Published in final edited form as:

IEEE/ACM Trans Comput Biol Bioinform. 2017 ; 14(3): 555–563. doi:10.1109/TCBB.2016.2591539.

Prognosis of Clinical Outcomes with Temporal Patterns and Experiences with One Class Feature Selection

Robert Moskovitch,
Department of Biomedical Inforamtics at Columbia University

Hyunmi Choi,
Department of Neurology at Columbia University

George Hripsack, and
Department of Biomedical Inforamtics at Columbia University

Nicholas Tatonetti
Department of Biomedical Inforamtics at Columbia University

Abstract

Accurate prognosis of outcome events, such as clinical procedures or disease diagnosis, is central in medicine. The emergence of longitudinal clinical data, like the Electronic Health Records (EHR), represents an opportunity to develop automated methods for predicting patient outcomes. However, these data are highly dimensional and very sparse, complicating the application of predictive modeling techniques. Further, their temporal nature is not fully exploited by current methods, and temporal abstraction was recently used which results with symbolic time intervals represetnation. We present Maitreya, a framework for the prediction of outcome events that leverages these symbolic time intervals. Using Maitreya, learn predictive models based on the temporal patterns in the clinical records that are prognostic markers and use these markers to train predictive models for eight clinical procedures. In order to decrease the number of patterns that are used as features we propose the use of three one class feature selection methdos. We evaluate the performance of Maitreya under several parameter settings, including the one-class feature selection, and compare our results to that of atemporal approaches. In general, we found that the use of temporal patterns outperformed the atemporal methods, when representing the number of patterns occurrences.

Keywords

Time Intervals Mining; Temporal Patterns; Prediction

1 Introduction

Prognosis is central to the practice of medicine. Accurate forecasting of a patient's disease state will help to identify the best course of treatment, mitigate side effects, ameliorate symptoms, and provide the patient with the best opportunity of recovery – effectively reducing morbidity and mortality. On the contrary, inaccurate forecasting may lead to mismanagement of the disease, increased pain and suffering, and increased costs to the healthcare system. The availability of electronic health record systems represents a novel opportunity to model disease prognosis. The longitudinal nature of the data enables the use of temporal modeling strategies to follow and predict patient trajectories. In order to develop accurate prognostic models we must first identify historical patterns that are indicative of disease. The burgeoning field of temporal data mining has laid the ground for discovering these patterns.

Temporal data mining is a sub-field of data mining, applied to time-oriented data to discover temporal knowledge. Unlike typical data mining methods, which often ignore the temporal dimension or use only concise statistical abstractions of it, temporal data mining (Esling and Agon, 2012) faces additional computational and methodological challenges.

The data in medical record systems, however, exhibits unique characteristics that challenge these algorithms. Namely, they contain mainly thousands of variables and are largely missing (Hripcsak and Albers, 2013). Because of these challenges temporal data mining has focused on relatively small, complete data sets and prognostic modeling in the EHR typically ignores time (Jensen, 2012). In order to address these challenges we implement efficient time intervals mining and patterns detection algorithms.

Mining symbolic time-intervals is a subfield of data mining and has emerged over the last fifteen years. These methods typically use a subset of Allen's temporal relations (Allen, 1983). In their most expressive and complete form, these relations are computationally demanding due to the large combinations of symbols and temporal relations in potential patterns (Villafane et al., 2000; Hoeppner, 2001; Papapetrou et al., 2009; Winarko and Roddick, 2007; Wu and Chen, 2007; Pei et al, 2001; Patel et al, 2008; Wu and Chen, 2007; Papapetrou et al, 2009). Moskovitch and Shahar (2015a) introduced KarmaLego that introduces an efficient data structure and exploits the transitivity property of the temporal relations to generate only realistic candidates. KarmaLego was faster than IE-Miner (Patel et al, 2008), ARMADA (Winarko and Roddick, 2007), and H-DFS (Papapetrou et al, 2009) methods. Unlike earlier methods, KarmaLego discovers the complete set of Time Intervals Related Patterns (TIRPs) (Moskovitch and Shahar, 2015b).

In this study we use only symbolic elements in the data, specifically conditions, procedures and drug exposures. Conditions are clinical diagnoses that are extracted from the physician's notes for billing purposes. Procedures are protocols (e.g. surgeries) that are executed to treat conditions. Drug exposures are prescription medications used to treat the condition or ameliorate side effects. We then use the KarmaLego algorithm (Moskovitch and Shahar, 2015a) for the discovery of frequent TIRPs from symbolic time intervals. Later to detect these TIRPs for a given patient in order to perform classification or prediction, we use SingleKarmaLego (Moskovitch and Shahar, 2015b), a highly efficient TIRPs detection algorithm that first indexes the time intervals to make the detection faster. Finally we introduce Maitreya, a framework for the prediction of outcome events that consists on KarmaLego and SingleKarmaLego, and demonstrate it on several Electronic Health Records (EHRs) datasets.

Maitreya discovers TIRPs using KarmaLego from a cohort of patients with the outcome event, and trains a predictive model learning only on this cohort. A control group of patients without the outcome event is selected for evaluation purposes. Due to that scheme we propose in this study a method for one class feature selection based on two strategies: those that are less correlated with other TIRP features; and those that increase the homogeneity of the cohort. Finally, we demonstrate the use of Maitreya on eight cohorts of patients derived from the New York-Presbyterian/Columbia University Medical Center Clinical Data Warehouse. We compare Maitreya to atemporal methods (i.e. ignoring the temporal order of the symbols) and also evaluate multiple TIRP representation methods. We found that Maitreya outperforms methods that do not account for temporal order of events, and that in some cases, one-class feature selection can be used to improve predictive performance.

2 Background

Multivariate temporal data exist across many scientific disciplines. These data are often characterized by sparse sampling, varying measurements frequencies and types of events, and having variable duration. The use of Symbolic Time Intervals representation, the development of fast and efficient Time Intervals Related Patterns algorithms, and their use for classification and prediction in multivariate time series to analyze these data will be covered in this section.

