On Formalism of Continuous Knowledge Discovery and Temporal Granularity

Ding Pan
Center for Business Intelligence Research, School of Management, Jinan University, Guangzhou, China
Email: tpanding@jnu.edu.cn

***Abstract*—The accessibility and abundance of data today makes knowledge discovery a matter of considerable importance and necessity. The process to discover continuously knowledge in evolving business domain is a challenge issue. A continuous knowledge discovery process is introduced for inducing the local first-order rules and global evolutional rules, to trace dynamic evolution patterns firstly. The definitions of main notions (event, sequence pattern, temporal rule) are proposed in a formal way, based on first-order linear temporal logic and temporal granularity. The measures of support and confidence about ranged degree of truth of a formula are established. The formalism defines the valuation on a linear state structure with time granules. By defining transition operation between temporal types, it is proved that only the independent information for unspanned-granule may be transferred without loss among different granularities. Otherwise, an aggregation mechanism was proposed to state sequence.**

***Index Terms*—continuous discovery process, temporal granularity, formal theory, first-order linear temporal logic**

I. Introduction

Today many organizations have more than large databases that change and grow continuously. The knowledge discovery (KDD) in such evolving data is an important challenge to the scientific and industrial communities. Before any attempt can be made to extract the useful knowledge, an overall approach that describes how to extract knowledge needs to be established. Moreover, time granulation can be regarded as an important step forward when dealing with complex problem.

Agrawal et al. showed an active data mining process, where the mining algorithm was applied to each of the partitioned data set and rules were induced [1]. Many different algorithms for incremental mining have been proposed in evolving database environment [2-4]. Gupta et al. presented a user-centric KDD process model [5]. Several researches have concentrated on fusion of domain knowledge with data mining system [6]. However, little attention has been paid to establishment of a general mechanism for supporting share of domain knowledge, intermediate results, and dynamically setting mining goal, data sources, and model selection.

The general framework researches of data mining focus more on mining algorithmic aspect, and less on the mining theoretical frameworks of knowledge discovery [7]. Cotofrei et al. investigated the form of temporal rules of time series, and presented a formalism of main terms and notions, based on the first-order temporal logic [8]. Bettini et al. presented a general framework to define time granularity systems [9].

In this paper we attempt to explore theoretical frameworks to make continuous knowledge discovery for evolving database environments. The formal theory is built upon a continuous model of the KDD process (C-KDD model), using session model. We attempt to expand the formalism theory of [8], apply the definition of primary concepts to C-KDD model in a formal way, and discuss the valuation of first-order formula and measure definition. This formalism is then extended to include the notion of temporal granularity and a discussion is made to investigate the formal relationships between the support measures of the same event with different granularities.

This paper is organized as follows: Section 2 introduces the C-KDD process model. Section 3 presents main notions of discovery process in a formal way. Section 4 introduces the general framework to define time granularity model. Section 5 discusses the definitions and theorems concerning the extension of the formalism towards a temporal granular logic. Finally, Section 6 concludes the paper.

II. C-KDD Process Model

We present a process model for the KDD process, which is based on the concept of active, continuous discovery and designed to integrate known knowledge and granules in order to support automatic discovery process [10].

To make use of data granules mechanism (or groups, classes, clusters of a universe) to support discovery process, it need create various levels of data, as well as inherent data structure, by partitioning data attributes into intervals. The different levels of granularities form data granule ontologies.

The model, called C-KDD (Cohesive KDD) model, is shown as Fig. 1 [9]. It consists of four stages: planning, session mining, merge mining, and post-processing.

Manuscript received December 25, 2010; revised January 18, 2011; accepted January 25, 2011.

doi:10.4304/jsw.6.7.1217-1224

During planning stage, the KDD process begins with business understanding, including business-aims and business-logic. Through interactive exploration and experimentation, discovery goals, business data, and subsequent processes are identified and the specification of discovery task schedule (TS) is generated.

The session mining stage performs select-transfer-premining and achieves partial data mining. It places emphasis on local and static rules induction, and executes induction on incremental data at regular intervals, e.g. month. As the functions are already specified in the TS, they are periodically repeated on incremental data as per the frequency or trigger condition, and whose outcome forms a rule bin (RB). Therefore, several session minings will generate the measure sequence of the inferred rule.

