# A 3-Window Framework for the Discovery and Interpretation of Predictive Temporal Functional Dependencies

Beatrice Amico(✉) and Carlo Combi

Department of Computer Science, University of Verona, Verona, Italy
{beatrice.amico,carlo.combi}@univr.it

**Abstract.** Clinical databases collect large volume of data. Relationships and patterns within these data could provide new medical knowledge. Temporal data mining has as major scope the discovery of potential hidden knowledge from large amounts of data, offering the possibility to identify different features less visible or hidden to common analysis techniques. In this work, we show how temporal data mining, precisely mining of functional dependencies, can be fruitfully exploited to improve clinical prediction. To develop an early prediction model, a window-based data aggregation approach could be a good starting point, therefore we introduce a new temporal framework based on three temporal windows designed to extract predictive information. In particular, we propose a methodology for deriving a new kind of predictive temporal patterns. We exploit the predictive aspect of the approximate temporal functional dependencies, formally introducing the concept of Predictive Functional Dependency (PFD), a new type of approximate temporal functional dependency. We discuss some first results we obtained by pre-processing and mining ICU data from the MIMIC III database, focusing on functional dependencies predictive of Acute kidney injury (AKI).

**Keywords:** Temporal data mining · Predictive patterns · Functional dependencies · Temporal windows

## 1 Introduction

The increasing use and availability of longitudinal electronic data provide the opportunity to discover new knowledge from multivariate, time-oriented data, by using various data mining methods.

Temporal data mining in medicine has been receiving considerable attention since it provides a way of revealing useful information hidden in the clinical data, extracting different temporal patterns. The analysis of such healthcare/medical data collections could greatly help to observe the health conditions of the population and extract useful information that can be exploited in the assessment

Student paper submission.


M. Michalowski et al. (Eds.): AIME 2022, LNAI 13263, pp. 299–309, 2022.
https://doi.org/10.1007/978-3-031-09342-5_29

of healthcare/medical processes [2]. For example, together with temporal data mining, clinical data sources enable us to rapidly generate prediction models for thousands of clinical problems, for identifying diagnoses, speed medical processes, risks prevention, prediction of mortality, and risk stratification [11].

Particularly, prediction of medical events, such as clinical procedures, is essential for preventing disease, understanding disease mechanism, and increasing patient quality of care. When we talk about prediction, we associate the well-known machine learning techniques and the already known black box problem. In the last two decades, several supervised learning methods have been introduced [1], but often, it is not possible to understand why machine learning algorithms are proposing specific predictions. On the contrary, temporal patterns represent an explainable way to study the intrinsic data dependencies, to allow physicians to focus on the most interesting and relevant discovered rules.

According to this scenario, the main novelty of this paper is the proposal of an original temporally-oriented data mining technique for the prediction of clinical diseases. Therefore, we propose a new type of functional dependencies, the *approximate predictive functional dependencies* (APFDs). They are evaluated within a new temporal framework based on three temporal windows: observation window, waiting window, prediction window.

The paper is organized as follows. In Sect. 2 we briefly describe some related work, relevant to the topic discussed in this paper. In Sect. 3 we introduce a new temporal framework based on three temporal windows: observation window, waiting window, prediction window. Then, we define the entire framework for the approximate predictive functional dependencies (APFDs), introducing two new error measures. In Sect. 4 we detail the first application of this framework on real clinical data from patients hospitalized in Intensive Care Units, using MIMIC III [6]. In Sect. 5 we draw some conclusions and discuss possible future work.

## 2 Related Work

In the context of temporal data mining, various techniques are applied to time-oriented data to discover knowledge about relationships among different raw data and abstract concepts, in which the temporal dimension is treated explicitly.

Associations discovery is one of the most common Data Mining (DM) techniques used to extract interesting knowledge from large datasets. Association rules enable the identification of correlations between the elements of a dataset. In literature, we find different methods to mine temporal association rules (TARs) aiming at providing a greater predictive and descriptive potential in different contexts, with a high number of contributions in the context of medicine and healthcare [13].