2.1 Outcome Events Prognosis in EHR

The development of predictive modeling in clinical medicine through data mining is an important and developing field (Bellazi and Zupan, 2008). With the increase in access to EHR data in the recent years there is a growth in studies dealing with the prediction, or prognosis, or outcome events based on patient’s history (Jensen et al, 2012; Ng et al, 2014).

Ng et al. (2014) had introduced a distributed platform for healthcare analytics for EHR data that consists on the Map-Reduce principles and parallelizes the entire process of cohort construction, feature construction and selection and classification in a cross validation fashion. There was no temporal analysis among the varying predictive events, and while several classifiers used the Random Forest was the best. Sun et al. (2014) used this framework to predict hypertension transition points in EHR data. They used the Random Forest and no temporal representation. Rana et al. (2015) had introduced a framework that models the change in interventions over time to predict outcome events considering the temporal evolution of the events, which was shown to be useful.

2.2 Symbolic Time Intervals Data and Mining

Conceptual representation in biomedicine is common for various purposes (Moskovitch et al, 2004; Boland et al, 2015). Representing multivariate temporal data using (conceptual) symbolic time intervals is becoming more common in the data mining literature too. This is especially true when applied to biomedical data (Patel et al, 2008; Batal et al, 2012) in which the variables, or concepts, are represented by time intervals in their raw form, or after some transformation process (Hoepenner, 2001; Moerchen, 2007). Methods for transformation from time point series into meaningful symbolic time intervals data, a process often called Temporal Abstraction, were proposed in the past decade by several researchers (Shahar, 1997; Hoppner, 2001; Lin et al, 2003; Moerchen and Ultsch, 2005; Moskovitch and Shahar, 2015c). In raw datasets in which the data is already represented by multivariate symbolic time intervals, other forms of abstractions are performed. In this case events that occur within a predefined gap will be concatenated into a single time interval. In this paper we followed the standard aggregation process as defined in the Observational Medical Outcomes Partnership (OMOP, 2009) to abstract EHR data. Additionally, we used the SNOWMED hierarchy to use more general concepts achieving higher generalizability and enabling the discovery of more temporal patterns. Whether the data is raw symbolic time intervals or abstracted from time point series, it can be mined to discover frequent time intervals patterns. In the past decade and a half several algorithms were proposed that improve the efficiency of the mining. Most of the methods use a subset of Allen's temporal relations (Allen, 1983) Early studies defined a pattern in an ambigious way (Villafane et al., 2000; Kam and Fu, 2000), while later studies defined a non ambigious definintion using a conjunction of all the pairwise relations. In this study we use KarmaLego that was introduced by Moskovitch and Shahar (2015a) that improves Papapetrou et al. (2009) and Patel et al (2008) by an efficient data structure and exploiting the transitivity property of the temporal relations. KarmaLego was found to be faster than IEMiner (Patel et al, 2008), ARMADA (Winarko and Roddick, 2007), and H-DFS (Papapetrou et al, 2009) methods.

2.4 Classification and Prediction with Temporal Patterns

Once temporal patterns, i.e., TIRPs, are discovered from longitudinal data they can be used for regression and classification. Quite simultaneously several groups had proposed using TIRPs, as features for classifying multivariate time series (Patel et al, 2008; Moskovitch et al, 2009; Batal et al, 2012), followed by recent studies (Moskovitch and Shahar, 2015b,c). Interestingly, all of the studies that reported the use of temporal abstraction and the use of TIRPs for classification applied to biomedical data (Patel et al, 2008; Batal et al, 2012; Moskovitch and Shahar, 2015b,c). Patel et al. (2008) proposed IEClassifier to classify patients Hepatitis patients using TIRPs (Patel et al, 2008). Batal et al. (2012) performed knowledge based temporal abstraction, but used only two relations: before and co-occur, which is a specific case of an apriori sequential mining algorithm called STF-Mine. They compared recent patterns and older to detect outcome events. Moskovitch and Shahar (2015a) introduced a framework for classification of multivariate time series using several discretizations, such as Equal Width Discretization (EWD), and SAX (Lin et al, 2003; Keogh et al, 2005).

Later proposed a supervised Temporal Discretization for Classification method that increases the accuracy by learning the cutoffs that increase the differences in the states according to their distribution in the classes (Moskovitch and Shahar, 2015c). This approach outperformed the unsupervised EWD and SAX methods. Several studies had shown the advantages of using TIRPs over atemproal representation in classifying multivariate temporal data (Patel et al, 2008; Batal et al, 2012; Moskovitch et al, 2015). Recent studies introduced several heuristics to decrease the number of discovered patterns that still maintain the same level of accuracy (Shknevsky et al., 2014).

2.5 One Class Feature Selection

Feature selection, the identification of the most informative variables, is an important step in supervised machine learning, especially when the number of possible features is very large. Often these are the features that correlate best with the classes. There has been a meaningful work on feature selection that takes into consideration the data of both classes, and several methods were proposed (Guyon and Elisseeff, 2003). However, in some cases there is only one class available, or the model is based only on one class. This is the case in our study in which we have a cohort of patients with an outcome, but no real alternative class, but rather a control group of patients that are for the evaluation and we would not like to base our models on. The other patients in our data, which we use as a comparator group, are not healthy individuals, but have their own conditions and outcomes. In our case we want to perform the selection based on the cohort class only, since the controls are for evaluation and not to rely the predictive model upon.