The merge mining is initiated by mining queries or a trigger event. The query contents are listed in consultation with the TS; user can commit them, according to his requests. A trigger event is occurred as the causes of time or rule rising. It places emphasis on overall and dynamic rules discovery, in the interaction paradigm, the rules are merged and refined from several RBs.

The post-processing stage begins with matching discovered rules and known knowledge, filters useless ones, then classified and ranked automatically interesting results according to interestingness. When a critical point threshold is reached, an alert will be triggered. Meanwhile, the user can review and confirm these findings. It would also integrate new interesting insights with the known knowledge, to perform knowledge evolution and presentation.

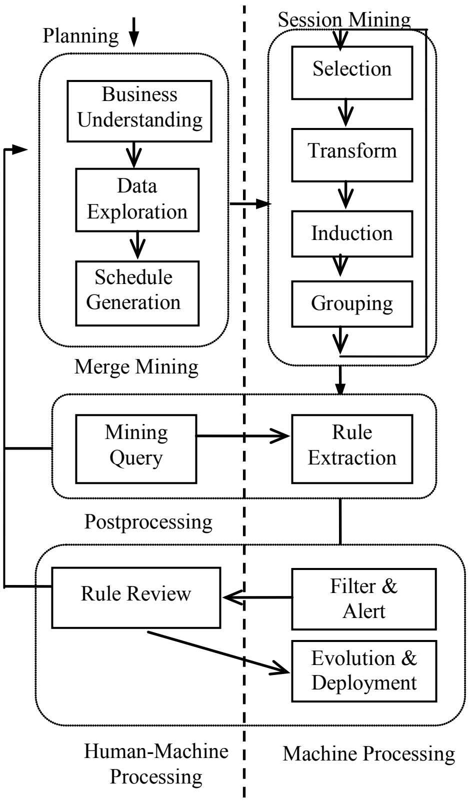

Figure 1. C-KDD process model.

Then, it forms a close-loop solution that helps to maintain the continuous knowledge discovery process. When unable to satisfy intelligence application or rule review, the process flow goes back to the planning stage to re-explore data.

III. FORMALISM OF DISCOVER PROCESS

In the continuous knowledge discovery, there is a large number of data with temporal dependencies, i.e. temporal sequence. Temporal data mining (TDM) can provide the solution to satisfy these processing requests.

We consider that a linearly ordered temporal domain is a structure $T=(T_D,<)$, where T_D is a set of discrete time instants and $<$ is a linear order relationship on T_D. For simplicity, we assume that the elements of T are strictly increasing, and $t_{i+1}-t_i=\Delta$ is a positive constant.

Given a temporal domain T, a non-empty attribute set $\{A_1,...,A_k\}$ denotes to each attribute of finite entity set Q (such as customer or stock), then a temporal sequence is a ordered item list $X=\{X_1, X_2,... X_i,...\}$, where X_i is a k+2-tuple $(q,t,a_1,...,a_k)$, $q\in Q$, $t\in T$, $a_i\in D_{A_i}$. A sequence space W_X consists of all instances of the sequence category. Mined data set w_X is a subset of W_X.

After the pre-processing of temporal sequence, a sequence X in w_X has been transformed into a linear ordered sequence of events X_e, consisted of some basic shapes or strings. Given a finite symbol set D_e of the basic shapes or strings, the feature function set $\{f_1,...,f_p\}$ ($p\geq 0$) and each corresponding domain D_{f_i}, then the event set of w_X is $E=\{(q,e,b_1,...,b_p)|\ q\in Q,\ e\in D_e,\ b_i= f_i,\ b_i\in D_{f_i}\}$, so $X_e=\{E_1, E_2,... E_j,...\}$, E_j is an event at t_j.

As inferring classification rule, there is a finite class symbol set D_g. The mapping $w_t\rightarrow D_g$, where $w_t\subset W_X$, must be specified prior to the induction process, where w_t is called as train set, eventually it infers $w_X\rightarrow D_g$ for supervised learning, while the induction process generates directly $w_X\rightarrow D_g$ for unsupervised learning.

