Transitive Sequential Pattern Mining for Discrete Clinical Data

Hossein Estiri[1,2(✉)], Sebastien Vasey[3], and Shawn N. Murphy[1,2]

[1] Massachusetts General Hospital, Boston, MA 02144, USA
hestiri@mgh.harvard.edu
[2] Harvard Medical School, Boston, MA 02155, USA
[3] Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

Abstract. Electronic health records (EHRs) contain important temporal information about disease progression and patients. However, mining temporal representations from discrete EHR data (e.g., diagnosis, medication, or procedure codes) for use in standard Machine Learning is challenging. We propose a transitive Sequential Pattern Mining approach (tSPM) to address the temporal irregularities involved in recording discrete records in EHRs. We perform experiments to compare the classification performance metrics for predicting "true" diagnosis between traditional sequential pattern mining (SPM) and the proposed tSPM algorithms across multiple diseases. We demonstrate that transitive approach is superior to the traditional SPM in mining temporal representations for diagnosis prediction.

Keywords: Transitive Sequential Pattern Mining · Electronic Health Records · Temporal representations

1 Introduction

In today's biomedical research, modern Machine Learning (ML) algorithms are being increasingly applied to data from electronic health records (EHRs). EHRs contain important temporal information about disease progression and patients. However, in the temporal nature of these data a multitude of convolutions are embodied that are created through what is known as the recording processes [11]. EHR observations are often acquired asynchronously across time (i.e., recorded at different time instants and irregularly in time) [2,5]. This property provides challenges for directly applying standard temporal analysis methods to clinical data recorded in EHRs. Specifically, traditional time-series analysis methods can not be directly applied to analyze the temporal dimensions of the discrete clinical data (e.g., diagnosis records) due to the irregular acquisition of these data. There are also concerns about the validity of diagnosis records in EHRs due to the payer-provider dynamics and healthcare processes [11].

Innovative methods that enable us to properly incorporate time and understand the complexities involved in the healthcare process can yield to interpretable findings from large scale clinical databases. Since the late 1990s, a body

M. Michalowski and R. Moskovitch (Eds.): AIME 2020, LNAI 12299, pp. 414–424, 2020.
https://doi.org/10.1007/978-3-030-59137-3_37

of work has accumulated on mining temporal representations. Temporal representation mining aims to transform temporal data into machine-readable representations, which can be used in standard Machine Learning (ML) techniques for temporal reasoning. Much of this work in medical domains explicitly has been thought of, formed, and developed around clinical measurements data that are commonly continuous in nature and have precise time stamps. From the data mining community, sequential pattern mining (SPM) approaches provide solutions for constructing temporal representations from discrete clinical data. However, SPM algorithms were primarily developed on transaction data, and thus, may require careful modifications for application in medicine.

In this paper, we propose modifications to the traditional SPM approach in a transitive sequential pattern mining (tSPM) algorithm. The goal in tSPM is to mine temporal data representations from discrete clinical data (e.g., diagnosis or medication codes) for application in downstream ML. We apply the tSPM and SPM to predicting "true" diagnosis records in the next visit for six diseases and demonstrate that the tSPM algorithm improves the prediction performance over the traditional SPM by up to 15%.

2 Related Work

2.1 Temporal Representation Mining

Temporal representation mining involves providing a machine-readable representation that formalizes the notion of time in the context of a set of events and temporal relationships [24]. In biomedical research, development and evolution of temporal representation approaches has been largely confined to the abstraction of temporal data from clinical measurements [23]. While numerical observations (e.g., laboratory test results) have explicit timestamps, precise time-stamped information is often unavailable in discrete clinical data such as records of diagnoses, medications, and procedures.

Sequential Pattern Mining (SPM) [1] is a viable approach for discrete data. The goal in SPM is to discover "relevant" sub-sequences from a large set of sequences (of events or items) with time constraints. The "relevance" is often determined by a user-specified occurrence frequency, known as the minimum support [14]. The frequent sequential pattern (FSP) problem is to find the frequent sequences among all sequences [6]. Apriori-based sequential pattern mining methods, such as sequential pattern discovery using equivalence classes (SPADE) [27] and sequential pattern mining (SPAM) [3], are popular in the healthcare domain. For example, Perer, Wang, and Hu (2015) used the SPAM algorithm for mining long sequences of events [21]. The apriori property is that if a sequence cannot pass the minimum support test (i.e., is not assumed frequent), all of its sub-sequences will be ignored. The goal in the apriori-based approach is to cut the number of features by finding the frequent temporal patterns among all patterns. For example, most frequent temporal patterns were utilized in Moskovitch & Shahar (2015); Batal et al. (2013, & 2016),

Moskovitch et al. (2015 & 2017), and Orphanou et al. (2018) [4,5,16–19]. For various reasons, temporal patterns that are mined based on frequency may not make clinical sense.