Thus, we learn only from the cohort class the patterns, and for that, perform feature selection only based on that one class. While one class classficiation were developed in the past years, one class feature selection is a new field with only few available studies (Jeong et al, 2012). In one class classification the intention is to learn to classify an instance into a single class and determine whether it belongs to it or not. The need for one class classifiers (Khan and Madden, 2004) is essential, since in many real world problems there is only one genuine class to learn, while the alternative class is all the rest, or the not-class, which is infinite. Through the development of classifiers, it was mainly binary classifiers that were developed, but real world problems aim at classifying into a specific class, rather than among two classes, or more. This is true for diagnosis of diseases, or almost any real concept learning. Another problem caused by the use of binary classifiers is that many classes have low rates of occurences, which result in the imbalance problem. Being able to learn a predictive model based only on the learned class examples without requiring having alternative class examples is ideal. Some “one class” classifiers were indeed developed based on binary classifiers, such as based on SVM (Manevitz and Yousef, 2001) or based on regression model (Sokolov et al, 2016) and more. However, this a research field with a lot of room for contribution. In this study we examine the ability to select features, or temporal patterns in our case, based only on the learned class, since we want to avoid relying on informatoin from the other class (controls), since in our case they are used for evaluation purposes and not for learning. Thus, the selection has to be done based solely on the predicted class examples. In a recent paper Lorena et al (2014) had proposed several methods for one class feature selection based on the minimizing the heterogeneity of the dataset after a feature is removed, or by selecting features with low correlations with the other features.

3 Methods

3.1 Definitions

We start with presenting the definitions underlying the presented framework and the algorithms. These definitions are based on the definitions framework presented in (Moskovitch and Shahar, 2015a,b,c). In our framework a symbolic time interval, I = <s, e, sym>, is an ordered pair of time points, start-time (s) and end-time (e), and a symbol (sym) that represents one of the domain's temporal concepts. In our study concepts can be a clinical procedure, a condition, or a drug exposure. Given a patient record the symbolic time intervals should be ordered lexicographically based on the start-time, end-time and the symbols.

To represent the temporal relations among a pair of symbolic time intervals, we use an abstracted version of Allen's temporal relations (1983). Allen proposed seven temporal relations to relate a pair of time intervals (i.e., *before, meets, overlaps, contains, finished-by, starts, equal* and their inverse (for example, *after* is the inverse of *before*), which are used by KarmaLego and described in details in (Moskovitch and Shahar, 2015a). Ordering the symbolic time intervals lexicographically, as defined in definition 3, enables the use of only the seven temporal relations, without their inverse. In previous work (Moskovitch et al, 2015) three broader temporal relations defined by disjunctions of Allen's temporal relations, including BEFORE [before ∨ meets], OVERLAP [overlaps] and CONTAIN [contains ∨ finished-by ∨ starts ∨ equal] were proposed. Since they were found to be more effective than the seven temporal relatioins (Moskovitch and Shahar, 2015b,c) or similar (Moskovitch et al, 2015) we used in this study the three relations.

Definition 1—A non-ambiguous lexicographic *Time Intervals Related Pattern* (*TIRP*) P is defined as $P = \{I, R\}$, where $I = \{I^1, I^2, .., I^k\}$ is a set of k symbolic time intervals and

$$\begin{aligned} R &= \bigcap_{i=1}^{k-1} \bigcap_{j=i+1}^{k} r(I^i, I^j) \\ &= \left\{ r_{1,2}(I^1, I^2), ., r_{1,k}(I^1, I^k), .., r_{k-1,k}(I^{k-1}, I^k) \right\} \end{aligned}$$

defines all the temporal relations among each of the $(k^2 - k)/2$ pairs of symbolic time intervals in I.

Figure 1 presents a typical TIRP, represented as a half-matrix of the conjunction of the temporal relations. On the left there is an illustration of a four sized symbolic time intervals TIRP, and on the right there is a half matrix that presents the temporal relations among them. For example, the relation among A and B is overlap (o), while the relation between C and D is finished-by.

Definition 2—Given a database of $|E|$ distinct entities (e.g., different patients), the *vertical support* of a TIRP P is denoted by the cardinality of the set E^P of distinct entities for which P holds, divided by the total number of entities (e.g., patients) $|E|$: $ver\_sup(P) = |E^P| / |E|$.

The vertical support is actually what is commonly used as support in association rules, itemset and sequential mining. A TIRP having above minimal vertical support threshold is referred to as *frequent*.

3.2 The KarmaLego Algorithm

The KarmaLego algorithm is a fast time intervals mining algorithm for the discovery of TIRPs through exploiting the transitivity of temporal relations that enables an efficient candidate generation mechanism (Moskovitch and Shahar, 2015a). KarmaLego consists on two main steps: Karma, in which the entire set of entities' time intervals data are scanned and indexed. Through that all the symbols are counted, and each pair of symbolic time intervals and the temporal relation among them are indexed in an index called DharmaIndex that contains all the frequent 2-sized TIRPs (k=2). The DharmaIndex (Moskovitch et al, 2015) will be used laterin the Lego phase to retrieve the relevant pairs through the TIRPs extension process.

In the second phase, referred to as the Lego algorithm, a recursive process extends the 2-sized TIRPs that are frequent. Based on the symbol and the relation *r*, a set of candidate TIRPs are generated highly efficiently, by exploiting the transitivity of the temporal relations, into a tree of longer frequent TIRPs. These consist of conjunctions of the 2-sized TIRPs that were discovered in the Karma phase. Lego receives a TIRP t that is extended by any of the frequent symbols in T^1 (S), and any temporal relation r of the R temporal relations, which holds between the new symbolic time interval and the last one in the extended TIRP. The result of this process is a frequent TIRPs enumeration tree (Moskovitch and Shahar, 2015a). More details about the Karma algorithm are in (Moskovitch and Shahar, 2015a). Unlike in previous studies, in this study we had thousands of types of symbols, thus, we implemented the DharmaIndex using hash-tables, instead of arrays as described in (Moskovitch et al, 2015).

3.3 The SingleKarmaLego Algorithm

In order to classify a patient based on TIRPs as features they have to be detected at the patients' records. For that we use SingleKarmaLego (Moskovitch and Shahar, 2015b)., an algorithm for fast TIRPs detection. Compared to a naïve TIRPs detection algorithm, due to the first phase of indexing the pairs of symbolic time intervals, SingleKarmaLego was shown to be much faster, especially when the number of TIRPs to detect is larger. Single-KarmaLego indexes all the pairs of the entity's symbolic time intervals into an efficient data structure that we call *DharmaIndex*. Similar to KarmaLego, this allows accessing the pairs index fast in the detection of longer TIRPs. SingleKarmaLego is applied on a single entity, thus, the DharmaIndex in this case does not contain the entity id.

