Journal of Biomedical Informatics 81 (2018) 74–82

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Incorporating repeating temporal association rules in Naïve Bayes classifiers for coronary heart disease diagnosis

Kalia Orphanou[a,*], Arianna Dagliati[c,f,g], Lucia Sacchi[b,c], Athena Stassopoulou[e], Elpida Keravnou[a], Riccardo Bellazzi[b,c,d]

[a] Department of Computer Science, University of Cyprus, Cyprus
[b] Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy
[c] Centre for Health Technologies, University of Pavia, Italy
[d] IRCCS Istituti Clinici Scientifici Maugeri, Pavia, Italy
[e] Department of Computer Science, University of Nicosia, Cyprus
[f] Centre for Health Informatics, University of Manchester, Manchester, United Kingdom
[g] The Manchester Molecular Pathology Innovation Centre, University of Manchester, Manchester, United Kingdom

ARTICLE INFO

Keywords:
Bayesian models
Time series classification
Temporal abstraction
Temporal reasoning
Temporal association rules

ABSTRACT

In this paper, we develop a Naïve Bayes classification model integrated with temporal association rules (TARs). A temporal pattern mining algorithm is used to detect TARs by identifying the most frequent temporal relationships among the derived basic temporal abstractions (TA). We develop and compare three classifiers that use as features the most frequent TARs as follows: (i) representing the most frequent TARs detected within the target class ('Disease = Present'), (ii) representing the most frequent TARs from both classes ('Disease = Present', 'Disease = Absent'), (iii) representing the most frequent TARs, after removing the ones that are low-risk predictors for the disease. These classifiers incorporate the horizontal support of TARs, which defines the number of times that a particular temporal pattern is found in some patient's record, as their features. All of the developed classifiers are applied for diagnosis of coronary heart disease (CHD) using a longitudinal dataset. We compare two ways of feature representation, using horizontal support or the mean duration of each TAR, on a single patient. The results obtained from this comparison show that the horizontal support representation outperforms the mean duration. The main effort of our research is to demonstrate that where long time periods are of significance in some medical domain, such as the CHD domain, the detection of the repeated occurrences of the most frequent TARs can yield better performances. We compared the classifier that uses the horizontal support representation and has the best performance with a Baseline Classifier which uses the binary representation of the most frequent TARs. The results obtained illustrate the comparatively high performance of the classifier representing the horizontal support, over the Baseline Classifier.

1. Introduction

Temporal abstraction (TA) is useful for abstracting time point data into interval-based sequences of events [1]. Temporal abstracted events were shown to be helpful in various clinical tasks and domains such as summarizing and managing patient data in oncology [2], monitoring of children's growth [3], management of insulin-dependent diabetes [4] and interpreting online patient data for monitoring purposes in intensive care units (ICUs) [5]. Bayesian networks (BNs) [6–8] belong to the family of probabilistic models and they were widely used in many clinical domains as they can handle well uncertainty in medical knowledge and data. Both Bayesian models and TAs demonstrated their effectiveness as standalone engines predominantly for medical problem solving and for medical data processing respectively, but not in conjunction. A detailed survey on TAs and BNs approaches applied to clinical domains and the benefits of their integration can be found in [9].

Temporal association rules (TARs) are special types of association rules extracted by applying a temporal operator between the antecedent and the consequent of the rule. TARs characterize the temporal relation between the time-interval events defined in the antecedent and consequent. Both the antecedent and consequent represent temporal abstraction events and the aim of the TARs is to extract complex abstractions that are mined from data in a knowledge-based fashion

* Corresponding author.
E-mail address: kalia.orphanou@gmail.com (K. Orphanou).

https://doi.org/10.1016/j.jbi.2018.03.002
Received 1 March 2017; Received in revised form 14 February 2018; Accepted 7 March 2018
Available online 16 March 2018

[11,10]. Related works [12,13] introduced symbolic time-interval events that extract knowledge-based abstractions such as TIRPs which are discussed in Section 2.

In our previous work [14], we developed a Naïve Bayes classifier [15,16], using as features the most meaningful TARs from patient records and utilize them to improve the classification performance. We applied the developed classifier for predicting a future risk of Coronary Heart Disease (CHD). The notion of horizontal support that was introduced in [12,13] and defines the number of times that a particular temporal pattern is found in some patient's record, was incorporated in the values of the features. As an extension of that work, the focus of this paper is to demonstrate how the classification performance changes due to the feature selection and the feature representation process. In this paper, we introduce two more classifiers where the features represent the most frequent TARs that better discriminate the target class and where the TARs that are not good predictors for the disease, based on the medical knowledge for the specific domain, were excluded from the selected features. We also develop a classifier that uses the mean duration of the TARs as its features and finally we compare the classifier with the best performance, to a Baseline Classifier which uses a binary representation of the TARs.

The rest of the paper is organized as follows: In Section 2, we present an overview of temporal data mining in biomedicine and temporal patterns mining, while in Section 3 we describe the methods used for the development of the classifiers. A discussion of our experiments for selecting the most frequent and predictive TARs for the disease, to be represented as features to the network is given in Section 4. In the same section, we describe the dataset used as a testbed. In Section 5, the obtained results for the evaluation of all the developed classifiers are presented. Finally, a discussion about the results is given in Section 6, while conclusions and future work are presented in Section 7.

2. Background

2.1. Temporal data mining in biomedical data

The medical history of some individual is a repository of time-stamped data/information of diverse format/content and storage medium. Such data are invariably expressed at different levels of semantic detail and sampling frequency, they could have gaps or be excessively voluminous, and they are not amenable to direct processing and reasoning with. The recent advances in technology enabled the collection and storage on electronic platforms (e.g. electronic health records (EHR)) of large volumes of such data, while the development of temporal data mining techniques enabled the analysis, representation, interpretation and reasoning with the EHR longitudinal data [17–20].

A variety of temporal data mining techniques is proposed in the literature to deal with biomedical data, such as phenotyping [18], machine learning techniques such as Gaussian kernel smoothing and differential entropy [21], temporal abstraction over time intervals [1,9,22] and dynamic Bayesian networks [23,24]. The aim of temporal data mining techniques is to induce new temporal knowledge from the time-series in EHR data and to develop accurate classification and predictive models.

2.2. Temporal abstraction and time intervals mining

Temporal Abstraction (TA) refers to a set of techniques that allow describing a set of time series data and external events through sequences of context-specific temporal intervals [1,25]. TAs can be divided into two main categories: basic and complex. Basic TAs take as input time point events and return as output time intervals on the basis of some predefined rules, known as TA mechanisms. The derived symbolic time intervals can then be combined into complex temporal patterns representing their temporal relationships (complex TAs).

Although mining time-intervals is a relatively young research field, many automated tools have already been proposed in the literature to automatically discover frequent temporal patterns derived as the conjunction of temporal relations between pairs of intervals. Most of the methods use a subset or all of Allen's 13 interval relations to discover the temporal relationships among events. Kam and Fu [26] were the first who proposed the discovery of temporal patterns using all of Allen's interval relations, however their discovered patterns were ambiguous, since only the temporal relations among all the pairs of consecutive intervals were defined. Following, Hoppner [27] resolves this issue, by defining a non-ambiguous representation where all the possible pairwise temporal relations are represented in a *k*-intervals pattern.