2.2 Discrete Clinical Event Prediction

Recurrent Neural Networks (RNNs) [22] and its variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are being increasingly applied to model discrete clinical events [13]. Applied a LSTM model to synthetic EHR data (derived from MIMIC-III dataset [12]) to predict the next medication administration events, lab results events, procedure events, and physiological result events. Their study, however, does not model diagnosis records, potentially due to the limitations of the dataset. [8] proposed Med2Vec, a scalable two layer neural network for learning lower dimensional representations to predict future medical codes and Clinical Risk Groups (CRG) levels. Evaluation of this work, however, is limited to prediction of inpatient visits. [9] developed the REverse Time AttentIoN model (RETAIN) that mined sequences of primary care visits to predict if a patient will be diagnosed with heart failure. On the one hand, RETAIN has been tested only on 1 disease. On the other hand, all these models assume that the existence of a diagnosis code in the medical records reflects true diagnosis. These models, therefore, at best can mimic clinical data, which may not represent the truth about patients' health. In this study, we predict the possibility that the patient will be "truly" diagnosed for a disease based on her past medical records, in six different diseases.

3 Methodology

We aim to leverage the diagnosis and medication histories in patients' electronic medical records to predict the "true" diagnosis record at the next visit. This presents an important difference from the related work in this space that aim to predict the next diagnosis records. To do so, we mine two types of sequential patterns (i.e., representations) and train and test prediction models on computer- and expert-annotated labels.

3.1 Traditional Sequential Pattern Mining

We first constructed a baseline method that applies the traditional sequential pattern mining (SPM) algorithm, using EHR observations as features for disease prediction. Given a list $O_1, O_2, \ldots, O_n$ of diagnosis or medication observations, for each patient p, we have recorded the times $t^p_{i1} \leq t^p_{i2} \leq \ldots \leq t^p_{ik^p_i}$ at which the observation O_i was recorded.

In the SPM approach, the features are all possible *pairs* of distinct observations (O_i, O_j), $i \neq j$. We now explain how their frequencies are counted. For a given patient p and a given time t, let $t' > t$ be minimal such that for some i and some $\ell \leq k^p_i$, $t^p_{i\ell} = t'$. In words, t' is the first time strictly bigger than t

at which some observation is made for the patient (it could be several observations are made at the same time). Now, for a given patient p and a given index $i \in \{1, 2, \ldots, n\}$, let S_{pi} be the set of all pairs (j, ℓ'), with $j \in \{1, 2, \ldots, n\}$ and $\ell' \leq k_j^p$, such that observation j is made right after observation i at time $t_{j\ell'}^p$. Formally, $(j, \ell') \in S_{pi}$ if and only if $k_i^p, k_j^p \geq 1$ and there exists $\ell \leq k_i^p$ such that $(t_{i\ell})' = t_{j\ell'}$.

For each $i, j \leq n$, $i \neq j$, let $r_{ijp} = |S_{ip}|$. We think of the r_{ijp}'s as samples of a random variable X_{ij}. Our goal is then to predict the class label Y given $(X_{ij})_{i \neq j}$.

3.2 Transitive Sequential Pattern Mining

To account for irregularity of clinical records and recording processes, we propose a transitive sequential pattern mining (tSPM) algorithm. In the tSPM algorithm, the features are again all possible pairs of distinct observations (O_i, O_j), $i \neq j$. For a fixed patient p and $i \neq j \leq n$, we set r_{ijp} to be 1 if $k_i^p \geq 1$, $k_j^p \geq 1$ and $t_{i1}^p \leq t_{j1}^p$, and 0 otherwise. In other words, r_{ijp} is 1 if and only if both O_i and O_j were recorded for the patient and the *first* record of observation i was before, or at the same time as, the first record of observation j. Again, for each fixed $i \neq j$, we think of the r_{ijp}'s as samples of a random variable X_{ij} and our goal is to predict the class label Y given $(X_{ij})_{i \neq j}$.

The use of first record (rather than all records) is a major difference in the way sequential patterns are mined in tSPM to handle the repeated problem list entries.