3.4 Maitreya - Prediction of Outcome Events

In this paper we propose a methodology for the prediction of outcome events using time intervals related patterns. Figure 2 illustrates the Maitreya framework workflow. Our methodology consists of discovering TIRPs only in the set of the cohort entities (i.e., patients) having the outcome event (the outcome cohort). This is in order to base the prediction model only on the cohort information, while the controls data will be used only for evaluation purposes. Since the discovered TIRPs can vary from different cohort of patients, so for that we are divide the data into three folds, on which we perform three iterations of mining. Thus, the TIRPs are discovered and consist only on one fold in the outcome cohort. Then the TIRPs that were discovered in the Cohort patients are going through a one class feature selection process, and the TIRPs that were selected are detected using Single-KarmaLego on the other two folds both in the cohort and in the control. The output of applying SingleKarmaLego is a matrix of patients and TIRPs and the appearances in three levels: Binary, Horizontal Support and Mean Duration. The classification model is induced from the matrix containing both the cohort and controls. Thus, the TIRPs are discovered from one fold that are detected in the other two folds and on which we perofrm a ten folds cross validation experiments. Finally after a prediction model is induced a new patient can be classified.

3.4.1 Data Creation—

A control set of entities (i.e., patients) is defined that are patients from the rest of the database, who do not have the outcome in their records and were selected randomly. In the cohort we took data from a prediction-time before the outcome event to an observation-time prior to the end of the prediction time. For the controls the same was made relatively to a procedure (that is not the outcome) in the patients data.

3.4.2 Model Learning and Evaluation Scheme—

As mentioned, since in TIRPs mining in each fold, or different population, a different set of frequent TIRPs may be discovered. Additionally, in order to avoid overfitting, we perform a rigorous evaluation strategy, in which the cohort and control datasets are divided into three folds. Then, in each iteration one fold of the cohort patients is mined for TIRPs and then the TIRPs are detected in the other two folds for the later classificiation experiment that includes ten folds cross valdidation. Note, the TIRPs are not discovered initially from the entire dataset (Moskovitch and Shahar, 2015b). Thus, in each iteration TIRPs are discovered from 1/3 of the cohort (one fold). Then SingleKarmaLego is applied to detect these TIRPs in the other 2/3 (two folds) of the cohort and the other 2/3 (two folds) of the control. Once the TIRPs' instances were detected a matrix of TIRP-features is created.

Thus, the TIRPs are learned from a distinct cohort population than the one used for the classification evaluation that includes ten folds cross validation. Overall each experiment includes three folds mining and ten fodls classiciation cross validation, which results in thirty prediction experiments. As an additional step, to avoid over-fitting we limit the number of features to the root of the number of patients in the cohort.

3.5 TIRPs Features Representation

After we apply SingleKarmaLego to the cohort and control to detect TIRPs' instances in the patient data, a matrix is constructed. The rows of this matrix are the entities, which in our case are patients, and the columns are the TIRP features. Each matrix entry will have a value representing the TIRP for a given patient. The default value is Boolean, which was used in previous studies (Patel et al, 2008; Batal et al, 2012). We used in addition to the default Boolean more expressive metrics, such as the *Horizontal Support* and *MeanDuration* that were introduced in (Moskovitch and Shahar, 2015b,c). Horizontal Support is the number of the TIRP's instances that were detected at the patient's records, and the Mean Duration is the average of the time length. These metrics are defined formally at (Moskovitch and Shahar, 2015b,c).

Note that what we refer to as *horizontal support* is the same as *term frequency* in text analytics (Salton et al, 1983), where the term frequency is the number of occurences of a term in text, horizontal support is the number of occurrences of a term in text, horizontal support.

3.6 One Class Feature Selection

Due to the large number of TIRPs that are discovered in the mining process that will become the features for the classification, we want to avoid overfitting. For that we reduce the number of TIRP predictors to the root of the number of patients in the outcome cohort, using feature selection.

As was explained earlier, since we don't want to rely on the controls in our modeling, we can not use traditional feature selection methods that consider both classes in their metrics. Thus, we wanted to apply feature selection that consists only on the data from the cohort class, which is called one class feature selection. For that we propose three strategies and evaluate them: (1) selecting those with the highest vertical support; (2) based on the correlation of the TIRP to other TIRPs, and selecting those with the lowest average correlation; (3) and based on the homogeneity of the cohort class (i.e., selecting those that their removal decreases the homogeneity). For that we used the following criterions.

3.6.1 Vertical Support based Selection—We used the vertical support that we defined in definition 2. Based on the vertical support metric each TIRP has its frequency in the cohort, and based on that sorted the TIRPs based on their vertical support in a descending order. The TIRPs having the highest vertical support were selected. This was our default criterion.

3.6.2 Correlation based Selection—For the Correlation based One Class Feature Selection the average correlation is measured using the Pearson measure for each TIRP using the following formula:

$$CorrOCFS(TIRP_i)=\left(\frac{1}{m}\sum_{j=1}^{m}\text{Person}(TIRP_j)\right)$$

In the shown formula given a TIRP i, all the m TIRPs correlation with the given TIRP i are calculated. Thus, we would like to choose the TIRPs, or features, that have the lowest averaged correlation across all the other TIRPs.

3.6.3 Homogeneity based Selection—The idea behind the homogeneity score is that the more the (cohort) class homogeneity is higher it is expected to be better for the classification. For the Homogeneity based One Class Feature Selection we calculate the homogeneity of the patients vectors. Thus, a TIRP that has a high homogeneity score means that its removal decereases less the homogeneity. The homogeneity score is calculated based on the cosine similarity of any pair of the entities vectors, after the removal of each TIRP accordingly:

$$HomogOCFS(TIRP_i)=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}Cosine(e^i,e^j)\right)$$

After that the TIRPs having the highest scores are selected, since the removal of the others (lowest) will decrase the homogeneity.