Example 1. Given a database containing daily price variations of the stocks, specified basic symbol set D_e= {peak, valley, flat}. Each event has form (q, e, b_1, b_2), where q is entity name, e is one of the strings {peak, valley, flat}, and b_1, b_2 represent the means respectively, the standard error. We assume that the price sequence of IBM stock is transformed into the event sequence ((IBM, flat, 3, 1.5), (IBM, peak, 10, 2.4), ..., (IBM, peak, 8, 1.4)). For classification task, there is class symbol set D_g={grow, stabile, wave, risk}.

For presenting definition and formalism of primary concepts used in discovery process, we begin by defining the various tasks that occupy the attention of researches.

- Search: Given a query sequence S, and some similarity/dissimilarity measure $D(S,C)$, find the nearest matching sequences C in data set w_X.
- Temporal association rule mining: Find the complete set of association rule and temporal features in data set w_X under measures such as a support and a confidence.

• Sequence mining: Find the complete set of frequently occurring ordered events or subsequences as patterns in data set w_X under a support measure.

• Clustering: Find natural groupings of the sequences in data set w_X under some similarity/dissimilarity measure $D(S,C)$.

• Classification: Given an unlabeled sequence S, assign it to one of two or more predefined classes.

A. Syntax

A first-order logic language contains usually constant, function, predicate, and general symbols like connectives, etc. For the requirement of formalism we consider a restricted first-order temporal logic language L, which contains constant symbols, variable symbol, function symbols, predicate symbols, relational symbol set $\{=,<,\leq,>,\geq\}$, quantifiers $\{\forall,\exists\}$, temporal operators $\{\Box,\Diamond\}$, logical connectives $\{\neg,\wedge\}$ and a temporal connective $\triangle_k$, $k\in\mathbb{Z}$, where $k>0$ denotes next k time instants, $k<0$ denotes last k time instants, $k=0$ denotes now.

The syntax of L defines the set of terms, atomic formulae (or atom) and formulae denoted respectively by *Term*(L), *Atom*(L) and *Form*(L).

Definition 1. (*Term*(L)) $t\in Term$(L), if and only if t is defined inductively (finite times) by the following rules:

(1) a, $y\in Term$(L), where a is a constant symbol, y is a variable symbol.

(2) If $t_1,t_2,\ldots,t_n\in Term$(L) and f is an n-ary function symbol then $f(t_1,t_2,\ldots,t_n)\in Term$(L).

Definition 2. (*Atom*(L)) An expression $A\in Atom$(L), if and only if A is defined by the following arbitrary rules:

(1) $P(t_1,t_2,\ldots,t_m)$, where P is an m-ary predicate symbol, $t_1,t_2,\ldots,t_m\in Atom$(L).

(2) Relational atom $t_1\ \rho\ t_2$, where t_1, $t_2\in Atom$(L), $\rho\in \{=, <, \leq, >, \geq\}$.

Definition 3. (*Form*(L)) An expression $F\in Form$(L), if and only if F is defined inductively (finite times) by the following rules, where F, F_1, $F_2\in Form$ (L).

(1) $Atom(L)\in Form$(L).

(2) $(F_1\wedge F_2)\in Form$(L).

(3) $\neg F\in Form$(L), $\triangle_k F\in Form$(L).

(4) $\Box F\in Form$(L), $\Diamond F\in Form$(L).

(5) If $F(y)\in Form$(L), where y is a variable symbol, then $\forall yF(y)\in Form$(L) and $\exists yF(y)\in Form$(L).

Based on the linear temporal logic, the formulas are true or false on computation paths, that is, sequences of states s_0, s_1, s_2, …. The formula $\Box F$ means that F is true at all states along the path. The formula $\Diamond F$ means that F is true at some state on the path. A quasi-Horn clause is a formula of form: $A_1\wedge\ldots\wedge A_k\Rightarrow A_{k+1}$, if and only if it is syntactically equivalent with the formula $A_1\wedge\ldots\wedge A_k\wedge A_{k+1}$, where A_i is a positive atom.