The work proposed by Batal et al. [28] follows the approach by Hoppner by deriving temporal patterns using a subset of Allen's temporal operators. They detect temporal patterns based on the sliding window method. This algorithm is based on the assumption that events occurring far enough from each other, have no temporal relationship. Adam et al. [29,30] proposed 'Frequence', a web-based interface that integrates a data mining algorithm with a visualization tool. The interface aims at discovering frequent patterns from temporal event sequences and then to present the patterns mined in a user friendly way. The 'Frequence' system considers the sequence and the duration of temporal events, but it does not explicitly use any temporal relation to mine the frequent patterns.

In [31], a fast symbolic time intervals mining algorithm, KarmaLego, is presented to mine Time Intervals Related Patterns (TIRPs). In that work, KarmaLego is included in a process that implements a knowledge based temporal abstraction (KBTA) [13] framework for deriving basic TAs and then the algorithm is iteratively applied to the derived abstractions to detect TIRPs. In addition, other methods have been proposed that do not use Allen's temporal relations [29,30,32].

In the current work, we use the method proposed by Sacchi et al. [11] to extract temporal patterns as a set of TARs. As mentioned in the introduction, TARs are a special type of association rules extracted by applying a temporal operator between the antecedent and the consequent of the rule. In a TAR, the members of the antecedent are characterized by a co-occurrence of the temporal patterns that compose it. In the current work, we only use the *Precedes* operator which synthesizes the *Before*, *Meets*, *Overlaps*, *Equal*, *Starts* and *Finished-by* temporal relations, and we only extract TARs consisting of two symbolic time intervals. From the related work, TIRPs are the most similar technique to the one we are presenting. Differently from TARs, in TIRPs mining the temporal operator is applied among each interval that builds up the pattern. According to its definition, any TAR involves only two elements (an antecedent and a consequent), and as such only one temporal relation is considered between the two, albeit a disjunctive one. In addition, while TIRPs are mined by using as input an interval based representation, to mine the TARs herein proposed, raw time series go through a temporal abstraction step that results in the definition of arbitrarily complex abstractions that are extracted from data in a knowledge-based fashion. Thanks to this procedure, the extracted TARs provide more compact patterns, where the necessary complexity is included in the temporal abstractions that make up the antecedent and the consequent, rather than in the temporal operator that links them. The idea of defining a set of complex abstractions of interest that are then used to create the antecedents and the consequents of the TARs allows pursuing a knowledge-based approach, where the users can define the patterns on the basis of what they have in mind and interpret the final results in terms of such knowledge.

2.3. Time intervals related patterns based classification

The use of the discovered temporal patterns as features for classification is becoming increasingly popular in data mining literature. The classification task has been performed by using different methods such as decision trees [33], Naïve Bayes [12,28,34], recurrent neural

networks [35] and random forest [12]. Patterns extracted by using temporal relations have been included as features for classification in [13,36–39]. Batal et al. [28] used two temporal relations, *Before* and *Co-occurs* to discover the most recent and frequent patterns and they proposed an Apriori-like mining algorithm, called STF-Mine. The aim of this approach is to find the most predictive temporal patterns based on the fact that the most recent ones are the most predictive and then to use them as features to the classifier.

Most recent approaches for using time interval related patterns to classify multivariate time series data were proposed in [33,35]. Nancya et al. [33] propose a new approach which handles the irregularity of time series data during the mining process and uses a decision tree as a classification method, while Che et al.'s [35] approach handles missing data and uses a recurrent neural network to classify the mined patterns. Another recent approach was proposed by Guo et al. [40] to combine temporal pattern mining with feature selection to identify temporal risk factors that can predict the acute ischemic stroke, as well as temporal treatment patterns.

Patel et al. [36] proposed a feature selection method, GAIN, for classifying multivariate data. GAIN is an entropy-based measure which defines that temporal patterns that occur in only one class are more discriminative. Moskovitch et al. [32] also present the detection of the most frequent temporal patterns in one class which usually is the target class. In this work, we also validate this assumption for our benchmark dataset.

The utilization of knowledge-based temporal abstractions to detect temporal relation patterns such as frequent time interval patterns, and to incorporate them as features for classification was recently proposed in [12,28,32,34,41–43].

3. Methods

TAs are divided into two main categories: basic and complex. Basic TAs are used to extract simple patterns in a time series, such as time intervals where a state or trend can be detected. Complex TAs represent the temporal relationship between the extracted basic or other complex TAs using temporal operators [44], e.g. "Diabetes *Before* Hypertension".

3.1. Basic TAs

In the current study, we use two different types of basic TA algorithms: state and trend TAs. State abstractions determine the state of a parameter over a time period. Each state is defined on the basis of thresholds selected by clinical experts. For instance, the value of Total cholesterol is defined as v = Normal at a specific time point t, when its raw value (rv) at t is $rv < 200$ mg, v = Borderline high when $200 < rv < 260$ mg, otherwise v = High. On the other hand, trend abstractions detect decreasing, increasing or stationary patterns in numerical time series, such as the value of Total cholesterol when its raw value at t_0 is 200 mg and at t_1 is 260 mg, is v = Increasing from $[t_0,t_1]$.

To extract symbolic time intervals from time series raw data, in the current study, we use the Java Time Series Abstractor (JTSA) [45] software. JTSA is a standalone application for the definition and execution of a complete time series analysis workflow to detect temporal patterns. The framework incorporates an algorithm taxonomy that includes both algorithms for time-series preprocessing and algorithms for symbolic time intervals detection. The resulting symbolic time intervals are defined by a pair: interval of occurrence and a label for the pattern. The following parameters should be defined before using the framework: granularity of the data, minimum trend slope for trend abstractions detection (or the thresholds for state TAs), the minimum length of a pattern and the maximum gap between consecutive time points to keep them in the same pattern.

Aggregation TA is a category of TA algorithms used in the JTSA tool [45] to merge consecutive time intervals. The intervals merged by an aggregation algorithm can either have the same label or two different labels, which we want in the end to have the same value (for example 'High' and 'Very High' to be simply called 'Out of Range'). To perform intervals aggregation, Aggregation TA algorithms take into consideration a specific parameter, the maximum gap. For instance, if the value of 'Total cholesterol' is $v = High$ at time t_0 and time t_1, (with $t_1 > t_0$) and $(t_1 - t_0) \leqslant maximumgap$ then the resulting aggregation TA is a symbolic time interval characterized by $I = [t_0,t_1]$, $L = High$, where I is the interval of occurrence of the time-interval and L is the label associated with the temporal pattern and *maximum gap*, is a user-defined parameter. For example, if the granularity of the dataset is MONTHS and maximum gap is set to 12, it means that two consecutive high values that are collected within a year can be merged in the same interval after ($I = [t_0,t_1]$, $L = High$).