4 Evaluation

4.1 Datasets

We extracted clinical data from the New York Presbyterian Columbia University Medical Center Hospital (NYP-CUMC) clinical data warehouse. In total, the CUMC-NYP EHR contains medical record data for approximately 1.5 million patients going back to 1989, containing approximately 30 million diagnosis billing codes, 20 million prescription orders, 9 million procedures, and 500 million laboratory results. Only coded data used for this analysis, including drug exposures, conditions (billing codes), and procedures. These concepts were mapped to RxNorm, SNOMED-CT, and ICD-9-Procedure, respectively, to conform to the Common Data Model made available by the Observational Medical Outcomes Partnership (OMOP). Medical concepts are then transformed into symbolic time intervals (called "eras") and further concatenated according to definition 3. Definition 3 is based on the OMOP standards, in which two time intervals ("eras") having the same concept id (symbol) and having a gap less than a given period of time are concatendated into a single time interval having the same symbol.

Definition 3. Abstraction function—Given two symbolic time intervals having the same symbol I^i and I^j, and I^i is before I^j according to the lexicographical order, and $I^j s - I^i e <$ max_time holds, it will be abstracted into a new single time interval having the start time of I^i and the end time of I^j. In our case max_time was set to 30 days, according to the OMOP standard.

$$\forall \boldsymbol{I^i}, \boldsymbol{I^j} \in \boldsymbol{TIS} \wedge (\boldsymbol{i}<\boldsymbol{j}) \wedge \left(\boldsymbol{I^i_{sym}}=\boldsymbol{I^j_{sym}}\right) \wedge \left(\boldsymbol{I^j_s} - \boldsymbol{I^i_e}< \underset{time}{\mathbf{max\_}}\ \mathbf{time}\right) : \boldsymbol{TIS} \leftarrow \boldsymbol{TIS} - \boldsymbol{I^i} \wedge \boldsymbol{TIS} \leftarrow \boldsymbol{TIS} - \boldsymbol{I^j} \wedge \boldsymbol{newI.s}=\boldsymbol{I^i_s} \wedge \boldsymbol{newI.e}=\boldsymbol{I^j_e} \wedge \boldsymbol{TIS} \leftarrow \boldsymbol{TIS}+\boldsymbol{newI}$$

4.2 Research Questions

For the evaluation of the outcomes prediction, we selected eight commonly performed clinical procedures targeting different organ systems, shown in Table 1. Because most of these procedures would have been performed either in outpatient or inpatient setting, we assumed that these procedures would allow for inclusion of patients with a range of disease severity or acuity. We wanted to answer the following questions: (1) whether using TIRPs as features is better than symbols; (2) whether the use of the one class classification feature selection methods is better than using the most frequent TIRPs (vertical support); and (3) whether the TIRPs representation will be better than the default Boolean that were used in previous studies. We compared our methods to the work of Batal et al (2012), in which prediction models for diseases based on time intervals patterns were learnt. In their work they used as a baseline the symbols and when TIRPs were used they were represented by a Boolean representation. These two were used as baseline in our evaluation. We used 20% minimal vertical support for all the procedures and did not limit the maximal gap. All the experiments were performed with three folds of mining cross validation, as was explained in section 3.4 and ten folds of classification cross validation using weka 3.6.11. Thus, each result that is reported is based on thirty experimental runs.

5 Results

We ran Maitreya on the eight clinical procedures prediction tasks datasets in Table 1, in order to answer the research questions in section 4.2. Table 2, 3 and 4 present the results of the procedures prediction with the three feature selection strategies. All the tables have the same structure of contents. Symbols (first column) refers to the use of no temporal analysis as features, thus, the symbols are the types of events in the data, and the next columns are of the TIRPs with three types of representations: Boolean, Horizontal Support and Mean Duration. The results are presented as means and 95% confidence intervals. The baseline methods: symbols and Boolean TIRPs are based on (Batal et al, 2012).

5.1 Outcomes Prediction Evalution

In this experiment the symbols and the TIRPs were selected having the highest vertical support. The results in Table 2 show that in average the use of Symbols was the worst, and when using the TIRPs with the Horizontal Support representation it was meaningfuly better, and for some procdures its signfiicant. However, for specific procedures the differences were bigger.

Table 3 presents the results when the homogeneity one-class feature selection was used to select the TIRPs. The magnitude of the results was similar in this experiment. Thus, the use of TIRPs with horizontal support was better than the Symbols and the others. The TIRPs with the Boolean representation was slightly worse than the Symbols. However, overall the use of the homogeneity one-class feature selection was less effective than the use of vertical support for feature selection.

Table 4 presents the results of the procedures prediction when using the correlation based one class feature selection to select the TIRPs features. Also here the behavior of the framework was the same, while the overall results were slightly worse than the use of homogeneity for selection. However, again the use of vertical support was more effective and the correlation based seems slightly worse than the homogeneity.

6. DISCUSSION AND CONCLUSIONS

The increasing availability of longitudinal data provides an opportunity for temporal knowledge discovery. However, it also brings many challenges due to the common sparsity and the various forms of temporal variables.

The first and default feature selection is based on the vertical support of each TIRPs, in which we favored the TIRPs having the highest vertical support to minimize the sparsity of the data. The second measures the homogeneity of the entire dataset by calculating the average similarity of each pair of patients after a TIRP, or feature, were removed.

In this paper we described a framework, called Maitreya, for the prediction of outcome events and applied it to Electronic Health Records. Maitreya was designed to learn a model based on the cohort class of the outcome event, while a control class of entities (i.e. patients) is selected for evaluation purposes.

Maitreya discovers TIRPs from the cohort using Karma-Lego, a fast TIRPs discovery algorithm, and later detects the TIRPs, which are used as features for classification, using SingleKarmaLego, a fast TIRPs detection algorithm. To avoid over-fitting Maitreya selects TIRPs features for which we proposed three approaches for one class features selection. Finally, the TIRPs are represented using three TIRP representations in addition to the previous used Boolean representation (Batal et al, 2012).

In this case the TIRPs that after their removal the homogeneity was the highest were removed first. The third method measured the average correlation between a given TIRP and the rest of the TIRPs. The TIRPs that were less correlated were chosen.