Definition 4. (*Event*) Given finite symbol set Q and D_e, an event is an atom formed by a $p+2$-ary predicate $E(q,e,f_1,\ldots,f_p)(p\geq 0)$, where $q\in Q$ is an entity symbol, e is a symbol representing the name of the event, q and e are constant or variable symbols, and $f_1,\ldots,f_p\in Term$(L) are the function symbols.

Definition 5. A constraint formula for the event $E(q,e,f_1,\ldots,f_p)$ is a conjunctive formula, $E(q,e)\wedge C_1\wedge\ldots\wedge C_m$. $E(q,e)$ is the event, each C_i $(1\leq i\leq m)$ is a relational atom $t\rho c$, where t is one of variable symbols $\{q,e,f_1,\ldots,f_p\}$ representing the name of entity, event or corresponding f_i respectively, c is a constant symbol, and $\rho\in\{=,<,\leq,>,\geq\}$. At least, it contains a relational atom $E(q,e)$, where q and e are constant symbols, denoted as a short constraint formula. A temporal constraint formula B_k denotes $\triangle_k(E(q,e)\wedge C_1\wedge C_2\wedge\ldots\wedge C_m)$, $k\in\mathbb{Z}$.

Definition 6. A sequence pattern (also called pattern) is a conjunctive formula of several ordered temporal constraint formula B_k, $B_{i_1}\wedge B_{i_2}\wedge\ldots\wedge B_{i_m}$, $i_1\leq i_2\leq\ldots\leq i_m\leq 0$.

Definition 7. A subsequence is a sequence pattern, $B_{i_1}\wedge B_{i_2}\wedge\ldots\wedge B_{i_m}$, $i_1\leq i_2\leq\ldots\leq i_m\leq 0$, where each B_{i_k} is a temporal short constraint formula.

Definition 8. A temporal rule is a formula of the form $B_{i_1}\wedge B_{i_2}\wedge\ldots\wedge B_{i_m}\Rightarrow H_{i_{m+1}}$, $i_1\leq\ldots\leq i_m\leq i_{m+1}\leq 0$, where B_{i_k} and $H_{i_{m+1}}$ are constraint formulas, prefixed by the temporal connectives Δ_{i_k} and $\Delta_{i_{m+1}}$. We denote such $B_{i_1}\wedge\ldots\wedge B_{i_m}$ as rule body, $H_{i_{m+1}}$ as rule head.

The definitions 4-8 have dealt with the local properties of sequences, and then clustering and classification rules involving the global properties are as follow:

Definition 9. A class is an atom formed by a $p+1$-ary predicate $G(g, f_1,\ldots,f_k)(k\geq 0)$, where $g\in D_g$ is an element of the class symbol, and $f_1,\ldots,f_k\in Term$(L) are the function symbols about class properties.

Definition 10. A classification rule is a formula of the form $B_{i_1}\wedge B_{i_2}\wedge\ldots\wedge B_{i_m}\Rightarrow G(g, f_1,\ldots f_k)$, $i_1\leq i_2\leq\ldots\leq i_m\leq 0$.

The results of clustering and classification task are represented in classification rules. The temporal rule, classification rule, subsequence and pattern are called first-order rule $R_F\in Form$(L), by a joint name. $|i_m-i_1|+1$ are the time interval of R_F. The temporal rule and classification rule are quasi-Horn clause prefixed by $\triangle_i$. When only order of the event is considered, the temporal connective $\triangle_i$ may be omitted in R_F.

B. Semantics

The semantics of terms and formulae of L is provided by an interpretation. Generally, for structure $U=(D,\{a^i\},\{f^i\},\{R^i\})$, where D is domain, a^i, f^i and R^i represent constant, total function and predicate respectively on D, the constant, function and predicate symbols of L must be mapped to U, respectively. Moreover, the individuals in D are assigned to interpreted freedom variables. The other symbols have common semantics or have been explained above. The interpretation and assignation are called valuation jointly. The valuation V: (i)$V(a),V(u)\in D$. (ii)$V(f):D^n\rightarrow D$. (iii)$V(R):D^m\rightarrow\{$ture,flase$\}$. The valuation V may be extended to arbitrary expression. For a formula p, the meaning of truth under valuation V is denoted for $V|\!=p$.