Mining time intervals data is another interesting research field, especially for the extraction of Time Intervals Related Patterns (TIRPs). In [4], the authors introduce TIRPClo, an efficient algorithm for the discovery of frequent closed TIRPs, a compact subset of all the frequent TIRPs based on which their complete

information can be revealed. In addition, it is possible to use patterns as features for classification. For example, in [11], the authors propose a framework for discovering TIRPs only from the cohort of patients having the outcome event. The results showed that representing the TIRPs using the horizontal support outperformed the binary and mean duration representations.

Among temporal abstractions, we also find the trend abstractions that focus on detecting changes in the temporal evolution. In [9], starting from the concept of Trend-Event Pattern [10] and moving through the concept of *prediction*, the authors propose a new kind of predictive temporal patterns, namely Predictive Trend-Event Patterns (PTE-Ps). The framework aims to combine complex temporal features to extract a compact and non-redundant predictive set of patterns composed by such temporal features.

Another type of temporal pattern is functional dependency. In literature there are different extensions, temporal functional dependency (TFD) [3], approximate functional dependency (AFD) [8], and approximate temporal functional dependencies (ATFDs) [2]. Temporal functional dependencies (TFDs) add a temporal dimension to classical functional dependencies (FDs) to deal with temporal data. In [3], Combi et al. propose a new formalism for the representation of TFDs, involving multiple time granularities. They identify four relevant classes of TFDs: Pure temporally grouping, Pure temporally evolving, Temporally mixed, and Temporally hybrid. Moving on, approximate functional dependency (AFD) derives from the concept of plain FD. Given a relation $r$ where an FD holds for most of the tuples in $r$, we may identify some tuples for which that FD does not hold. In [8], Kivinen and Mannila introduce three measures, known as $G_1$, $G_2$ and $G_3$ considering the number of violating couples of tuples, the number of tuples that violate the functional dependency, and finally the minimum number of tuples in $r$ to be deleted for the FD to hold. In [2], the authors propose the concept of approximate temporal functional dependencies (ATFDs), which are defined and measured either on temporal granules or on sliding windows, considering the psychiatry and pharmacovigilance domains.

## 3 Predictive Functional Dependencies

In this section, we describe a new temporal framework and detail definitions to mine the approximate predictive functional dependencies (APFDs).

### 3.1 A 3-Window Framework for the Interpretation of Predictive Temporal Data

As previously said, data mining in medicine has great potential for discovering hidden patterns in data sets from the medical domain. In such conditions, we face the challenge of the extraction of hidden predictive information from large databases. To develop an early prediction model, a window-based data aggregation approach could be a good starting point. As far as we know, the prediction models exploit the use of two-time windows. The first one, called data collection

or observation window, concerns the collection of data that allows us to predict the problem of interest, and the second one, the prediction window, in which the key event occurs. Here, we generalize an approach based on three (possibly moving) time windows. So that the prediction becomes effective, it is necessary that clinicians can act before a clinical decline has occurred by: (i) delivering insights on preventable conditions; (ii) offering contextual information to help clinical decision-making; (iii) being generally applicable across a different cohort of patients. The anticipation of a future event is obviously relevant, but the more significant facet is the time needed to anticipate a future event, that is a key aspect, especially in medicine. Acknowledge these prerequisites, in this paper we propose a framework based on three windows: (i) an observation window (OW); (ii) a waiting window (WW); and (iii) a prediction window (PW). The OW is considered as a time interval, where the information is collected, and ends when an event of interest occurs. The WW is held to be the minimum time interval required to act in order to prevent the event in the prediction window. Finally, the PW, the time interval when the predicted event occurs.

