models built on one of the two different representations of the frequent pattern: each frequent sequence is represented as its corresponding number of occurrences in each data-sequence (count) or alternatively represented as a boolean feature that is 0 when the frequent sequence does not occur in the data-sequence and 1 otherwise (boolean).

For the case study dataset examined, in which the frequent patterns are short and expected to occur a variant number of times across samples, the count representation is expected to yield higher accuracies since the boolean representation is far less informative. Results show that this hypothesis is valid for all affective states but anxiety. Apparently, the boolean representation provides sufficient information for predicting the reports of anxiety whereas it proves to be insufficient for the remaining of the user states examined.

6.2.2 *Comparison Between Sequential and Statistical Features*

Figure 2 depicts the average accuracies of ten ANN models built on subsets of features selected by GFS from a set of statistical features, a set of sequential features and a set combining both feature sets. Results show that the accuracies of models built on sequential patterns are comparable to models trained on statistical features and even significantly higher on fun — 1.48% higher accuracy ($p - value < 0.1$)— and challenge — 6.48% higher accuracy ($p - value < 0.01$) — showing that information about short intervals of the user experience across modalities is as valuable as information about the overall user experience for the prediction of affective self-reports.

Some of the information gathered by sequential patterns overlap with the information contained in statistical features (e.g. the final score of the player can be inferred from the 1-sequences ($) and ($E$)); however, most features are specific to one of the feature extraction approaches and provide complementary information about the multimodal interaction. Thus, it is surprising that models that can combine both sources of information cannot, in general, outperform models built on one of the sets of features only. Figure 2 shows that only in frustration and anxiety such models present a slightly higher average accuracy, which is not statistically

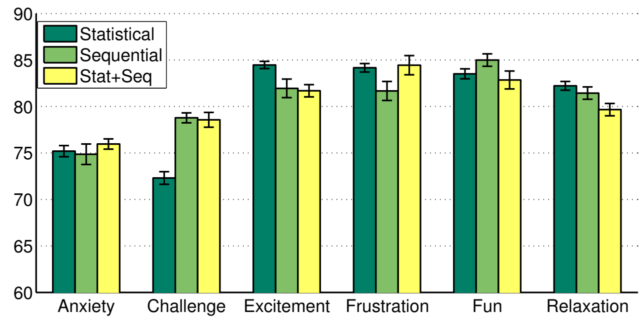

Figure 2: Average accuracy (3-fold cross validation) and standard error of 10 runs of GFS in statistical and sequential multimodal features.

significant in either dataset. This could indicate that in this particular dataset — as already reported in [20] — it might be impossible to find a subset that yields a higher performing predictor with a simple model such as a single layer perceptron. Additionally, the search space for automatic feature selection is increased substantially when joining the two sets of features, complicating the task of effectively finding the *optimal* subset of ANN inputs through genetic search. This is, in part, supported by the lower accuracy of the excitement, fun and relaxation models trained on both input sets than only trained on each one of them.

Table 4 depicts the best models trained for challenge and fun on sequential features. The challenge model combines sequences of physiological responses with only two gameplay sequences and no multimodal sequences. The sequence $(s^{\Downarrow})(s^{\Uparrow})$ associated with a negative weight suggests that repeated changes on SC — typically a decrease follows an increase — contribute to a lower challenge value. This relationship is consistent with results reported in [20] where higher SC variance is linked to lower challenge values. However, the effect on challenge is inverse when the sequence is preceded by a positive change in the RR intervals — $(r^{+50})(s^{\Downarrow})(s^{\Uparrow})$ — suggesting that sudden changes in sympathetic arousal indicated by both HR and SC variability lead to higher challenge values. A somewhat surprising result is that the predicted challenge is lower when pellets are picked and simultaneously an enemy hits the player. This could be a sign that some players do not try to get around the enemies to pick the pellets (which is a difficult and time consuming task) and instead prefer to pick a pellet as quickly as possible and run towards the next one even if an enemy is next to it. This further suggests that the negative reward scheme for enemies should to be increased for a more balanced game design.

Observing the best-performing ANN model of fun (see Table 4) it seems that the total number of pellets picked ($) has a positive impact on reported fun as well as when a sudden peak in SC is generated just before or just after picking a pellet — $(s^{\Uparrow})(s^{\Downarrow})(\$)$ and $(\$\ s^{\Uparrow})(s^{\Downarrow})$ — which could be related to a heightened arousal state when the player is about to pick a pellet. Sudden increases on the RR intervals length $(r^{+50}\ r^{+50})$ have also a positive impact on reported fun. On the other hand, enemy hits accompanied by a sudden decrease on SC $(E\ s^{\Downarrow})$, not surprisingly, seem to decrease the level of fun. Note, that a single event E would not necessarily have a negative effect on fun as it is a fundamental

Table 4: Input features and corresponding connection weights for the highest performing ANN models for challenge and fun