Given $D=w_X \cup D_e \cup D_f \cup D_g$, the feature function set $\{f_1,...,f_p\}$ is defined on D. To determine meaning based on a first-order linear temporal logic, we create a structure having a temporal dimension and capable of valuating the relationship between a specific moment and the valuation V, according to *Kripke* structure.

Definition 11. Given L and a domain D, a linear state structure is a triple $K=(S, \sigma, V)$, where S is a finite non-empty set of states, $\sigma=\{\sigma_i|\sigma_i=(s_i,s_{i+1},...,s_j,...),\ s_j \in S,\ j \in \mathbb{N}\}$ is a non-empty set of infinite sequence of states and V: $Form(\mathrm{L})\times\sigma\rightarrow\{\text{true, false}\}$ is a function that associates with the sequence σ_i a valuation V_{σ_i} of all formulae of L.

The sequence set w_X forms the set of the event sequence $\Omega=\{S^1,...,S^k,...,S^n\}$, after the pre-processing. To define a linear state structure $K=(S, \sigma, V)$, we specify a state s_i as an event, the S as the set all event and σ_i as the event sequence. For simplicity, we assume the state s_i means the sequence σ_i beginning with s_i, viz. $V_{\sigma_i} = V_{s_i}$, hereinafter.

Given a linear state structure K and Ω, we denote the $V|=p$ of a sequence S^k at a state s_i by $(K,S^k,i)|=p$. For a temporal rule, we denote the rule body in S^j and the rule head in S^k by $(K,S^{j,k},i_1)|= B_{i_1} \wedge B_{i_2} \wedge ... \wedge B_{i_m} \Rightarrow H_{i_{m+1}}$ or simply $i_1|=p$, if there is no confusion for K and $S^{j,k}$. We assume the rule body and head in the same sequence, hereinafter.

Using this definition, we can also define: $i|=p \wedge q$ if and only if $i|=p$ and $i|=q$; $i|=\triangle_k p$ if and only if $i+k|=p$. So, $i|=E(q,e,f_1,...,f_p)$ denotes that for entity q an event with the name e and the features $V(f_1),..., V(f_p)$ occurs at state s_i. Analogously, a temporal constraint formula B_k is true at state s_i if and only if $i|=E(q,e)$ and all $i|=C_j$; a subsequence is true at state s_i if and only if all $i= B_{i_k}$; a pattern is true at state s_i if and only if all $i|= B_{i_k}$; a temporal rule is true at state s_i if and only if $i|= B_{i_1} \wedge B_{i_2} \wedge ... \wedge B_{i_m} \wedge H_{i_{m+1}}$. Moreover, if a classification rule is true, then there is $i|= B_{i_1} \wedge B_{i_2} \wedge ... \wedge B_{i_m}$ and the predicate $G(g, f_1,...,f_k)$. It means that the sequence having the features $B_{i_1} \wedge B_{i_2} \wedge ... \wedge B_{i_m}$ belonged to class g.

We can establish some measures about ranged degree of truth of a formula $V|=p$ on the Ω. Now assuming that for each formula p in L, there is an algorithm that calculates the value of $V(p)$ for every state on mined dataset, in a finite number of steps.

Definition 12. Given L and a linear state structure K, for every formula p, on the set of state sequence Ω, a real set function $P(p)=|A|/n$, where $n=|\Omega|$ and $A=\{k\in \{1,...,n\} |(K,S^k,i)|=p\}$.

Theorem 1. Given L and a linear state structure K, for a formula p, the real set function $P(p)$ is a probability of $V|=p$ on the Ω.

Proof. By applying K, suppose $\Omega=\{S^1,...,S^k,...,S^n\}$ as sample space. Let $F=2^{\Omega}$, then $\Omega\in F$ and $Q\in F$, where Q is an arbitrary subset of Ω. Therefore, F is an σ-algebra.