We evaluated Maitreya on eight procedures from the EHR domain for which we created datasets from Columbia University Medical Center EHR. We compared the use of TIRPs as features to the use of symbols without any temporal representation, and to the use of TIRPs with Boolean representation as proposed by (Batal et al, 2012). Our results show that representing TIRPs with Boolean representation, describing whether a TIRP was detected for the patient or not, performed slightly better than the symbols baseline. Representing the TIRPs using the horizontal support performed the best and was significantly better than the Symbols, and using it with mean duration was better than the Symbols but slightly worse than the Horizontal Support. However, our experiments had shown that using the one class feature selection approaches did not improve the performance relatively to the use of the vertical support for feature selection. For future work would like to further explore other one class feature selection approaches, as well as analyze the clinical meaning of the predictive TIRPs.

Acknowledgments

The authors would like to thank Drs Papapetrou, Moerchen, and Lin, for random discussions on temporal abstraction, time intervals mining and their potential use. This work was supported in part by a grant from the NIGMS R01GM107145.

Biographies

Robert Moskovitch holds a B.Sc., M.Sc., and a Ph.D. in Information Systems Engineering from Ben Gurion University and is currently a postdoctoral research scientist at the Department of Biomedical Informatics at Columbia University. Prior to that, he headed several Research and Development projects in Information Security at Deutsche Telekom Innovation Laboratories at BGU. He has served on several journal editorial boards, as well as on program committees of several conferences, such as SIGKDD, and workshops in Biomedical Informatics and Information Security. He published more than sixty peer-reviewed papers in leading journals and conferences, several of which had won best-paper awards.

Hyunmi Choi is a board-certified physician in neurology, neurophysiology, and epilepsy. She holds a degree in B.S (‘94), MD (‘96), and M.S. in Biostatistics (‘2004). An Associate Professor of Neurology at Columbia University Medical Center, her primary area of research interest includes health services research, particularly in the area of decision science.

George Hripcsak George Hripcsak, MD, MS, is a board-certified internist with degrees in chemistry (BS 1981), medicine (MD 1985), and biostatistics (MS 2000). He is Vivian Beaumont Allen Professor and Chair of Columbia University's Department of Biomedical Informatics and Director of Medical Informatics Services for NewYork-Presbyterian Hospital/Columbia Campus. He co-chaired the Meaningful Use Workgroup of U.S. Department of Health and Human Services's Office of the National Coordinator of Health Information Technology; it defines the criteria by which health care providers collect incentives for using electronic health records. He led the effort to create the Arden Syntax, a language for representing health knowledge that has become a national standard. Dr. Hripcsak chaired the U.S. National Library of Medicine's Biomedical Library and Informatics Review Committee, and he is a fellow of the National Academy of Medicine, the American College of Medical Informatics, and the New York Academy of Medicine. He has published over 250 papers.

Nicholas Tatonetti is Herbert Irving assistant Professor of biomedical informatics in the Departments of Biomedical Informatics, Systems Biology, and Medicine and is Director of Clinical Informatics at the Herbert Irving Comprehensive Cancer Center at Columbia University. He holds degrees in molecular biosciences/biotechnology (BS '08) and computational mathematics (BS '08) from Arizona State University and biomedical informatics (MS '11, PhD '12) from Stanford University. His primary research interests focus on the development of novel statistical and computational methods for observational data mining. He applied these methods to detect, explain, and validate drug effects and drug interactions from large-scale data.

References

Allen JF. Maintaining knowledge about temporal intervals. Communications of the ACM. 1983; 26(11):832–843.

Batal, I., Fradkin, D., Harrison, J., Moerchen, F., Hauskrecht, M. Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data. Proceedings of Knowledge Discovery in Databases (KDD); Beijing, China. 2012.

Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. International Journal of Medical Informatics. 2008; 77(2)

Boland M, Jacunski A, Lorberbaum T, Romano J, Moskovitch R, Tatonetti N. Systems Biology Approaches for Identifying Adverse Drug Reactions and Elucidating their Underlying Biological Mechanisms. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2015

Esling P, Agon C. Time Series Data Mining. ACM Computing Surveys. 2012; 45(1):12.

Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research. 2003; 3:1157–1182.

Höppner F. Learning Temporal Rules from State Sequences. Proceedings of WLTSD. 2001

Hripcsak G, Albers DJ, Perotte A. Exploiting Time in Electronic Health Records. Journal of American Medical Association. 2013; 20:2.

Hripcsak G, Albers D. Next-Generation Phenotyping of Electronic Health Records. Journal of American Medical Informatics Association. 2013; 20:117–121.

Hripcsak G, Albers DJ, Perotte A. Parameterizing Time in Electronic Health Record Studies. Journal of the American Medical Informatics Assofication. 2015

Hu B, Chen Y, Keogh E. Time Series Classification under More Realistic Assumptions. Proceedings of SIAM Data Mining. 2013

Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics. 2012; 13(6)

Jeong YS, Kang IH, Jeong MK, Kong D. A New Feature Selection Method for One-Class Classification Problems. IEEE Transactions on systems, man, and cybernetics part C. 2012; 42:6.

Keogh, E., Lin, J., Fu, A. IEEE International Conference on Data Mining. Houston, Texas: 2005. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence; p. 226-233.

Lin J, Keogh E, Lonardi S, Chiu B. A Symbolic Representation of Time Series with Implications for Streaming Algorithms. 8th ACM SIGMOD DMKD workshop. 2003

Manevitz L, Yousef M. One Class SVMs for Codument Classification. Journal of Machine Learning Research. 2001; 2:139–154.

Mörchen F, Ultsch A. Optimizing Time Series Discretization for Knowledge Discovery. Proceeding of SIGKDD. 2005

Moerchen F. Unsupervised Pattern Mining from Symbolic Temporal Data. SIGKDD Explorations. 2007; 9(1):41–55.

Moskovitch R, Hessing A, Shahar Y. Vaidurya-a concept based, context sensitive search engine for clinical guidelines. Medinfo. 2004

Moskovitch R, Shahar Y. Medical Temporal Knowledge Discovery via Temporal Abstraction. American Medical Informatics Association. 2009

Moskovitch R, Shahar Y. Fast Time Intervals Mining Using Transitivity of Temporal Relations. Knowledge and Information Systems. 2015a; 42:1.