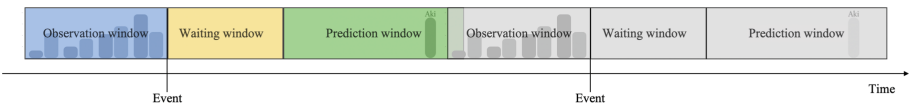



**Fig. 1.** 3 window-based framework

### 3.2 Towards the Definition of Predictive Functional Dependencies

A FD is composed of the antecedent (X) and the consequent (Y). Informally, for all the couples of tuples $t$ and $t'$ showing the same value(s) on X, the corresponding value(s) on Y for those tuples are identical. In our specific case, we describe the antecedent as a set of attributes ordered on VT, called *predictive attributes*, and the consequent defined as the *predicted attribute*.

Let us consider a database schema DB as a set of temporal relational schemas $\{R_1, ..., R_n\}$ and a set of corresponding instances $\{r_1, ..., r_n\}$. Any schema $R_i$ has attributes $ZU_i \cup \{VT\}$, where $\forall R_i, R_j$ with $i \neq j$ it holds $U_i \cap U_j = \emptyset$. $U_i$ is a set of attributes representing properties of a patient, which is identified by attributes $Z$ (hereinafter patient identification attributes). VT is the attribute representing the temporal dimension of the tuples.

Given a set of relations $\{r_1, ..., r_m\}$ according to the data schema $DB$, a **State expression** with schema $ZU \cup \{VT\}$ where $U \equiv U_i U_j ... U_m$ is defined as

$$SE \equiv r | r_i \bowtie r_j \bowtie ... r_m$$

**Definition 1 (K-State evolution expression (KSE)).** *Given a set of State Expressions* $\{SE_\alpha(ZU_\alpha \cup \{VT\}), ..., SE_\omega(ZU_\omega \cup \{VT\})\}$*, a* ***K-State evolution expression*** *with schema* $Z\overline{U}^0_\alpha \overline{U}^1_\beta .. \overline{U}^k_\kappa \cup \{\overline{VT}^0_\alpha, \overline{VT}^1_\beta, .., \overline{VT}^k_\kappa\}$ *is defined as:*

$$KSE \equiv \Theta(\overline{SE}^0_\alpha, \overline{SE}^1_\beta, ...\overline{SE}^k_\kappa) \text{ with } \overline{VT}^0 < \overline{VT}^1 < .. < \overline{VT}^k$$
$$\text{and } \overline{SE}^0_\alpha = \rho_{U_\alpha, VT \to \overline{U}^0_\alpha, \overline{VT}^0} SE_\alpha, \ \overline{SE}^1_\beta = \rho_{U_\beta, VT \to \overline{U}^1_\beta, \overline{VT}^1} SE_\beta, \ \ldots \quad (1)$$

*Function $\Theta$ allows different evolutions of the same State Expressions. For example, it can represent suitable join of different State Expressions according to the patient attributes, possibly using the same attribute at different states. It can allow also the join of tuples at distance k (for k = 1, it joins pairs of consecutive corresponding tuples) or allow one to join pairs of successive (concerning the values they take on attribute VT) tuples.*

Let us now move to consider the attributes to be considered in the prediction, i.e., boolean attributes representing the presence/absence of a pathological state. We can join such attributes to a KSE to build a K-State Prediction Expression.

**Definition 2 (K-State Prediction Expression (KSPE)).** *Given a schema $R \in DB$, with attributes $ZU_p \cup \{VT\}$, a K-State Prediction Expression (KSPE) is defined as:*

$$\sigma_{\substack{0 \le (\overline{VT}^k - \overline{VT}^0) < OW \wedge \\ (\dot{VT} - \overline{VT}^0 \ge OW + WW) \wedge \\ (\dot{VT} - \overline{VT}^0 < OW + WW + PW)}} KSE \bowtie \dot{R}$$
$$\text{and } \dot{R} = \rho_{U_p, VT \to \dot{U}_p, \dot{VT}} R \quad (2)$$

Analogously to the previous renaming, $\dot{U}_p$ represents the overall renaming of the attribute set of $\dot{R}$, where $\dot{R}$ is the patient relation.