3.2. Temporal association rules

According to [10], a TAR can be viewed as a temporal pattern that describes an association between two events, the antecedent and the consequent. Differently from traditional association rules [46,47], the relationship that holds between the antecedent and the consequent of a TAR is expressed through a temporal operator (e.g. *Before*). Frequent TARs can be mined from data through an algorithm that relies on the notions of confidence and support [11]. The support is defined as the proportion of cases verifying the TAR (SR) over the total number of cases involved in the study (S) (Eq. (1)) while the confidence is defined as the ratio between the support of a TAR ($A \rightarrow C$) and the support of its antecedent (A) as defined in Eq. (2). More specifically, the confidence represents the conditional probability for a case to verify the TAR given that the antecedent is detected for that case.

$$Support = SR/S \tag{1}$$

$$Confidence = Support(A \rightarrow C)/Support(A) \tag{2}$$

TARs are automatically extracted using a miner software tool [10]. In this paper, we apply the TARs miner software tool, to discover frequent TARs (support and confidence > 0.9) among the basic TA episodes on the basis of the given temporal relationship. We exclusively use the temporal operator *Precedes* as defined in [11,48] and illustrated in Fig. 1. To compute support in the TARs mining phase, we refer to the definition given in Concaro et al. [10], where the number of subjects that verify a specific TAR (SR) is computed as the number of subjects who satisfy TARs that meet a minimum duration threshold ($span_{th}$), added to the number of subjects who satisfy TARs that meet a minimum frequency threshold (f_{th}). In this way, the symbolic time interval events supporting a rule can be either long-lasting symbolic interval events with low frequency or highly frequent short symbolic interval events.

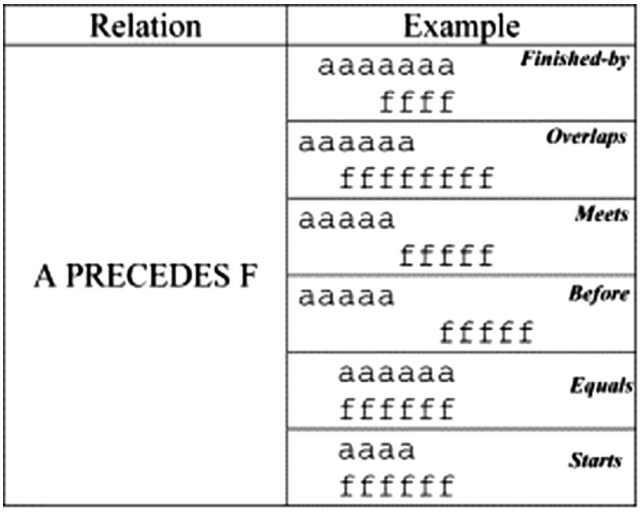

Fig. 1. *Precedes* temporal operator.

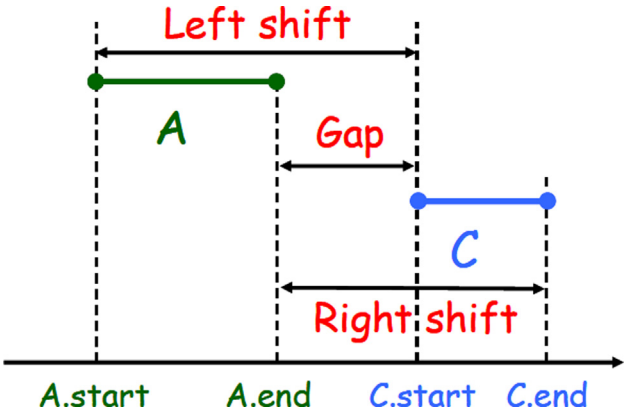

Fig. 2. Representation of the three temporal constraints (left shift, gap, right shift) used in the TARs mining tool, to constrain the mutual distances between the antecedent (A) and the consequent (C).

Moreover, the notion of a frequent rule includes both an evaluation of the rule along the temporal dimension of the sequences ($f > f_{th}$ or $span > span_{th}$) and an evaluation of the frequency of the rule on the population, given by the proportion of subjects supporting the rule (support > minsup). In this work, we have set $f_{th} = span_{th} = 1$, thus uniforming our definition of support to the definition of horizontal support given by Moskovitch et al. [12,13].

Besides the temporal operators that define the temporal relationships among the TAs, the software tool also uses three temporal constraints (*left shift, right shift, gap*) to properly control the mutual distance between the antecedent (A) verified on the interval $IA = [A.start, A.end]$ and the consequent (C) of a TAR verified on the interval $IC = [C.start, C.end]$ [11]. The parameter *left shift*, is defined as the maximum allowed distance between the start time of the antecedent ($A.start$) and the start time of the consequent ($C.start$). The *right shift*, is defined as the maximum allowed distance between the end time of the antecedent ($A.end$) and the end time of the consequent ($C.end$). The *gap* is similarly defined as the maximum allowed distance between $A.end$ and $C.start$. Using any temporal operators apart from 'MEETS' and 'BEFORE', the *gap* is a negative number since $C.start < A.end$. These three parameters can be conveniently used to constrain the mutual position of the operands in the relation, being careful to comply with the specific characteristics of the temporal operator that is being used, as detailed in [11]. Fig. 2 graphically displays the meaning of these three temporal constraints.

3.3. Naïve Bayes classifier

As a classification method, we use the Naïve Bayes network [15,16] which is a simple Bayesian network with the "naïve" independence assumption that the effect of an attribute value on a given class is independent of the values of the other attributes. It is one of the most widely used classifiers with many applications in medical expert systems for different clinical domains [49–51].

The parameters of the classifier are learned from data using the maximum likelihood algorithm [52]. Once the network structure is defined and the network is quantified with the learned parameters, the next step is to predict the probability of the class variable. Each feature in the network is instantiated with the corresponding feature value. Then the model derives the belief:

$$P(Disease|tar1, tar2, tar3, ..., tarN)$$

As shown in Fig. 3,[1] in this paper we use the disease (CHD event) as the class variable, whereas all the TARs mined on the dataset are used as attributes and connected to the class. The *Disease* is a binary node taking the values: (i) 0: Absence of CHD event and (ii) 1: Presence of CHD event. The goal is to classify the presence or absence of a CHD event, given the repeated occurrences of each TAR in the analyzed history.

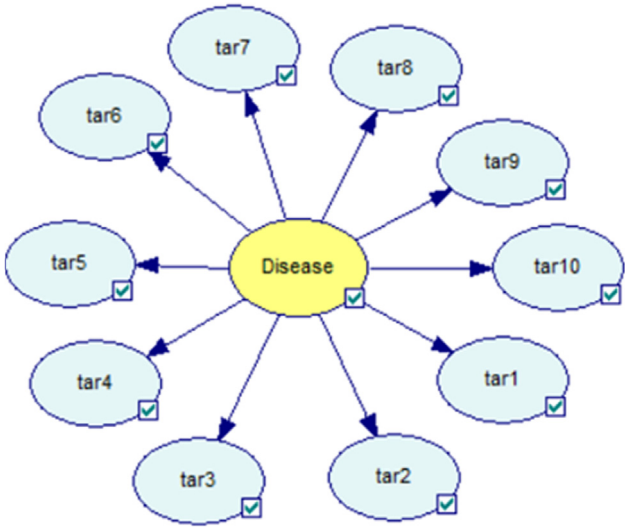

Fig. 3. Naïve Bayes structure representing TARs as features.