It is important to emphasize that we call the sequential pairs in the tSPM approach transitive sequences as they embody distinctive modifications to the conventional sequential pattern mining (SPM). Imagine a sequential pattern where observation A happened right before B, and B happened right before C ($A \rightarrow B \rightarrow C$). SPM mines subsequences $A \rightarrow B$ and $B \rightarrow C$. To account for the potential biases in EHRs, the transitive sequencing algorithm mines subsequences $A^* \rightarrow B^*$, $B^* \rightarrow C^*$, but also $A^* \rightarrow C^*$ from the sequence $A^* \rightarrow B^* \rightarrow C^*$, where A^*, B^*, and C^* are the first records of A, B, C in the database.

3.3 Dimensionality Reduction

Since the numbers of pairs (O_i, O_j) with $i \neq j \leq n$ is exactly $\frac{n(n-1)}{2}$, the number of sequential features is quadratic in the number of observations. This leaves us with a highly dimensional vector of representations. The usual approach when using SPM is to keep only the first n most frequent features, where n is a hyperparameter. We call this approach FSPM.

We attempt to improve on this heuristic with a formal dimensionality reduction procedure that aims to minimize sparsity and maximize relevance (MSMR). To minimize sparsity, we removed any feature that has a prevalence smaller than one percent. On the remaining features we compute the empirical mutual information using the entropy of the empirical probability distribution [10,15].

Mutual information provides a measurement of the mutual dependence between two random variables, which unlike most correlation measures can capture non-linear relationships [10,20]. We ranked the data representations based on their mutual information with the labeled outcome and kept the first 10'000 (in ties, we used prevalence to determine the ranking).

We further scrutinized the relevance using joint mutual information (JMI) [25]. The algorithm starts with a set S containing the top feature according to mutual information, then iteratively adds to S the feature X maximizing the *joint mutual information score*

$$J_{jmi}(X) = \sum_{X^* \in S} I(XX^*; Y)$$

Here, $I(Z;Y)$ denotes the mutual information between random variables Z and Y (a measure of the information shared by Z and Y—it can be expressed as the entropy of Z minus the entropy of Z given Y). The random variable XX^* is simply the random variable corresponding to the joint distribution of X and X^*. In the end, we select the first 40–500 features that were added to the set S.

The idea of using the joint mutual information score (as opposed to just the mutual information) is that it also takes into account the redundancy between the features: two features could each be highly relevant on their own, but also be strongly correlated. On the other hand computing joint mutual information is quadratic in the number of features, hence the initial dimensionality reduction using mutual information.

Joint mutual information belongs to a large zoo of feature selection scores J, see for example [7] for a theoretical analysis and some experimental results suggesting that JMI is a reasonable default choice. They compare the performance of 9 different feature selection scores across 22 data sets, and suggest JMI as the "best trade-off [...] of accuracy and stability". It is one of few scores that satisfies three theoretical criteria: reference to a conditional redundancy term, keeping the relative size of the redundancy term from swamping the relevancy term, and whether it is a low-dimensional approximation, hence usable with small sample sizes.

4 Experimental Setting

We ask three questions. The first question involves examining the efficiency of temporal representation mining algorithms; (1) can a classifier trained on transitive sequences (tSPM) improve true disease prediction over the classifier trained on traditional sequences (SPM)? As we discussed, a possible shortfall of traditional sequential pattern mining methods in clinical domains is the use of frequency-based criterion to mine temporal representations. The second question aims to examine the proposed MSMR dimensionality reduction algorithm; (2) does an entropy-based criteria improve the frequency-based criterion for feature selection? Finally, the third question examines whether together the proposed temporal representation mining and dimensionality reduction can improve true

disease prediction; (3) Does tSPM+MSMR provide an overall better prediction that the Frequent traditional sequential pattern mining (FSPM)?

We used EHR data from the Mass General Brigham Biobank on six diseases: congestive heart failure (CHF), type 1 and 2 diabetes mellitus (T1DM and T2DM), rheumatoid arthritis (RA), chronic obstructive pulmonary disease (COPD), and ulcerative colitis (UC). In each disease cohort, the medical records end with an encounter that includes International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code for the respective disease. We excluded data from the last encounter to predict. There is no time constraint for the next encounter that we aim to predict. For each of the diseases, we had a random subset of patients for whom we had curated gold-standard labels via in-depth expert review of clinical notes. The gold-standard labels identified if the patient truly had (or did not have) the disease. We used the gold-standard labels as the test sets. For the rest of the patients in each cohort, we had silver-standard labels that were curated by ML algorithms for training. The use of data for this study was approved by the Mass General Brigham Institutional Review Board (2017P000282). Table 1 shows the basic data statistics for each disease cohort.