**Definition 3 (Predictive Functional Dependency (PFD)).** *Given a K state prediction expression KSPE with schema $Z\overline{U}^0_\alpha \overline{U}^1_\beta .. \overline{U}^k_\kappa \dot{U}_p \cup \{\overline{VT}^0, \overline{VT}^1, .., \overline{VT}^k, \dot{VT}\}$, a Predictive Functional Dependency is a FD, of the following form:*

$$\overline{X}^h \overline{S}^i ... \overline{W}^j \to \dot{B} \text{ with } 0 \le h < i ... < j \le k$$
$$\text{where } \overline{X}^h \subseteq \overline{U}^h, \overline{S}^i \subseteq \overline{U}^i, \overline{W}^j \subseteq \overline{U}^j \text{ and } \dot{B} \in \dot{U}_p \text{ is the predicted (Boolean) attribute.} \quad (3)$$

Table 1 represents an example of a general KSPE, where the PFD reported below holds.

**Table 1.** A KSPE where $\overline{Resprate}^0, \overline{SpO_2}^1, \overline{Drug}^2 \rightarrow A\dot{K}I$ holds.

| Tuple # | Patient | $\overline{RespRate}^0$ | $\overline{VT}^0$ | $\overline{SpO_2}^1$ | $\overline{VT}^1$ | $\overline{Drug}^2$ | $\overline{VT}^2$ | $A\dot{K}I$ | $\dot{VT}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Mark Jones | High | 1 | Low | 2 | Aspirin | 3 | F | 4 |
| 2 | Mark Jones | High | 3 | Medium | 4 | Indapamide | 5 | T | 6 |
| 3 | Mark Jones | Medium | 5 | Medium | 6 | Metolazone | 7 | T | 8 |
| 4 | Mark Jones | High | 8 | Medium | 9 | Indapamide | 10 | T | 11 |
| 5 | Viola Thompson | Low | 2 | High | 3 | Aspirin | 4 | F | 6 |
| 6 | Viola Thompson | Low | 3 | Medium | 4 | Indapamide | 5 | F | 7 |
| 7 | Viola Thompson | Low | 4 | Low | 5 | Aspirin | 2 | F | 8 |
| 8 | Paul Walker | Medium | 1 | High | 2 | Ibuprofen | 3 | T | 5 |
| 9 | Paul Walker | Medium | 1 | High | 2 | Sulindac | 3 | T | 5 |
| 10 | Paul Walker | Medium | 2 | Medium | 3 | Indapamide | 5 | T | 8 |

### 3.3 Discovering Approximate PFDs

The term approximation is about the approximate satisfaction of a normal PFD $\overline{X}^h \overline{S}^i ... \overline{W}^j \rightarrow \dot{B}$.

An APFD $f$ requires the PFD to be satisfied by most tuples of temporal relation $r$. It allows a very small portion of tuples of $r$ to violate the dependency. If it is less than or equal to the satisfaction threshold $\varepsilon$, f is approximately satisfied on r. Several methods have been proposed to calculate the error measure. In the context of PFDs, we reconsider a measure proposed in [8] and we introduce two other error measures, specifically tailored to the predictive purpose of APFDs.

Considering a general KSPE $w$ over a schema $Z\overline{U}_\alpha^0 \overline{U}_\beta^1 .. \overline{U}_\kappa^k \dot{U}_p \cup \{\overline{VT}^0, \overline{VT}^1, .., \overline{VT}^k, \dot{VT}\}$ and any set $s \subseteq w$, where the PFD holds, we define three error measures. We start from $G_3$ that considers the minimum number of tuples in r to be deleted to obtain a relation where the PFD holds. This measure is defined as follows:

**Definition 4 (Error measure $G_3$).** *Given a PFD expressed as in Definition 3, the error measure G3 is expressed as:*

$$G_3 = |w| - max\left\{|s| \mid s \subseteq w \wedge s \models \overline{X}^h \overline{S}^i ... \overline{W}^j \rightarrow \dot{B}\right\} \tag{4}$$

The related *scaled measurement* $g_3$ is defined as $g_3 = G_3/|w|$.