Moskovitch R, Shahar Y. Classification of Multivariate Time Series via Temporal Abstraction and Time Intervals Mining. Knowledge and Information Systems. 2015b; 45(1):35–74.

Moskovitch R, Shahar Y. Classification Driven Temporal Discretization of Multivariate Time Series. Data Mining and Knowledge Discovery. 2015c; 29(4):871–913.

Moskovitch, R., Walsh, C., Wang, F., Hripsack, G., Tatonetti, N. Outcomes Prediction via Time Intervals Related Patterns. IEEE International Conference on Data Mining (ICDM); Atlantic City, USA. 2015.

Ng K, Ghoting A, Steinhubl SR, Stewart WF, Malin B, Sun J. PARAMO: A PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. Journal of Biomedical Informatics. 2014; 48:160–170. [PubMed: 24370496]

Papapetrou P, Kollios G, Sclaroff S, Gunopulos D. Mining frequent arrangements of temporal intervals. Knowledge and Information Systems. 2009; 21(2)

Patel D, Hsu W, Lee M. Mining Relationships among Interval-based Events for Classification. Proceedings of the ACM SIGMOD international conference on Management of data. 2008

Rana S, Gupta S, Phung D, Venkatesh S. A predictive framework for modeling healthcare data with evolving clinical interventions. Statistical Analysis and Data Mining. 2015 In Press.

Ratanamahatana C, Keogh EJ. Three Myths about Dynamic Time Warping Data Mining. Proceedings of Siam Data Mining. 2005

Reisinger SJ, Ryan PB, O’Hara DJ, Powell GE, Painter JL, Pattishall EN, Morris JA. Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. Journal of American Medical Informatics Association. 2010; 17(6)

Roddick J, Spiliopoulou M. A survey of temporal knowledge discovery paradigms and methods. IEEE Transactions on Knowledge and Data Engineering. 2002; 4(14)

Sacchi L, Larizza C, Combi C, Bellazzi R. Data mining with temporal abstractions: learning rules from time series. Data Mining and Knowledge Discovery. 2007; (15)

Salton, G., McGill, M. Introduction to Modern Information Retrieval. New York, NY: McGraw-Hill Book Company; 1983.

Shahar Y. A framework for knowledge-based temporal abstraction. Artificial Intelligence. 1997; 90(1–2)

Khan SS, Madden MG. One Class Classification: Taxonomy of Study and Review of Techniques. The Knowledge Engineering Review. 2004

Shknevsky, A., Moskovitch, R., Shahar, Y. Semantic Considerations in Time Intervals Mining. ACM KDD on Workshop on Connected Health at Big Data Era; NYC, USA. 2014.

Stopel D, Boger Z, Moskovitch R, Shahar Y, Elovici Y. Application of Artificial Neural Networks Techniques to Computer Worm Detection. Proceedings of the International Joint Conference on Neural Networks. 2006

Sokolov, A., Paul, EO., Stuart, JM. One Class Detection of Cell States in Tumor Subtypes. Pacific Symposium on Biocomputing; Hawaii. 2016.

Sun J, McNaughton CD, Zhang P, Perer A, Gkoulalas-Divanis A, Denny JC, Kirby J, Lasko T, Saip A, Malin BA. Predicting Changes in Hypertension Control Using Electronic Health Records from a Chronic Disease Management Program. Journal of the American Medical Informatics Association. 2014; 21:337–344. [PubMed: 24045907]

Villafane R, Hua K, Tran D, Maulik B. Knowledge discovery from time series of interval events. Journal of Intelligent Information Systems. 2000; 15(1)

Winarko E, Roddick J. Armada - an algorithm for discovering richer relative temporal association rules from interval-based data. Data and Knowledge Engineering. 2007; 1(63)

Wu S, Chen Y. Mining Non-ambiguous Temporal Patterns for Interval-Based Events. IEEE Transactions on Knowledge and Data Engineering. 2007; 19(6)

Wu J, Roy J, Stewart WF. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical Care. 2010; 48:106–13.

Yi-Cheng C, Ji-Chiang J, Wen-Chih P, Suh-Yin L. An Efficient Algorithm for Mining Time Interval-based Patterns in Large Databases. Proceedings of CIKM. 2010

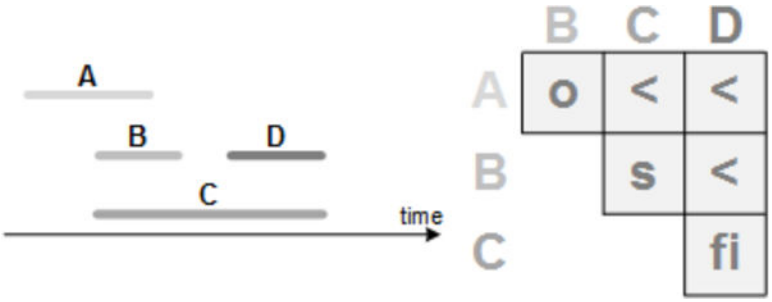

Figure 1.

An example of a Time Intervals Related Pattern (TIRP), containing a sequence of four lexicographically ordered symbolic time intervals and all of their pair-wise temporal relations shown on the right half matrix.

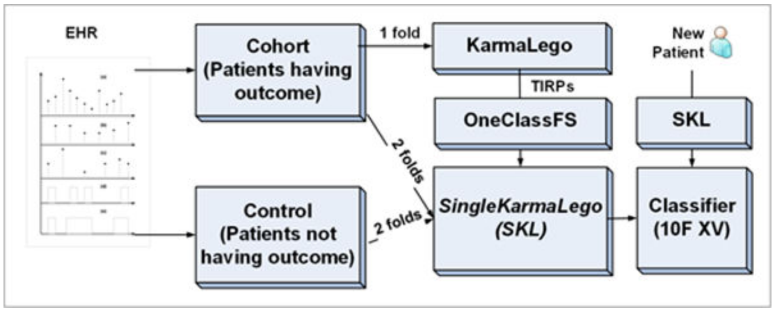

Figure 2.