Table 1. Basic statistics from the disease cohorts.

Disease	Cohort population	Unique records	Average depth	Average age	Test set size
CHF	6,857	25,480	16.24 yrs	68 yrs	49
COPD	5,107	29,880	9.14 yrs	60 yrs	61
RA	4,015	20,315	7.83 yrs	52 yrs	72
T1DM	3,107	23,001	8.59 yrs	57 yrs	58
T2DM	9,500	27,099	7.62 yrs	55 yrs	74
UC	1,560	14,682	6.68 yrs	46 yrs	45

From each representations, a nested set of the top 50, 100, 200, and 400 features obtained from the MSMR algorithm are used for prediction. We applied Logistic Regression with L1 regularization to the training sets to predict the true diagnosis of the target disease (Dx_i) at the next encounter, using bootstrap cross-validation. To evaluate the classifiers, we utilized the gold-standard labels on the test sets. We used the area under the receiver operating characteristic curve (AUC ROC) to compare the algorithms' prediction performances. We also compared the Youden's J statistic [26], that aims to capture both sensitivity and specificity in a single statistic. We computed the all possible Youden's J statistics on the receiver operating characteristic curve, and used the best J statistic a classifier had to offer.

5 Results

As expected, the number of tSPM sequences were about 10-fold the number of SPM sequences. Table 2 shows number of unique sequential patterns mined by

Table 2. Representations mined by each algorithm.

Disease	SPM representations	tSPM representations
CHF	3,555,577	30,969,742
COPD	3,520,689	37,137,257
RA	1,863,405	18,310,561
T1DM	2,209,541	22,294,391
T2DM	3,480,550	29,397,627
UC	863,097	9,062,784

the SPM and tSPM algorithms. The prediction performances from each algorithm are presented in Table 3. Results (Table 3) showed that on average, transitive sequential pattern mining (tSPM) improved equivalent SPM metrics by between 6% to 15% in AUC ROC and 7% to 10% in Youden's J statistics, regardless of the feature selection approach. Frequent transitive sequential representations mined in the FtSPM improved AUC ROC by 6 and Youden's J by 7% compared with the frequent sequential representations in the FSPM algorithm. Using the MSMR dimensionality reduction approach, the tSPM representations offered an average 15% improvement in AUC ROC and 10% in Youden's index over the SPM representations. Across the six diseases, there was some variability in the magnitude of improvements transitive sequencing provided over traditional sequencing. For instance, in COPD and RA, the difference between frequency-based tSPM (FtSPM) and SPM (FSPM) in AUC ROC was close to zero, while this performance difference was 25% in T2DM between the MSMR-based algorithms. Nevertheless, the overall improvement from using tSPM is supported by the results.

We also found that in the traditional sequencing (SPM), the improvement in prediction is almost non-existent. That is, the MSMR algorithm did not offer any benefit over using the most frequent traditional sequential representations. For transitive sequences, however, the MSMR algorithm provided an average of 9% improvement in the AUC ROC and 4% improvement in Youden's J over the frequency-based criterion. Across the six diseases, there was some level of variability in both magnitude and direction (positive/negative change) of the results that impacted the averages. For example, in COPD and RA, the MSMR algorithm for dimensionality reduction decreased the AUC ROC and Youden's J statistic obtained from the SPM algorithms. This resulted in the minimal difference in the effect of MSMR on representations mined from the traditional sequential pattern mining. Variability was less of an issue in the overall effectiveness of the MSMR algorithm on transitive sequences. In four of the six diseases, the transitive sequential representations selected by the MSMR algorithm significantly improved the classifiers' AUC ROC over using the most frequent transitive sequential representations.

Regarding the third question, we found that, on average, the proposed tSPM+MSMR approach improved the AUC ROC by 15% and Youden's J statis-

Table 3. AUC ROC and Youden's J statistics across diseases and by algorithms.