Secondly, we introduce $H_3$ that considers the maximality focused on the number of patients that we accept to loose for the sake of the PFD. This maximality permits to delete patients with a very low number of tuples, which could generate noise in our dataset. $H_3$ can be formalized as follows:

**Definition 5 (Error measure $H_3$).** *Given a PFD expressed as in Definition 3, the error measure H3 is expressed as:*

$$H_3 = |\{t\,[Z] \mid \exists t \in w\}| - \max_s \left\{\left|\left\{t\,[Z] \mid \exists t \in s \wedge s \subseteq w \wedge s \models \overline{X}^h \overline{S}^i ... \overline{W}^j \rightarrow \dot{B}\right\}\right|\right\} \tag{5}$$

The related *scaled measurement* $h_3$ is defined as $h_3 = H_3/|\{t\,[Z] \mid \exists t \in w\}|$.

Finally, we can formalize $J_3$ that focuses on the number of tuples for each patient we accept to delete in order to satisfy the PFD. This error is very useful to ensure to maintain enough information for each patient, ensuring to be consistent.

**Definition 6 (Error measure $J_3$).** *Given a PFD expressed as in Definition 3, the error measure $J_3$ can be formalized as follows:*

$$J_3 = \max_{(v \in \{t[Z] | t \in s\})} \{ |\{t[Z] | t \in w \wedge t[Z] = v\}| - |\{t[Z] | t \in s \wedge t[Z] = v\}| \} \quad (6)$$

The related *scaled measurement* $j_3$ weights each term of $J_3$ with respect to the number of tuples in $w$ having value $v$ for $t[Z]$.

After the introduction of these three error measures, we are now ready to define the approximate predictive functional dependency as follows:

**Definition 7 (Approximate Predictive Functional Dependency (APFD)).** *Let $w$ be a relationship over a K-state prediction expression: let $X, \dot{B} \subseteq R$ be sets of attributes of $R$. Relation $w$ fulfills the functional dependency $\overline{X}^h \overline{S}^i ... \overline{W}^j \xrightarrow{\varepsilon} \dot{B}$ (written as $w \models \overline{X}^h \overline{S}^i ... \overline{W}^j \xrightarrow{\varepsilon} \dot{B}$) if $G(\overline{X}^h \overline{S}^i ... \overline{W}^j \xrightarrow{\varepsilon} \dot{B}, w) \leq \varepsilon$, where $\varepsilon =< \varepsilon_g, \varepsilon_h, \varepsilon_j >$ and $0 \leq \varepsilon < 1$ is the maximum acceptable error defined by the user. $G$ is the corresponding error of the previously introduced measures.*

Among the several APFDs that can be detected over a relation $w$, the minimal APFD is particularly interesting. We thus define the minimal APFD as follows:

**Definition 8 (Minimal APFD).** *An APFD $\overline{X}^h \overline{S}^i ... \overline{W}^j \xrightarrow{\varepsilon} \dot{B}$ is minimal for $w$, if $w \models \overline{X}^h \overline{S}^i ... \overline{W}^j \xrightarrow{\varepsilon} \dot{B}$ and $\forall \overline{V} \subset \overline{X}^h \overline{S}^i ... \overline{W}^j$ we have that $w \not\models \overline{V} \xrightarrow{\varepsilon} \dot{B}$.*

As an example, according to the data in Table 1, the APFD $\overline{SPO_2}^1, \overline{Drug}^2 \xrightarrow{\varepsilon} \dot{AKI}$ holds with $\varepsilon = 0.1$. Indeed it is enough the delete tuple 6, to have the corresponding PFD satisfied.