For a formula p, let $Q=\{S^k\ |(K,S^k,i)|=p\}$, $A=\{k|S^k\in Q\}$, then $|A|\geq 0$, $P(p)\geq 0$. When $Q=\Omega$, $|A|=n$, then $P(p)=1$. Again let $|A_j|=k_j$ $(\leq n)$, then corresponding $P(p_j)= k_j/n$ $(j=1,2,...,m)$. If the A_js are not intersectant, then for the corresponding p_j:

$$P(\sum_{j=1}^{m} p_j) = (\sum_{j=1}^{m} k_j)/n = \sum_{j=1}^{m} \frac{k_j}{n} = \sum_{j=1}^{m} P(p_j)$$

Therefore, (Ω, F, P) is a probability space.

Definition 13. Given language L and a linear state structure K, a measure for the $V|=p$ of a formula p is a function $Supp(p)=P(p)$. The measure is usually called the support of the p.

For a temporal rule, there is another useful measure about the degree of truth of the implication between the rule body and head.

Definition 14. Given language L and a linear state structure K, for the $V|=p$ of a temporal rule p, a measure of p is a function $Conf(p)=P(p)/P(p_b)$, where p_b is the rule body. The $Conf(p)=0$ if $P(p_b)=0$. The measure is usually called the confidence of the temporal rule p.

Those measures of pattern and temporal rule describe the local characteristic. In the other hand, the support measure of classification rule describes the degree of truth about a class. When a sequence belongs to several classes, the priority of the rules determines the class of the sequence. Analogously, we can define the accuracy measure of classification rule.

Now supposing classification rule set, $C=\{C_g|g\in D_g\}$, classifies uniquely the sequences in the Ω, where C_g is the set of classification rule about g.

Given language L, a linear state structure K and $\Omega=\{S^1,...,S^k,...,S^n\}$, the accuracy measure of C_g is $accu(C_g)=|\{k\in A|(M,S^k,i)|=p,\ p\in C_g\}|/|A|$, where $A=\{k\in \{1,...,n\}|(S^k,g)\}$, (S^k,g) represents that the S^k belongs actually to class g.

For supervised classification, the set of the event sequence Ω should be independent of the train set. While on the contrary, for unsupervised classification, the judgment of the (S^k,g) usually depends on the practical application.

Theorem 2. Given language L and a linear state structure K, the accuracy of classification rule set C is as follow:

$$accu(C)= \sum_{g_j\in D_g} accu(C_{g_j})\times P(g_j)$$

where $P(g_j)$ is the probability of (S^k,g) in the Ω.

Proof. According to the definition, $accu(C_{g_j})=|TA_j|/|A_j|$, where $TA_j=\{k\in A_j|(K,S^k,i)|=p,\ p\in C_{g_j}\}$, $A_j=\{k\in\{1,...,n\}|(S^k,g_j)\}$.

$$accu(C) = \frac{1}{n}\sum_{g_j\in D_g}|TA_j| = \sum_{g_j\in D_g}(\frac{|TA_j|}{|A_j|}\times\frac{|A_j|}{n})$$

$$= \sum_{g_j\in D_g}(accu(C_{g_j})\times P(g_j)).$$

C. Session Model

In practical application, the user has no access to the entire sequence, or the sequences mined have only finite time intervals. Therefore, the measures should be calculated in a finite linear state structure, i.e. a session.

Definition 15. Given L and a linear state structure K, a session for K is a structure $\widetilde{K}=(\widetilde{\sigma},sl)$, where sl is length of the session, the sequence $\widetilde{\sigma}=(s_{i_1},s_{i_2},...,s_{i_{sl}})$ and the s_{i_j} $(1\le j\le sl)$ is a accessible state.

For a formula p and a sequence $S^k\in\Omega$, we have the meaning of truth under valuation $V_{\widetilde{K}_i}$ on $\widetilde{K}_i=(\widetilde{\sigma}_i,sl)$. Moreover, for a formula p and Ω, we can estimate some measures about ranged degree of truth on $\widetilde{K}$ as follow, where $\widetilde{\Omega}=\{\widetilde{S}^k=(s^k_{i_1},s^k_{i_2},...,s^k_{i_{sl}})\mid \widetilde{S}^k\subseteq S^k\}$, $1\le k\le n$.