		AUC ROC	FSPM	FTSPM	SPM+MSMR	Youden's J statistic	FSPM	FTSPM	SPM+MSMR
CHF	FSPM	0.672				0.682			
	FtSPM	0.689	3%			0.671	-2%		
	SPM+MSMR	0.723	8%	5%		0.709	4%	6%	
	tSPM+MSMR	0.826	23%	20%	14%	0.792	16%	18%	12%
COPD	FSPM	0.751				0.726			
	FtSPM	0.740	-1%			0.767	6%		
	SPM+MSMR	0.644	-14%	-13%		0.682	-6%	-11%	
	tSPM+MSMR	0.788	5%	6%	22%	0.702	-3%	-8%	3%
RA	FSPM	0.692				0.671			
	FtSPM	0.694	0%			0.784	17%		
	SPM+MSMR	0.659	-5%	-5%		0.654	-3%	-17%	
	tSPM+MSMR	0.793	15%	14%	20%	0.759	13%	-3%	16%
T1DM	FSPM	0.888				0.846			
	FtSPM	0.926	4%			0.904	7%		
	SPM+MSMR	0.865	-3%	-7%		0.875	3%	-3%	
	tSPM+MSMR	0.904	2%	-2%	4%	0.904	7%	0%	3%
T2DM	FSPM	0.596				0.626			
	FtSPM	0.728	22%			0.705	13%		
	SPM+MSMR	0.647	9%	-11%		0.671	7%	-5%	
	tSPM+MSMR	0.808	36%	11%	25%	0.796	27%	13%	19%
UC	FSPM	0.782				0.783			
	FtSPM	0.832	6%			0.813	4%		
	SPM+MSMR	0.804	3%	-3%		0.800	2%	-2%	
	tSPM+MSMR	0.849	9%	2%	6%	0.863	10%	6%	8%
Mean	FtSPM		6%$^{\intercal}$				7%$^{\intercal}$		
	SPM+MSMR		0%‡	-6%			1%‡	-5%	
	tSPM+MSMR		15%*	9%‡	15%$^{\intercal}$		12%*	4%‡	10%$^{\intercal}$

$^{\intercal}$ question 1: transitive sequencing (tSPM) vs. traditional sequencing (SPM)?
‡ question 2: MSMR vs. the frequency-based criterion
* question 3: transitive sequencing with MSMR vs. traditional frequency-based sequencing

tic by 12% over the FSPM. The only exception was in COPD where Youden's J decreased by 3% as compared with the FSPM (AUC ROC improved by 5%), in which case the frequent transitive sequential patterns provided the best overall performance. Given the benchmark of frequency-based sequential pattern mining (FSPM), both performance indices improved consistently by utilizing the MSMR driven transitive sequential patterns for predicting the diagnosis.

6 Conclusion

In this work, we argue that discrete clinical data in EHRs are fundamentally different from transaction data (for which the SPM algorithm and its often used frequency-based feature selection criteria were developed). We proposed a transitive sequential pattern mining (tSPM) algorithm as well as a dimensionality reduction algorithm (MSMR) to construct and apply temporal representations from discrete clinical data into standard Machine Learning. Our results showed that the tSPM representations improve prediction performance over the traditional sequential pattern mining (SPM). In other words, we showed that the assumption of transitivity improves prediction power in sequential pattern mining with EHR data. More research is needed to evaluate other use cases of transitive sequential pattern mining. We also showed that selecting features using the MSMR algorithm improved prediction using tSPM representations. Another novelty of this research is in its prediction task, where we aim to predict a verified disease diagnosis record. This is different and possibly harder than the prior clinical prediction tasks, such as [9], that aim to predict the next diagnosis record. We believe that it is naive to assume that the diagnosis records in EHRs are always correct and reflect clinical practice. The tSPM and MSMR algorithms together provide methodological pathway to making sense of data in electronic health records.

Acknowledgements. The work in this paper was supported by NIH grant R01-HG009174. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

References

1. Agrawal, R., Srikant, R., et al.: Mining sequential patterns. In: ICDE, vol. 95, pp. 3–14 (1995)
2. Albers, D.J., Hripcsak, G.: Estimation of time-delayed mutual information and bias for irregularly and sparsely sampled time-series. Chaos, Solitons Fractals **45**(6), 853–860 (2012)
3. Ayres, J., Flannick, J., Gehrke, J., Yiu, T.: Sequential pattern mining using a bitmap representation. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2002, pp. 429–435. ACM, New York (2002)
4. Batal, I., Cooper, G.F., Fradkin, D., Harrison, J., Moerchen, F., Hauskrecht, M.: An efficient pattern mining approach for event detection in multivariate temporal data. Knowl. Inf. Syst. **46**(1), 115–150 (2015). https://doi.org/10.1007/s10115-015-0819-6
5. Batal, I., Valizadegan, H., Cooper, G.F., Hauskrecht, M.: A temporal pattern mining approach for classifying electronic health record data. ACM Trans. Intell. Syst. Technol. **4**(4) (2013)