3.4. Feature selection and representation methods

In the current study, we use feature selection to explore possible improvement in the classification performance. In particular, we use two ways of selecting the most frequent and the most predictive TARs to the disease, which are then incorporated as features to the Naïve Bayes classifiers. The two feature selection strategies are better detailed in the following:

1. **TARs Selection A**: Selection of the most frequent TARs, based on the support and confidence values, which are detected only on the set of patients who have been diagnosed with the disease (and not for the patients who did not experience the disease). The aim is to discover frequent TARs that better discriminate the target class: "Disease = Present".
2. **TARs Selection B**: Detection of the frequent TARs on the target class, "Disease = Present" and then we remove the ones which are not good predictors for the disease. The aim is to compare the performance of the classifier in the presence or absence of TARs that define the precedence of two decreasing variables such as "Total cholesterol decreasing *Precedes* Triglycerides decreasing". As it is known from medical literature [53,54], such associations are not considered good predictors for the CHD because they represent risk factors with low impact on the disease.

Another aim of this study is to investigate how the feature representation method can affect the classification performance. Thus, we compare the use of the horizontal support of the TARs against the use of the mean duration which defines the average duration of each TAR within a single patient entity. In order to test our claim that for medical domains where long time periods are significant, considering the horizontal support of the TARs can yield better performance, we develop another Naïve Bayes classifier whose features are binary TARs, expressing the occurrence or not of the same set of TARs.

[1] The models presented in this paper were created and tested using the GeNIe Modeler available at: http://www.bayesfusion.com/ [Date accessed: 30 September 2017].

4. Evaluation

4.1. Research questions

There are three main research questions that we aim to validate through this paper. The first two refer to the feature selection method whereas the last one refers to the feature representation method used for modeling the TARs into the classifiers.

- **Research Question 1:** Will the classification performance be improved by selecting as features only the most frequent TARs that better discriminate the target class? In that case the target class represents the patients who have been diagnosed with the disease.
- **Research Question 2:** Will the classification performance be improved by selecting as features the most frequent TARs after removing the ones which are not good predictors for the disease, based on medical knowledge?
- **Research Question 3:**Which feature representation method is better? A binary representation, representing the existence or not of TARs, the horizontal support or the mean duration of the TARs?

4.2. Experiments plan

Regarding the first type of TARs selection (**TARs Selection A**), we develop two classifiers where the features are selected as follows:

- **Classifier** *top*10*freq*−*allClasses*: We select the 10 most frequent TARs on the entire dataset. The selected TARs are displayed in Table 4.
- **Classifier** *top*10*freq*−*targetClass*: We select the 10 most frequent TARs on the target class: "Disease = Present", as displayed in Table 5.

From both of the classifiers, we select the 10 most frequent TARs, in order to avoid overfitting by selecting a large number of features but also to have a satisfying number of features for the classification. Regarding the second type of TARs selection (**TARs selection B**), we developed another classifier by excluding those TARs that, on the basis of the clinical knowledge, are known to be not good predictors of the disease:

- **Classifier** *top*10*freq*−*targetClassPred*: We select the same TARs as on the classifier *top*10*freq*−*targetClass*. Out of these, we then select the most predictive ones for the disease by excluding those TARs that define the precedence of two decreasing variables (or HDL increasing, since HDL is considered as a risk factor when its levels are low). The resulting number of TARs is nine and they are displayed in Table 6.

To assign a value to each of the selected features for each patient, the horizontal support of each TAR is used. More specifically, discretization techniques are used for categorizing the horizontal support of each TAR [55]. One well-known measure which characterizes the purity of the class membership of different variable states is entropy [56]. The number and range of values which result in the minimum total weighted entropy are chosen to quantize each variable. This minimum entropy principle is applied on all the variables (nodes) of the network, as illustrated in the following example.

For classifier *top*10*freq*−*allClasses*, for instance, for all the TARs *tar*1−*tar*10, the discretization resulted in four values, corresponding to their potential horizontal support, as follows:

- 0: the TAR does not occur
- 1: the TAR occurs once or twice
- 2: the TAR recurs between three and four times
- 3: the TAR recurs at least five times

Regarding the mean duration feature representation, another classifier has been developed, *top*10*freq*−*targetClassPredMD*. Since the values of the mean duration of each TAR vary from 1 to 7 years, the minimum entropy principle is applied on all the variables in the network, on a similar way as for the horizontal support and the discretization resulted in four values.

Following, we compare the performance of the classifier with the highest classification performance, *top*10*freq*−*targetClassPred*, against that of the Baseline Classifier, that uses a binary representation of the TARs. All of the classifiers have the same network structure (see Fig. 3), however, the nodes of the Baseline Classifier are binary taking the following two values:

- 0: The TAR does not occur at all in the relevant patient history
- 1: The TAR occurs at least once in the relevant patient history (but the recurrence pattern is not categorized)

The goal of both classifiers is to predict the value of the class variable.

TARs are detected and selected as features from the whole dataset, according to the feature selection and representation method. The 10-fold cross validation is then applied to the training of the classification models on the data instances representing the selected TARs.

4.3. Dataset overview

In the current study, we evaluate the performance of the framework on CHD diagnosis. CHD occurs when atherosclerosis affects the arteries of the heart, and is a disease that leads to a large number of deaths worldwide [57]. The benchmark dataset used for the evaluation of the framework is the STULONG[2] which was collected from a longitudinal study of atherosclerosis primary prevention.

The dataset includes 1427 male patients who were 38–53 years old, at their first examination. The number of visits of a single patient ranges from 1 to 20 and the follow-up time spans from 1 to 24 years. The first patient visit included blood pressure measuring, basic anthropometric measurements (e.g. weight and height) and an electrocardiogram test (ECG). Furthermore, patients were asked about their level of education and responsibility in job, their general habits such as smoking, physical activity, alcohol drinking as well as family and personal medical history related to cardiovascular diseases, chest and legs pain, and breathlessness.

4.3.1. Data preprocessing and feature selection

In medical datasets, usually, the most recent patterns are the most significant ones, since the temporal observations that are close to the time of the disease event are typically the most important for prediction [58,59]. On the basis of this assumptions and relevant medical literature, we detected TARs in a time window of 10 year before the last observation of the patient [58,59]. For each patient, we consider the last observation to be the one before the first diagnosis of the disease (for patients who were diagnosed) or the last visit (for patients who were not diagnosed with CHD).

In addition, the derivation of basic TAs would require a time period of at least two observations, i.e. at least two years, while the derivation of complex TAs would require even longer time periods than the basic TAs. As such, the selected dataset is further reduced by removing records of patients who had less than three observations, i.e. that spanned less than three years. The resulting target group consists of 709 patients, out of which 154 were diagnosed with the disease.

The data are characterized with missing values and in order to impute them, we use the *missforest* method [60], a non-parametric

[2] The data resource is available at: http://euromise.vse.cz/challenge2004/ [Date accessed: 30 September 2017].