Maitreya, a framework for outcomes prediction. A cohort set of patients having the outcome is constructed, and a corresponding control group of patients who don't have the outcome. One fold is used to discover frequent TIRPs using KarmaLego. SingleKarma-Lego detects the TIRPs that were selected by the One Class Feature Selection at the other two folds of the cohort and controls are used to detect these TIRPs. Then a matrix is constructed on which 10 folds cross validation classification is performed. Finally, once a new patient arrives his TIRPs are detected using SIngleKarmaLego and given to the induced classifier classifier.

Table 1

The outcome clinical procedures datasets

Procedure Code	**Procedure Description**	**#cases**
VEN_CATH	Venous catheterization	4264
ENDS_SML	Endoscopy of small intestine	1420
COLONSCPY	Colonoscopy	2506
ABDMN_ULTRA	Abdomen diagnostic ultrasound	1798
BRAN_IMAG	Brain Magnetic resonance imaging	1808
INTRW_EVAL	Interview and evaluation	1364
CRD_VAS_TST	Cardiovascular stress test	1194
MECH_VENT	Continuous invasive mechanical ventilation	1210

Table 2

Using the vertical support as one class feature selection.

Procedure	Symbols [baseline] (mean AUC −/+ 95% ci)	TIRPs (mean AUC −/+ 95% CI)		
		Boolean [baseline]	Horizontal Support	Mean Duration
VEN_CATH	0.684 −/+ 0.012	0.688 −/+ 0.018	0.727 −/+ 0.015	0.714 −/+ 0.015
ENDS_SML	0.644 −/+ 0.024	0.669 −/+ 0.027	0.685 −/+ 0.021	0.649 −/+ 0.032
COLONSCPY	0.577 −/+ 0.015	0.574 −/+ 0.021	0.609 −/+ 0.023	0.588 −/+ 0.026
ABDMN_ULTRA	0.656 −/+ 0.028	0.618 −/+ 0.030	0.676 −/+ 0.028	0.664 −/+ 0.034
BRAN_IMAG	0.655 −/+ 0.027	0.678 −/+ 0.020	0.699 −/+ 0.025	0.673 −/+ 0.025
INTRW_EVAL	0.790 −/+ 0.025	0.801 −/+ 0.023	0.834 −/+ 0.021	0.799 −/+ 0.023
CRD_VAS_TST	0.601 −/+ 0.035	0.637 −/+ 0.027	0.699 −/+ 0.026	0.660 −/+ 0.027
MECH_VENT	0.659 −/+ 0.031	0.668 −/+ 0.031	0.694 −/+ 0.031	0.684 −/+ 0.039
Mean	0.658 −/+ 0.025	0.667 −/+ 0.025	0.703 −/+ 0.024	0.679 −/+ 0.028

Table 3

Using the homogeneity one class feature selection.

Procedure	Symbols [baseline] (mean AUC −/+ 95% ci)	TIRPs (mean AUC −/+ 95% CI)		
		Boolean [baseline]	Horizontal Support	Mean Duration
VEN_CATH	0.684 −/+ 0.012	0.693 −/+ 0.014	0.732 −/+ 0.016	0.712 −/+ 0.015
ENDS_SML	0.644 −/+ 0.024	0.638 −/+ 0.031	0.662 −/+ 0.030	0.654 −/+ 0.028
COLONSCPY	0.577 −/+ 0.015	0.571 −/+ 0.021	0.614 −/+ 0.015	0.602 −/+ 0.027
ABDMN_ULTRA	0.656 −/+ 0.028	0.658 −/+ 0.022	0.691 −/+ 0.021	0.672 −/+ 0.028
BRAN_IMAG	0.655 −/+ 0.027	0.658 −/+ 0.023	0.689 −/+ 0.022	0.660 −/+ 0.026
INTRW_EVAL	0.790 −/+ 0.025	0.799 −/+ 0.024	0.834 −/+ 0.019	0.792 −/+ 0.028
CRD_VAS_TST	0.601 −/+ 0.035	0.632 −/+ 0.032	0.685 −/+ 0.027	0.631 −/+ 0.030
MECH_VENT	0.659 −/+ 0.031	0.662 −/+ 0.028	0.716 −/+ 0.033	0.677 −/+ 0.035
Mean	0.658 −/+ 0.025	0.664 −/+ 0.025	0.703 −/+ 0.023	0.675 −/+ 0.027

Table 4

Using the one class correlation based feature selection.

Procedure	Symbols [baseline] (mean AUC −/+ 95% ci)	TIRPs (mean AUC −/+ 95% CI)		
		Boolean [baseline]	Horizontal Support	Mean Duration
VEN_CATH	0.684 −/+ 0.012	0.688 −/+ 0.018	0.736 −/+ 0.016	0.713 −/+ 0.015
ENDS_SML	0.644 −/+ 0.024	0.685 −/+ 0.018	0.698 −/+ 0.025	0.684 −/+ 0.024
COLONSCPY	0.577 −/+ 0.015	0.570 −/+ 0.021	0.606 −/+ 0.021	0.596 −/+ 0.024
ABDMN_ULTRA	0.656 −/+ 0.028	0.649 −/+ 0.031	0.697 −/+ 0.022	0.668 −/+ 0.025
BRAN_IMAG	0.655 −/+ 0.027	0.676 −/+ 0.022	0.686 −/+ 0.020	0.673 −/+ 0.027
INTRW_EVAL	0.790 −/+ 0.025	0.802 −/+ 0.023	0.841 −/+ 0.019	0.799 −/+ 0.025
CRD_VAS_TST	0.601 −/+ 0.035	0.560 −/+ 0.036	0.592 −/+ 0.030	0.564 −/+ 0.034
MECH_VENT	0.659 −/+ 0.031	0.664 −/+ 0.033	0.692 −/+ 0.031	0.686 −/+ 0.029
Mean	0.658 −/+ 0.025	0.662 −/+ 0.025	0.693 −/+ 0.023	0.673 −/+ 0.025