Challenge		Fun	
$(s^{\Uparrow}\ s^{\Downarrow})(r^{-50})$	5.00	$(\$)$	4.80
$(s^{\Uparrow}\ r^{+50})(r^{-50})(s^{\Downarrow})$	-5.00	$(r^{+50}\ r^{+50})$	4.69
$(s^{\Downarrow})(s^{\Uparrow})$	-4.82	$(E\ s^{\Downarrow})$	-4.56
$(\$\ E)$	-3.61	$(s^{\Downarrow}\ r^{-50})(r^{+50})$	-3.49
$(r^{-50})(s^{\Downarrow})$	-2.22	$(s^{\Uparrow})(s^{\Downarrow})(\$)$	3.06
$(r^{+50})(s^{\Downarrow})(s^{\Uparrow})$	2.21	$(\$\ s^{\Uparrow})(s^{\Downarrow})$	2.50
$(s^{\Downarrow}\ r^{+50}\ r^{-50})$	1.67	$(\$\ m^6)$	-1.19
(r^{-50})	1.52	$(s^{\Uparrow})(r^{-50})(r^{+50})$	0.67
$(\$\ m^6)$	-0.05	$(r^{+50})(r^{-50})(s^{\Downarrow})$	0.09
		$(s^{\Uparrow}\ r^{+50})(s^{\Downarrow}\ r^{-50})$	-0.18

part of the game; however, a decrease on SC might indicate a lowered level of the player's arousal as consequence of the game event.

7. CONCLUSIONS

In this paper we propose sequential pattern mining as a method to explore the relationship between asynchronous signals from different modalities for the discovery of frequent event sequences across modalities. We applied the GSP sequence pattern mining algorithm to a game dataset that includes physiological signals (blood volume pulse and skin conductance), context-based game metrics (e.g. key board presses) and self-reported affective preferences. This study served as a starting point to highlight the main advantages, limitations and practical considerations of this approach when applied to the analysis of multimodal interactions.

Additionally, as an alternative to standard statistical features, the frequent sequences mined are presented as inputs of affect detectors assisting the process of finding more accurate models of user affect and experience. The resulting affective models are analysed and compared against models trained on statistical features. The analysis of the affective models trained on the sequences found in the game dataset reveals relationships between sudden arousal level changes across physiological signals, critical game events and reported affect. Sequences outperform statistical features in two (fun and challenge) out of the six affective states examined in this study when used as inputs of affect detectors; the sequential features yield similar performances to standard statistical features in the other four affective states: anxiety, excitement, frustration and relaxation. Future work will aim to validate the proposed method in dissimilar datasets, including different modalities of user input, and explore a richer set of multimodal events for sequence mining.

8. REFERENCES

[1] R. Agrawal and R. Srikant. Mining sequential patterns. In *icde*, page 3. Published by the IEEE Computer Society, 1995.

[2] A. Ben-Hur and D. Brutlag. Remote homology detection: a motif based approach. *Bioinformatics*, 19(suppl 1):i26-i33, 2003.

[3] A. Drachen, L. Nacke, G. N. Yannakakis, and A. L. Pedersen. Correlation between heart rate, electrodermal activity and player experience in First-Person Shooter games. In *SIGGRAPH'10*, 2010.

[4] P. Ferreira and P. Azevedo. Protein sequence classification through relevant sequence mining and bayes classifiers. *Progress in AI*, pages 236–247, 2005.

[5] J. Fürnkranz and E. Hüllermeier. Preference learning. *Künstliche Intelligenz*, 19(1):60–61, 2005.

[6] J. Goldberger, S. Challapalli, R. Tung, M. Parker, and A. Kadish. Relationship of heart rate variability to parasympathetic effect. *Circulation*, 103(15):1977, 2001.

[7] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[8] N. Lesh, M. Zaki, and M. Ogihara. Mining features for sequence classification. In *Proc. of the fifth ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 342–346. ACM, 1999.

[9] R. Mandryk, K. Inkpen, and T. Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25(2):141–158, 2006.

[10] R.L. Mandryk, and M.S Atkins. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, 65(4):329–347, 2007.

[11] H. P. Martinez, K. Hullett, and G. N. Yannakakis. Extending Neuro-evolution Preference Learning through Player Modeling. In *Proc. of 2010 IEEE CIG*, 2010.

[12] H. P. Martínez and G. N. Yannakakis. Genetic search feature selection for affective modeling: a case study on reported preferences. In *Proc. of the 3rd AFFINE workshop*, pages 15–20. ACM, 2010.

[13] S. Mcquiggan, B. Mott, and J. Lester. Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *UMUAI*, 18(1):81–123, 2008.

[14] M. Pantic and L. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, 2003.

[15] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *2005 IEEE International Conference on Multimedia and Expo*, page 5. IEEE, 2005.

[16] N. Ravaja, T. Saari, J. Laarni, K. Kallinen, M. Salminen, J. Holopainen, and A. Järvinen. The psychophysiology of video gaming: Phasic emotional responses to game events. In *Proc. of DiGRA'05*, 2005

[17] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen. Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology*, 8(4):343–367, 2006.

[18] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. *Advances in Database Technology - EDBT'96*, pages 1–17, 1996.

[19] G. N. Yannakakis. Preference Learning for Affective Modeling. In *Proc. of ACII'09*, 2009.

[20] G. N. Yannakakis, H. P. Martínez, and A. Jhala. Towards affective camera control in games. *UMUAI*, 20:313–340, 2010.