Table 1
Features selected with the minimum and maximum value of the continuous features.

Variable	Min value	Max value
Smoking	Non smoker	Current smoker
Medicines for reducing cholesterol	Taken	Not taken
Medicines for reducing blood pressure	Taken	Not taken
Systolic blood pressure (mmHg)	80	240
Diastolic blood pressure (mmHg)	40	130
Glucose levels (mmol/l)	0.1	51
Family history	Absent	Present
History of coronary heart disease	Absent	Present
Body mass index (kg/m^2)	17	40
Low-density lipoprotein cholesterol (LDL mg)	11	306
Triglycerides (mg)	17	5129
High-density lipoprotein cholesterol (HDL mg)	9	391
Total cholesterol (TCH mg)	102	878
Age (years)	38	62
Diet	Following a diet	Not following a diet
Exercise	Exercising	Not exercising

imputation method based on the random forest algorithm [61].

Regarding the feature selection process, we base our selection of features on the domain knowledge that we acquired from a CHD expert and from the medical literature. The selected features which are CHD risk factors are displayed in Table 1.

4.3.2. Extracted symbolic time intervals

State, trend and aggregation TAs are derived using the JTSA software tool [45], on all the mentioned risk factors for CHD, using a fixed six-months granularity (finest granularity). State TAs along with their determined thresholds are displayed in Table 2 and trend TAs in Table 3. As displayed in Table 3, the sorted list of trend values in ascending order, based on the impact of the variables on the disease is: (i) decreasing, (ii) stationary and (iii) increasing. It should be noted that, contrary to the rest of the variables, HDL is considered a CHD risk factor when its levels are low, thus its sorted list of trend values is: (i) increasing, (ii) stationary and (iii) decreasing. In the current study, aggregation TA algorithms have only been used to merge consecutive time intervals with the same label.

4.3.3. Selected TARs

The selected TARs for the classifiers: *top*10*freq*−*allClasses*, *top*10*freq*−*targetClass* and *top*10*freq*−*targetClassPred* are displayed in Tables 4–6 respectively.

In our testbed dataset, patients who did not suffer a CHD event are the majority class (4:1) [62]. The approach that we follow in this study, for tackling this issue, is to use the undersampling based on clustering (SBC) technique [63] to remove examples from the majority class of the dataset in order to select a balanced sample. The ratio of the whole dataset was 555/154 (around 4:1) whereas the resulting balanced sample ratio is 154/154. The technique is applied to all the classifiers until a balanced subset of the dataset is obtained.

Table 3
Trend TA variables and their labels.

Variable code	Variable name	L = 1	L = 2	L = 3
T1	BMI	Decreasing	Stationary	Increasing
T2	Systolic blood pressure (SBP)	Decreasing	Stationary	Increasing
T3	Diastolic blood pressure (DBP)	Decreasing	Stationary	Increasing
T4	Smoking	Decreasing	Stationary	Increasing
T5	Low-density lipoprotein cholesterol (LDL)	Decreasing	Stationary	Increasing
T6	Triglycerides (TRIG)	Decreasing	Stationary	Increasing
T7	High-density lipoprotein cholesterol (HDL)	Increasing	Stationary	Decreasing
T8	Total cholesterol (TCH)	Decreasing	Stationary	Increasing

5. Results

For the evaluation of the performance of all the developed classifiers, we adopt metrics that are commonly used for imbalanced datasets: precision, recall, the F_1 score, the Matthews correlation coefficient (MCC) and also the area under ROC curve (AUC) [55].

As illustrated in Table 7, the classifier *top*10*freq*−*targetClass* has the best obtained results, compared to *top*10*freq*−*allClasses*, which indicates that by discovering the TARs only on the patients who have been diagnosed with the disease, the classification performance is improved.

Table 8 displays the results of comparing the classifier *top*10*freq*−*targetClass* against the Classifier *top*10*freq*−*targetClassPred*. As displayed in Table 8, the classifier *top*10*freq*−*targetClassPred* has the best obtained results, compared to *top*10*freq*−*allClasses*. This indicates that by excluding those TARs, that define the precedence of two decreasing variables is significant for the diagnosis of CHD, and as a result the injection of medical knowledge influences the improvement of the performance of the classifier.

Table 9 displays the results of comparing the classifier *top*10*freq*−*targetClassPred* against the classifier *top* 10*freq*−*targetClassPredMD*.

As displayed in Table 9, the use of the horizontal support in the feature representation obtained higher classification results rather than the use of the mean duration.

Table 10, displays the results of comparing the classifier *top*10*freq*−*targetClassPred* against the Baseline Classifier.

As displayed in Table 10, the classifier *top*10*freq*−*targetClassPred* has the best obtained results, compared to the Baseline Classifier. The higher performance of the classifier *top*10*freq*−*targetClassPred*,

Table 2
State TA variables and their values.

Variable code	Variable	$v = 1$	$v = 2$
Q1	Medicines for blood pressure	Taken	Not taken
Q2	Total cholesterol (TCH)	Normal ($Q2 < 260$)	High ($Q2 \geqslant 260$)
Q3	High-density lipoprotein cholesterol (HDL)	Low ($Q3 < 40$)	High ($Q3 \geqslant 40$)
Q4	Low-density lipoprotein	Normal ($Q4 < 130$)	High ($Q4 \geqslant 130$)
Q5	Triglycerides	Normal ($Q5 < 199$)	High ($Q5 \geqslant 199$)
Q6	Obesity	Absent ($Q6 \leq 25$)	Present ($Q6 > 25$)
Q7	Age	Young ($Q7 < 45$)	Old ($Q7 \geqslant 45$)
Q8	Diet	Following a diet	Not following a diet
Q9	Exercise	Exercising	Not exercising
Q10	Diabetes	Absent	Present
Q11	Systolic blood pressure (SBP)	Normal ($Q11 < 140$)	Hypertension ($Q11 \geqslant 140$)
Q12	Diastolic blood pressure (DBP)	Normal ($Q12 < 90$)	Hypertension ($Q12 \geqslant 90$)

Table 4
Selected TARs for classifier *top10freq−allClasses*.

TAR Code	Trend TA Code 1	Relation Operator	Trend TA Code 2	Confidence	Support
tar1	HDL = increasing	*Precedes*	Triglycerides = increasing	0.962	0.925
tar2	Triglycerides = increasing	*Precedes*	LDL = decreasing	0.957	0.942
tar3	SBP = decreasing	*Precedes*	LDL = increasing	0.957	0.931
tar4	SBP = increasing	*Precedes*	Total cholesterol = decreasing	0.955	0.925
tar5	HDL = increasing	*Precedes*	LDL = increasing	0.955	0.918
tar6	DBP = decreasing	*Precedes*	SBP = decreasing	0.954	0.911
tar7	SBP = decreasing	*Precedes*	Triglycerides = increasing	0.952	0.927
tar8	SBP = decreasing	*Precedes*	Total cholesterol = increasing	0.952	0.922
tar9	SBP = increasing	*Precedes*	Triglycerides = increasing	0.952	0.922
tar10	HDL = increasing	*Precedes*	Total cholesterol = increasing	0.952	0.915

Table 5
Selected TARs for classifier *top10freq−targetClass*.