## 4 Deriving APFDs: an Experimental Evaluation

### 4.1 Dataset and Data Transformation

To illustrate the relevance and the potential meaning of our proposal, we consider a real-world example from the domain of Intensive Care Unit (ICU) with patients suffering from Acute Kidney Injury (AKI). Acute Kidney Injury is a frequent clinical problem, associated with a mortality of 50–80%, characterized by a sudden loss of the ability of the kidneys to excrete wastes, concentrate urine, store electrolytes, and maintain fluid balance [12]. A ground-truth label for the diagnosis of AKI is added using the internationally accepted KDIGO criteria [7]. A patient receives the diagnosis of AKI if one of the following criteria is valid: (i) an increase in serum creatinine by $\geq 0.3$ mg/dl ( $\geq 26.5\,\mu$mol/l) within 48 h, (ii) an increase in serum creatinine to $\geq 1.5$ times baseline within the previous 7 d and (iii) a urine volume $\leq 0.5$ ml/kg/h for 6 h.

Our methodology is applied using the MIMIC III (Medical Information Mart for Intensive Care) dataset [6], a freely accessible relational database of de-identified patients, hospitalized in the intensive care units at Beth Israel Deaconess Medical Center between 2001 and 2012.

An ETL (Extract, Transform, Load) process is necessary to transform the MIMIC-III raw data in a form useful for mining the APFDs. To obtain SEs, we use four tables. *Prescriptions* provides information about the administered medications, for a given patient. We mainly consider the following categories: diuretics, Non-steroidal anti-inflammatory drugs (NSAID), radio contrast agents, and angiotensin. *Chartevents* contains information about vital signs measured at the bedside. We mainly consider diastolic blood pressure, glucose, heart rate and temperature. *D_items* is the reference table needed to label every measure related to a patient. *Labevents* was used to extract the information about serum creatinine and urine, useful for the diagnosis of AKI.

Because of the high number of measures, we applied three aggregate functions, minimum, maximum and average every specified time interval. Moreover, *Chartevents* and *Labevents* contain numerical variables, so we categorize the measures into "low, medium, high" according to clinical literature.

To summarize, there is a new table for each considered vital sign, which contains: icustay_id, minimum, maximum, and average of the measure and the valid time expressed by an interval. From these tables, we create different SEs in order to obtain a KSE. PFDs are retrieved from a KSPE, generated joining a KSE and the patient table $\dot{R}$, through an algorithm inspired by TANE [5], a popular approximate functional dependency detection algorithm.

### 4.2 Results

The final scope of these experiments is to find significant PDFs for the AKI diagnosis. We illustrate the results obtained with the following 3-window framework: an observation window of 48 h, where we collect all the measures related to each patient, a waiting window of 12 h where we do not consider any event, and then a prediction window where there is the onset of the illness according to one of the KDIGO criteria or the discharge from the ward with any criteria satisfied, of 72h hours. Starting from a cohort of 50711 patients, we extract subjects that receive a diagnosis or a discharge from the ICU, at least after 60 h from the admission in ICU, namely the duration of the observation and the waiting windows, thus obtaining 7930 patients. Among these patients, there are 6024 controls and 1906 cases. Starting from the literature [14], we considered six measures: creatinine, glucose, administered drugs, respiratory rate, diastolic pressure, oxygen saturation, and body temperature. We take into consideration the average of each measure every 6 h, a time interval sufficiently long to reduce the number of records, obtaining significant results.

We generate three different KSPEs based on three different $\Theta$ expressions:

- a KSE with four states where except for the first state composed of one measure, the other three states involve two measures recorded at the same valid time, temporally ordered, i.e., $\overline{VT}^0 < \overline{VT}^1 < \overline{VT}^2 < \overline{VT}^3$.

– a KSE with four states where, except for the first state composed of one measure, the other three states involve two measures recorded at the same valid time, temporally ordered where $\overline{VT}^k = \overline{VT}^{k-1} + 1$ for $k = 1, .., 3$.
– a KSE with seven states, temporally ordered.