6. Berlingerio, M., Bonchi, F., Giannotti, F., Turini, F.: Mining clinical data with a temporal dimension: a case study. In: 2007 IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007), pp. 429–436, November 2007
7. Brown, G., Pocock, A., Zhao, M.J., Luján, M.: Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. J. Mach. Learn. Res. **13**, 27–66 (2012)
8. Choi, E., et al.: Multi-layer representation learning for medical concepts. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2016, pp. 1495–1504. Association for Computing Machinery, New York, August 2016
9. Choi, E., Bahadori, M.T., Sun, J., Kulas, J., Schuetz, A., Stewart, W.: RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 29, pp. 3504–3512. Curran Associates, Inc. (2016)
10. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Hoboken (2012)
11. Hripcsak, G., Albers, D.J., Perotte, A.: Exploiting time in electronic health record correlations. J. Am. Med. Inform. Assoc. **18**(Suppl 1), i109–15 (2011)
12. Johnson, A., Pollard, T., Shen, L., et al.: MIMIC-III, a freely accessible critical care database. Sci. Data **3**, 160035 (2016). https://doi.org/10.1038/sdata.2016.35
13. Lee, J.M., Hauskrecht, M.: Recent context-aware LSTM for clinical event time-series prediction. In: Riaño, D., Wilk, S., ten Teije, A. (eds.) AIME 2019. LNCS (LNAI), vol. 11526, pp. 13–23. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21642-9_3
14. Mabroukeh, N.R., Ezeife, C.I.: A taxonomy of sequential pattern mining algorithms. ACM Comput. Surv. **43**(1), 41 p. (2010). Article 3. https://doi.org/10.1145/1824795.1824798
15. Meyer, P.E.: Information-theoretic variable selection and network inference from microarray data. Ph.D. thesis, Université Libre de Bruxelles (2008)
16. Moskovitch, R., Choi, H., Hripcsak, G., Tatonetti, N.: Prognosis of clinical outcomes with temporal patterns and experiences with one class feature selection. IEEE/ACM Trans. Comput. Biol. Bioinform. **14**(3), 555–563 (2017)
17. Moskovitch, R., Polubriaginof, F., Weiss, A., Ryan, P., Tatonetti, N.: Procedure prediction from symbolic electronic health records via time intervals analytics. J. Biomed. Inform. **75**, 70–82 (2017)
18. Moskovitch, R., Shahar, Y.: Classification-driven temporal discretization of multivariate time series. Data Min. Knowl. Disc. **29**(4), 871–913 (2014). https://doi.org/10.1007/s10618-014-0380-z
19. Orphanou, K., Dagliati, A., Sacchi, L., Stassopoulou, A., Keravnou, E., Bellazzi, R.: Incorporating repeating temporal association rules in Naïve Bayes classifiers for coronary heart disease diagnosis. J. Biomed. Inform. **81**, 74–82 (2018)
20. Paninski, L.: Estimation of entropy and mutual information. Neural Comput. **15**(6), 1191–1253 (2003)
21. Perer, A., Wang, F., Hu, J.: Mining and exploring care pathways from electronic medical records with visual analytics. J. Biomed. Inform. **56**, 369–378 (2015)
22. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Others: learning representations by back-propagating errors. Cogn. Model. **5**(3), 1 (1988)
23. Stacey, M., McGregor, C.: Temporal abstraction in intelligent clinical data analysis: a survey. Artif. Intell. Med. **39**(1), 1–24 (2007)

24. Sun, W., Rumshisky, A., Uzuner, O.: Temporal reasoning over clinical text: the state of the art. J. Am. Med. Inform. Assoc. **20**(5), 814–819 (2013)
25. Yang, H., Moody, J.: Data visualization and feature selections: new algorithms for non-Gaussian data. In: Advances in Neural Information Processing Systems, vol. 12 (1999)
26. Youden, W.J.: Index for rating diagnostic tests. Cancer **3**(1), 32–35 (1950)
27. Zaki, M.J.: Parallel sequence mining on shared-memory machines. J. Parallel Distrib. Comput. **61**(3), 401–426 (2001)