TAR Code	Trend TA Code 1	Relation Operator	Trend TA Code 2	Confidence	Support
tar1	HDL = decreasing	*Precedes*	Triglycerides = increasing	0.922	0.979
tar2	HDL = increasing	*Precedes*	LDL = increasing	0.922	0.973
tar3	HDL = increasing	*Precedes*	SBP = increasing	0.922	0.973
tar4	SBP = decreasing	*Precedes*	LDL = increasing	0.928	0.966
tar5	SBP = decreasing	*Precedes*	Triglycerides = increasing	0.928	0.966
tar6	HDL = increasing	*Precedes*	Triglycerides = decreasing	0.916	0.966
tar7	HDL = decreasing	*Precedes*	LDL = decreasing	0.909	0.966
tar8	DBP = decreasing	*Precedes*	Triglycerides = increasing	0.909	0.966
tar9	SBP = increasing	*Precedes*	Triglycerides = increasing	0.948	0.960
tar10	LDL = increasing	*Precedes*	Triglycerides = decreasing	0.942	0.960

Table 6
Selected TARs for classifier *top10freq−targetClassPred*.

TAR Code	Trend TA Code 1	Relation Operator	Trend TA Code 2	Confidence	Support
tar1	HDL = decreasing	*Precedes*	Triglycerides = increasing	0.922	0.979
tar2	HDL = increasing	*Precedes*	LDL = increasing	0.922	0.973
tar3	HDL = increasing	*Precedes*	SBP = increasing	0.922	0.973
tar4	SBP = decreasing	*Precedes*	LDL = increasing	0.928	0.966
tar5	SBP = decreasing	*Precedes*	Triglycerides = increasing	0.928	0.966
tar6	HDL = decreasing	*Precedes*	LDL = decreasing	0.909	0.966
tar7	DBP = decreasing	*Precedes*	Triglycerides = increasing	0.909	0.966
tar8	SBP = increasing	*Precedes*	Triglycerides = increasing	0.948	0.960
tar9	LDL = increasing	*Precedes*	Triglycerides = decreasing	0.942	0.960

Table 7
The performance for the classifiers: *top10freq−targetClass* and *top10freq−allClasses*. The best results are marked in bold.

	Classifier *top10freq−targetClass*	Classifier *top10freq−allClasses*
Precision	0.72	**0.80**
Recall	0.79	**0.80**
F-score	0.75	**0.80**
AUC	0.79	**0.84**
MCC	0.43	**0.60**

Table 8
The performance for the classifier *top10freq−targetClass* and *top10freq−targetClassPred*. The best results are marked in bold.

	Classifier *top10freq−targetClass*	Classifier *top10freq−targetClassPred*
Precision	0.80	**0.81**
Recall	0.80	**0.83**
F-score	0.80	**0.82**
AUC	0.84	**0.86**
MCC	0.60	**0.64**

compared to the Baseline Classifier further supports our belief that the incorporation of the horizontal support of the TARs can improve the classification of CHD. The difference between the performance of the two classifiers is assessed using the McNemar's test [64] ($p \ll 0.01$) and shown to be statistically significant. Thus, it is important for the classifier to detect and consider the recurrence patterns of the discovered TARs, and not just to detect the occurrence or not of a TAR.

6. Discussion

In this paper, we described the incorporation of frequent TARs as features to a Naïve Bayes classifier as an extension of our previous work [14]. We exploited different methods of selecting the most predictive TARs to CHD. TARs are represented as features in the network either using the binary representation, or using the horizontal support, or the mean duration, of each TAR in the patient history, as it was defined in [12,13].

The discovered TARs represent associations that combine symbolic time intervals using exclusively the temporal relation *Precedes*, which is a disjuctive temporal relation of a number of Allen's interval relations. The symbolic time interval events were extracted by applying knowledge-based TA techniques to the raw data, using the JTSA framework [45]. Several data driven discretization techniques were used in related works to extract symbolic time intervals such as EWD and SAX [12,65],

Table 9
The performance for the classifier *top*10*freq*−*targetClassPred* and *top*10*freq*−*targetClassPredMD*. The best results are marked in bold.

	Classifier *top*10*freq*−*targetClassPred*	Classifier *top*10*freq*−*targetClassPredMD*
Precision	**0.81**	0.61
Recall	**0.83**	0.64
F-score	**0.82**	0.63
AUC	**0.86**	0.62
MCC	**0.64**	0.19

Table 10
The performance for the classifier *top*10*freq*−*targetClassPred* and the baseline classifier. The best results are marked in bold.

	Classifier *top*10*freq*−*targetClassPred*	Baseline classifier
Precision	**0.81**	0.75
Recall	**0.83**	0.68
F-score	**0.82**	0.71
AUC	**0.86**	0.80
MCC	**0.64**	0.46

however the focus of this paper was to utilize knowledge-based temporal abstractions for the detection of TARs and to incorporate knowledge-based TARs as features to the classifier.

The selection of the most frequent TARs was based on their support and confidence values while for the selection of the most predictive TARs, we used two different methods. We discovered frequent TARs (a) on each class separately in order to detect the ones that better discriminate the target class which was also tested in related works [32,36] and (b) by removing TARs which have low impact on the CHD and they are not good predictors (i.e. representing the precedence of decreasing variables). The choice of removing this type of abstractions was mainly driven by the medical knowledge on the CHD domain. When extending the methodology to other clinical domains, this choice should be carefully adapted by considering the available knowledge, in order to prevent the exclusion of some interesting patterns.

In the current work, for classification purposes, feature selection (in the form of TARs extraction) is performed on the whole dataset before training the model, which is then validated using 10-fold cross validation. However, according to [12], the classification performance would be more accurate, if the selection of the TARs is repeated over each fold. Considering that this is a limitation of our approach, as a future work, we are going to implement the cross validation by selecting a different subset of the features (TARs) on every run.

Considering the results from our experiments, the detection of the frequent TARs only from the class of interest had higher classification performance than the detection of the frequent TARs from both classes, which indicates that the most predictive TARs for the disease are extracted from the target class (from the patients who have been diagnosed with the disease). In addition, by removing the TARs that define the precedence of two decreasing variables, the performance of the classifier was improved. This points out that by excluding those TARs, that define the precedence of two decreasing variables is significant for the diagnosis of CHD, and as a result the injection of medical knowledge influences the improvement of the performance of the classifier.

In contrast to the results presented in Moskovitch et al. [12], in the present work, the representation of the mean duration of the TARs as features to the classifier, had lower performance rather than the representation of the horizontal support. Therefore, for the comparison with the Baseline Classifier, we used the classifier with the best performance, using the horizontal support representation, *top*10*freq*−*targetClassPred*, which had the best obtained results, compared to the Baseline Classifier. The McNemar's test was also used and the results validate the significance of the difference between the performance of the two classifiers. This further supports our belief that the incorporation of complex TARs can improve the classification of CHD.