We mine the minimal APFDs, where the approximation is given by $G_3$. In the three KSPEs, we obtain results using a margin error over 0.2. To achieve functional dependencies with more than one antecedent, we have to consider a margin error between 0.2 and 0.3. Getting closer to 0.3, the temporal states keep dropping until the results of functional dependencies consist of a single antecedent.

In Table 2, we report some functional dependencies regarding the three KSPEs. As we discussed before, it is possible to observe the drop which is inversely proportional to the epsilon value. For each KSPE, we select a PFD to show some value combinations peculiar of the cases, reported in Table 3.

**Table 2.** A list of PFDs valid on one of the three KPSEs, with a certain epsilon value.

| PFD | Epsilon | KSPE |
|---|---|---|
| $\overline{Creat}^0, \overline{Drug}^1, \overline{Diastolic}^1, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$ | 0,27 | KSPE 1 |
| $\overline{Creat}^0, \overline{Drug}^1, \overline{Diastolic}^1, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$ | 0,28 | KSPE 1 |
| $\overline{Diastolic}^1, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$ | 0,29 | KSPE 1 |
| $\overline{Creat}^0, \overline{Drug}^1, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$ | 0,25 | KSPE 2 |
| $\overline{Creat}^0, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3 \rightarrow A\dot{K}I$ | 0,28 | KSPE 2 |
| $\overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$ | 0,30 | KSPE 2 |
| $\overline{Creat}^0, \overline{Drug}^1, \overline{Diastolic}^2, \overline{RespRate}^3, \overline{Glucose}^4, \overline{Temperature}^5, \overline{SpO_2}^6 \rightarrow A\dot{K}I$ | 0,23 | KSPE 3 |
| $\overline{Creat}^0, \overline{Diastolic}^2, \overline{RespRate}^3, \overline{Glucose}^4, \overline{Temperature}^5 \rightarrow A\dot{K}I$ | 0,26 | KSPE 3 |
| $\overline{Creat}^0, \overline{Diastolic}^2, \overline{RespRate}^3, \overline{Temperature}^5 \rightarrow A\dot{K}I$ | 0,28 | KSPE 3 |

**Table 3.** (a) $\overline{Creat}^0, \overline{Drug}^1, \overline{Diastolic}^1, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3, \overline{SpO_2}^3 \rightarrow A\dot{K}I$, (b) $\overline{Creat}^0, \overline{RespRate}^2, \overline{Glucose}^2, \overline{Temperature}^3 \rightarrow A\dot{K}I$, (c) $\overline{Creat}^0, \overline{Diastolic}^2, \overline{RespRate}^3, \overline{Temperature}^5 \rightarrow A\dot{K}I$

| Num. | Value comb. | F | T |
|---|---|---|---|
| #1 | high, diu, low, med, med, low, low | 13 | 39 |
| #2 | high, nsaid, low, med, high, low, med | 0 | 23 |
| #3 | low, diu, med, high, high, high, med | 15 | 35 |
| #4 | high, diu, low, med, high, low, med | 30 | 50 |
| #5 | high, diu, med, high, med, low, med | 28 | 47 |

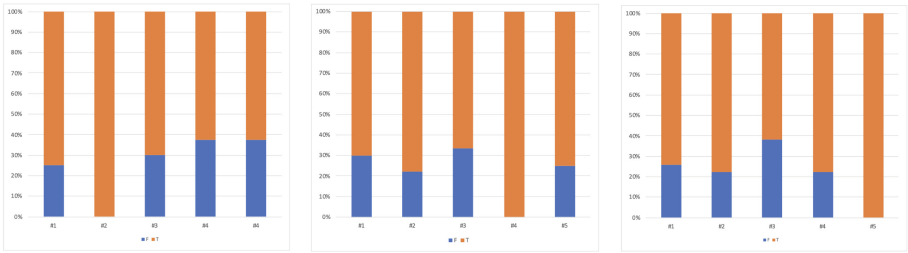
(a) Value combinations of KSPE 1

| Num. | Value comb. | F | T |
|---|---|---|---|
| #1 | high, high, med, low | 3 | 7 |
| #2 | low, high, med, high | 2 | 7 |
| #3 | high, med, high, low | 2 | 4 |
| #4 | low, med, high, med | 0 | 4 |
| #5 | high, med, low, med | 1 | 3 |

(b) Value combinations of KSPE 2

| Num. | Value comb. | F | T |
|---|---|---|---|
| #1 | high, low, high, low | 7 | 20 |
| #2 | high, low, med, low | 4 | 14 |
| #3 | med, high, high, med | 8 | 13 |
| #4 | med, med, low, high | 2 | 7 |
| #5 | med, high, high, low | 0 | 6 |

(c) Value combinations of KSPE 3



(a) Value comb. of KPSE 1 (b) Value comb. of KPSE 2 (c) Value comb. of KPSE 3

**Fig. 2.** Graphic representation of value combinations, reported in Table 3

## 5 Conclusions

In this paper, we introduced and discussed a novel framework for APFDs. The approach fits well into the context of the approximate temporal functional dependencies, adding a new aspect that has never been formalized before. It differs from the previous work because we dealt with the potential predictiveness of the approximate temporal functional dependencies, considering the possibility to exploit data dependencies for the prediction. As future work, we plan to consider the algorithmic aspects taking into account all the proposed error measures.

## References

1. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 161–168 (2006)
2. Combi, C., et al.: Mining approximate temporal functional dependencies with pure temporal grouping in clinical databases. Comput. Biol. Med. **62**, 306–324 (2015)
3. Combi, C., Montanari, A., Sala, P.: A uniform framework for temporal functional dependencies with multiple granularities. In: Pfoser, D., Tao, Y., Mouratidis, K., Nascimento, M.A., Mokbel, M., Shekhar, S., Huang, Y. (eds.) SSTD 2011. LNCS, vol. 6849, pp. 404–421. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22922-0_24
4. Harel, O.D., Moskovitch, R.: Complete closed time intervals-related patterns mining. Proc. AAAI Conf. Artif. Intell. **35**(5), 4098–4105 (2021)
5. Huhtala, Y., Kärkkäinen, J., Porkka, P., Toivonen, H.: TANE: an efficient algorithm for discovering functional and approximate dependencies. Comput. J. **42**(2), 100–111 (1999)
6. Johnson, A.E., et al.: Mimic-iii, a freely accessible critical care database. Sci. Data **3**(1), 1–9 (2016)
7. Khwaja, A.: Kdigo clinical practice guidelines for acute kidney injury. Nephron Clin. Pract. **120**(4), c179–c184 (2012)
8. Kivinen, J., Mannila, H.: Approximate inference of functional dependencies from relations. Theoret. Comput. Sci. **149**(1), 129–149 (1995)
9. Mantovani, M., Amico, B., Combi, C.: Discovering predictive trend-event patterns in temporal clinical data. In: Proceedings of the 36th Annual ACM Symposium on Applied Computing, pp. 570–579 (2021)

10. Mantovani, M., Combi, C., Zeggiotti, M.: Discovering and analyzing trend-event patterns on clinical data. In: 2019 IEEE International Conference on Healthcare Informatics (ICHI), pp. 1–10. IEEE (2019)
11. 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)
12. Schrier, R.W., Wang, W., Poole, B., Mitra, A., et al.: Acute renal failure: definitions, diagnosis, pathogenesis, and therapy. J. Clin. Investig. **114**(1), 5–14 (2004)
13. Segura-Delgado, A., Gacto, M.J., Alcalá, R., Alcalá-Fdez, J.: Temporal association rule mining: an overview considering the time variable as an integral or implied component. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. **10**(4), e1367 (2020)
14. Xu, Z., et al.: Identifying sub-phenotypes of acute kidney injury using structured and unstructured electronic health record data with memory networks. J. Biomed. Inform. **102**, 103361 (2020)