7. Conclusions and future work

The proposed classifier is a Naïve Bayes model where its features represent TARs and their recurrence patterns, constructed for the purpose of CHD diagnosis. A strength of our approach is that the developed classifier can work efficiently with irregularly sampled temporal datasets [66,67].

The framework handled successfully the class imbalance problem, in the training and evaluation stages. We compared different methods of selecting and representing the most frequent and predictive TARs, and the performance of the best classifier was compared against that of a Baseline Classifier. Considering the results from these experiments, the detection of the frequent TARs only from the class of interest and by removing the TARs with low impact on the disease, improved the performance of the classifier. Regarding the feature representation, horizontal support outperforms the mean duration. Furthermore, the better results of the classifier representing the horizontal support of the TARs (complex TARs) against the Baseline Classifier, support the claim that the incorporation of complex TARs, as features to the classifier can improve medical problem solving in domains where long time periods are significant.

As future work, we are going to derive all the possible features from the extracted basic TAs and to apply wrapper feature selection methods, running the entire experiment for each wrapper iteration, to select those to be represented in the classifier. In addition, we are going to use a different number of selected TARs as features to compare the performance of the classifiers. Some other factors can be taken into consideration combined with horizontal support and mean duration, such as the time period of the TARs occurrence by assigning a higher impact to the most recent TARs, or the duration of the most recent TAR, or the confidence of each TAR on each patient instance. Methods for handling consecutive events such as the ones proposed in [68] will also be considered for future work. In addition, we plan to compare the proposed framework with the integration of TARs with other classification methods such as decision trees and neural networks. Exploring new medical and other application domains is also a future consideration.

Conflict of interest

No conflict of interest.

References

[1] Y. Shahar, A framework for knowledge-based temporal abstraction, Artif. Intell. 90 (1–2) (1997) 79–133.
[2] Y. Shahar, M. Musen, Résumé: a temporal-abstraction system for patient monitoring, Comput. Biomed. Res. 26 (1993) 255–273.
[3] M.M. Kuilboer, Y. Shahar, D.M. Wilson, M.A. Musen, Knowledge reuse: temporal-abstraction mechanisms for the assessment of children's growth, Proceedings of the Annual Symposium on Computer Application in Medical Care, 1993, pp. 449–453.
[4] Y. Shahar, M. Musen, Knowledge-based temporal abstraction in clinical domains, Artif. Intell. Med. 8 (3) (1996) 267–298.
[5] L. Fugan, VM2: Representing Time-Dependent Relations in a Medical Setting, Ph.D. thesis Stanford University, 1980.
[6] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman, 1988.
[7] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
[8] P. Weber, G. Medina-Oliva, C. Simon, B. Iung, Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas, Eng. Appl. Artif. Intell. 25 (4) (2012) 671–682.
[9] K. Orphanou, A. Stassopoulou, E. Keravnou, Temporal abstraction and temporal Bayesian networks in clinical domains: a survey, Artif. Intell. Med. 60 (3) (2014) 133–149.
[10] S. Concaro, L. Sacchi, C. Cerra, P. Fratino, R. Bellazzi, Mining healthcare data with temporal association rules: improvements and assessment for a practical use, Artificial Intelligence in Medicine, Springer, 2009, pp. 16–25.

[11] L. Sacchi, C. Larizza, C. Combi, R. Bellazzi, Data mining with temporal abstractions: learning rules from time series, Data Min. Knowl. Disc. 15 (2) (2007) 217–247.
[12] R. Moskovitch, Y. Shahar, Classification of multivariate time series via temporal abstraction and time intervals mining, Knowl. Inform. Syst. 45 (1) (2015) 35–74.
[13] R. Moskovitch, Y. Shahar, Medical temporal-knowledge discovery via temporal abstraction, in: AMIA 2009, American Medical Informatics Association Annual Symposium, San Francisco, CA, USA, 2009.
[14] K. Orphanou, A. Dagliati, L. Sacchi, A. Stassopoulou, E. Keravnou, R. Bellazzi, Combining naive Bayes classifiers with temporal association rules for coronary heart disease diagnosis, in: IEEE International Conference on Healthcare Informatics, ICHI 2016, Chicago, IL, USA, October 4–7, 2016, 2016, pp. 81–92.
[15] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[16] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley-Interscience, 2000.
[17] M. Stacey, C. McGregor, Temporal abstraction in intelligent clinical data analysis: a survey, Artif. Intell. Med. 39 (1) (2007) 1–24.
[18] D.J. Albers, N. Elhadad, E. Tabak, A. Perotte, G. Hripcsak, Dynamical phenotyping: Using temporal analysis of clinically collected physiologic data to stratify populations, PLOS ONE 9 (6) (2014) 1–19.
[19] G. Hripcsak, Physics of the Medical Record: Handling Time in Health Record Studies, Springer International Publishing, Cham, 2015 pp. 3–6.
[20] G. Hripcsak, D.J Albers, A. Perotte, Parameterizing time in electronic health record studies, vol. 22, 2015.
[21] A. Perotte, G. Hripcsak, Temporal properties of diagnosis code time series in aggregate, IEEE J. Biomed. Health Inform. 17 (2) (2013) 477–483.
[22] M. Verduijn, L. Sacchi, N. Peek, R. Bellazzi, E. de Jonge, B.A. de Mol, Temporal abstraction for feature extraction: a comparative case study in prediction from intensive care monitoring data, Artif. Intell. Med. (2007) 1–12.
[23] K. Orphanou, A. Stassopoulou, E.T. Keravnou, DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis, IEEE J. Biomed. Health Inform. 20 (3) (2016) 944–952.
[24] K. Murphy, Dynamic Bayesian Networks, Probabilistic Graphical Models, 2002.
[25] C. Larizza, R. Bellazzi, A. Riva, Temporal abstractions for diabetic patients management, Artif. Intell. Med. (1997) 319–330.
[26] P.-s. Kam, A.W.-C. Fu, Discovering temporal patterns for interval-based events, Proceedings of the Second International Conference on Data Warehousing and Knowledge Discovery, DaWaK 2000, Springer-Verlag, London, UK, UK, 2000, pp. 317–326.
[27] F. Höppner, Learning temporal rules from state sequences, in: IJCAI Workshop on Learning from Temporal and Spatial Data, vol. 25, 2001.
[28] I. Batal, G.F. Cooper, D. Fradkin, J. Harrison, F. Moerchen, M. Hauskrecht, An efficient pattern mining approach for event detection in multivariate temporal data, Knowl. Inform. Syst. 46 (1) (2016) 115–150.
[29] A. Perer, F. Wang, Frequence: interactive mining and visualization of temporal frequent event sequences, Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI '14, ACM, New York, NY, USA, 2014, pp. 153–162.
[30] A. Perer, F. Wang, J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics, J. Biomed. Inform. 56 (Supplement C) (2015) 369–378.
[31] R. Moskovitch, Y. Shahar, Fast time intervals mining using the transitivity of temporal relations, Knowl. Inform. Syst. 42 (1) (2015) 21–48.
[32] R. Moskovitch, H. Choi, G. Hripcsak, N.P. Tatonetti, Prognosis of clinical outcomes with temporal patterns and experiences with one class feature selection, IEEE/ACM Trans. Comput. Biol. Bioinform. 14 (3) (2017) 555–563.
[33] J.Y. Nancy, N.H. Khanna, A. Kannan, A bio-statistical mining approach for classifying multivariate clinical time series data observed at irregular intervals, Exp. Syst. Appl. 78 (Supplement C) (2017) 283–300.
[34] I. Batal, H. Valizadegan, G.F. Cooper, M. Hauskrecht, A temporal pattern mining approach for classifying electronic health record data, ACM Trans. Intell. Syst. Technol. 4 (4) (2013) 63:1–63:22.
[35] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent Neural Networks for Multivariate Time Series with Missing Values, 2016.
[36] D. Patel, W. Hsu, M.L. Lee, Mining relationships among interval-based events for classification, Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD '08, ACM, New York, NY, USA, 2008, pp. 393–404.
[37] R. Moskovitch, F. Polubriaginof, A. Weiss, P. Ryan, N. Tatonetti, Procedure prediction from symbolic electronic health records via time intervals analytics, J. Biomed. Inform. 75 (C) (2017) 70–82.
[38] A. Shknevsky, Y. Shahar, R. Moskovitch, Consistent discovery of frequent interval-based temporal patterns in chronic patients data, J. Biomed. Inform. 75 (2017) 83–95.
[39] N. Nissim, Y. Shahar, Y. Elovici, G. Hripcsak, R. Moskovitch, Inter-labeler and intra-labeler variability of condition severity classification models using active and passive learning methods, Artif. Intell. Med. (AIME 2017) 81 (2017) 12–32.
[40] S. Guo, X. Li, H. Liu, P. Zhang, X. Du, G. Xie, F. Wang, Integrating temporal pattern mining in ischemic stroke prediction and treatment pathway discovery for atrial fibrillation, in: AMIA Summit on Clinical Research Informatics (CRI), 2017.
[41] I. Batal, L. Sacchi, R. Bellazzi, M. Hauskrecht, A temporal abstraction framework for classifying clinical temporal data, AMIA Annual Symposium Proceedings, vol. 2009, American Medical Informatics Association, 2009.
[42] R. Moskovitch, F. Polubriaginof, A. Weiss, P. Ryan, N. Tatonetti, Procedure prediction from symbolic electronic health records via time intervals analytics, J. Biomed. Inform. (2017).
[43] R. Moskovitch, C. Walsh, F. Wang, G. Hripcsak, N. Tatonetti, Outcomes prediction via time intervals related patterns, in: IEEE International Conference on Data Mining, 2015, pp. 919–924.
[44] H.-J. Brckert, B. Nebel, Reasoning about temporal relations: a maximal tractable subclass of Allen's interval algebra, JACM 42 (1) (1995) 43–66.
[45] L. Sacchi, D. Capozzi, R. Bellazzi, C. Larizza, JTSA: an open source framework for time series abstractions, Comput. Meth. Programs Biomed. 121 (3) (2015) 175–188.
[46] P.-N. Tan, et al., Introduction to Data Mining, Pearson Education India, 2006.
[47] G. Williams, Association Analysis, Springer, New York, 2011 pp. 193–203.
[48] R. Bellazzi, C. Larizza, P. Magni, R. Bellazzi, Temporal data mining for the quality assessment of hemodialysis services, Artif. Intell. Med. 34 (1) (2005) 25–39.
[49] B. Karlik, E. Öztoprak, Personalized cancer treatment by using naive Bayes classifier, Int. J. Mach. Learn. Comput. 2 (3) (2012) 339.
[50] G.F. Cooper, C.F. Aliferis, R. Ambrosino, J.M. Aronis, B.G. Buchanan, R. Caruana, M.J. Fine, C. Glymour, G.J. Gordon, B.H. Hanusa, J.E. Janosky, C. Meek, T.M. Mitchell, T.S. Richardson, P. Spirtes, An evaluation of machine-learning methods for predicting pneumonia mortality, Artif. Intell. Med. 9 (2) (1997) 107–138.
[51] P. Miasnikof, V. Giannakeas, M. Gomes, L. Aleksandrowicz, A.Y. Shestopaloff, D. Alam, S. Tollman, A. Samarikhalaj, P. Jha, Naive Bayes classifiers for verbal autopsies: comparison to physician-based classification for 21,000 child and adult deaths, BMC Med. 13 (1) (2015) 1–9.
[52] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Mach. Learn. 29 (2–3) (1997) 103–130.
[53] G. De Backer, J. Besseling, J. Chapman, G.K. Hovingh, J.J. Kastelein, K. Kotseva, K. Ray, Ž. Reiner, D. Wood, D. De Bacquer, et al., Prevalence and management of familial hypercholesterolaemia in coronary patients: an analysis of EUROASPIRE IV, a study of the European Society of Cardiology, Atherosclerosis 241 (1) (2015) 169–175.
[54] S.M. Grundy, G.J. Balady, M.H.e.a. Criqui, Primary prevention of coronary heart disease: guidance from Framingham, Circulation 97 (18) (1998) 1876–1887.
[55] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, first ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.
[56] T.M. Mitchell, Machine Learning, McGraw-Hill Series in Computer Science, McGraw-Hill, Boston (Mass.), Burr Ridge (Ill.), Dubuque (Iowa), 1997.
[57] P.A. McCullough, Coronary artery disease, Clin. J. Am. Soc. Nephrol. 2 (3) (2007) 611–616.
[58] D.M. Lloyd-Jones, P.W. Wilson, M.G. Larson, A. Beiser, E.P. Leip, R.B. D'Agostino, D. Levy, Framingham risk score and prediction of lifetime risk for coronary heart disease, Am. J. Cardiol. 94 (1) (2004) 20–24.
[59] J. Hippisley-Cox, C. Coupland, J. Robson, P. Brindle, Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: Cohort study using QResearch database, BMJ 341.
[60] D.J. Stekhoven, P. Buehlmann, Missforest - non-parametric missing value imputation for mixed-type data, Bioinformatics 28 (1) (2012) 112–118.
[61] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[62] G.E. Batista, R.C. Prati, M.C. Monard, Balancing strategies and class overlapping, Advances in Intelligent Data Analysis VI, Springer, 2005, pp. 24–35.
[63] S.-J. Yen, Y.-S. Lee, Cluster-based under-sampling approaches for imbalanced data distributions, Exp. Syst. Appl. 36 (3) (2009) 5718–5727.
[64] Q. McNemar, Note on the sampling error of the difference between correlated proportions or percentages, Psychometrika 12 (2) (1947) 153–157.
[65] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, DMKD '03, ACM, New York, NY, USA, 2003, pp. 2–11.
[66] U. Nodelman, C. Shelton, D. Koller, Continuous time Bayesian networks, in: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI), 2002, pp. 378–387.
[67] M. Ramati, Irregular Time Markov Models, Ph.D. thesis Ben Gurion University, 2010.
[68] M.-R. Shknevsky, A.Y. Shahar, Semantic considerations in time intervals mining, in: The notes of BigCHat: The KDD 2014 Workshop on Connected Health at Big Data Era, held as part of The 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2014.