Definition 16. Given L and a session $\widetilde{K}$ for K, an estimator of $Supp(p)$ of a formula p is $ES(p, \widetilde{K})=|A^e|/n$, where $|\widetilde{\Omega}|=n$, $A^e=\{k\in\{1,...,n\}|(\widetilde{K},S^k,i)|=p\}$.

Definition 17. Given L and a session $\widetilde{K}$ for K, an estimation of the $Conf(p)$ of temporal rule p is $EC(p, \widetilde{K})=ES(p, \widetilde{K})/ES(p_b, \widetilde{K})$, where p_b is a rule body, if $ES(p_b, \widetilde{K})=0$, then $EC(p, \widetilde{K})=0$.

So, a session mining SM_F for $\widetilde{K}$ is a sextuple ($Task$, T_{st}, w_X, $Stat$, DK, R), where $Task$ is a task of mining, T_{st} is a start time of mining, w_X is a mined data set, $Stat$ is a threshold of the measure, DK is domain knowledge, $R=\{r\in R_F|Stat(r)\wedge DK(r)\}$ is induced rule set. According to the definition, the estimator sequence of the measures, called measure sequence, is generated across various sessions, for the formula p. For example, the support sequence, $ES_1,ES_2,...,ES_r,...$, and the confidence sequence, $EC_1,..., EC_r,...$.

Definition 18. Given L and a linear state structure K, a sequence S is consistent for a formula p, if the limit $\lim_{m\to\infty}\frac{|B|}{m}$ exists, where $B=\{i\in\{1,...,m\}|(K,S,i)|=p\}$. The set of the event sequence Ω is a p consistent set if every sequence in Ω is consistent for the p.

Theorem 3. Given L and a session $\widetilde{K}$, if a formula p exists the consistent set Ω, then for the p, when sl of the $\widetilde{K}$ is long enough, $\lim_{r\to\infty}ES_r=P(p)=|A|/n$, where $|\Omega|=n$, $A=\{k\in\{1,...,n\}|(K,S^k,i)|=p\}$.

Proof. Let $\Omega=\{S^1,...,S^n\}$ is a p consistent set, then for a $S^k\in\Omega$, the limit $\lim_{m\to\infty}\frac{|B|}{m}=\alpha_k$ exists, where $B=\{i\in\{1,...,m\}|(K,i)|=p\}$. Therefore, there is $\alpha(p)=\{\alpha_1,...,\alpha_n\}$ in the Ω. Let $\alpha=min(\{\alpha_j\in\alpha(p)|\alpha_j>0\})$, $sl=max(1/\alpha$, the time interval of p).

Let $\widetilde{K}=(\widetilde{\sigma},sl)$, the estimator sequence of support is $ES_1,ES_2,...,ES_r,...$. Obviously, $A^e_r\subseteq A$. If $j\in A=\{k\in\{1,...,n\}|(K,S^k,i)|=p\}$ for arbitrary S^j, then $\alpha_j>0$. When r is large enough, $|\{i\in\{r_1,...,r_{sl}\}|(\widetilde{K},\widetilde{S}^j,i)|=p\}|>0$ for the subsequence $\widetilde{S}^j=(s^j_{i_1},s^j_{i_2},...,s^j_{i_{sl}})$, so $j\in A^e_r$, viz. $A\subseteq A^e_r$. Therefore,

$$\lim_{r\to\infty}ES_r=\lim_{r\to\infty}\frac{|A^e_r|}{n}=\frac{1}{n}\lim_{r\to\infty}|A^e_r|=\frac{|A|}{n}=P(p).$$

Generally, there are three categories of basic trend for the measure sequence of: ascend, descend and fluctuation. High order rule is used to describe the dynamic characteristic of the first-order rule [4]. Its syntax is the same as the definition of the first-order rule above.

IV. The Time Granularity Model

The concept of a temporal type to formalize the notion of time granularities, as described in [9]. It is a generalization of most definitions of linear time granularities.

Definition 19. Let (I, <) (index) be a discrete linearly ordered temporal set isomorphic to a subset of the integers with the usual order relation, and let (T, <) (absolute time) be a linearly ordered set. Then, a temporal type on (I, T) is a mapping μ: $I\rightarrow 2^T$, for $i<j$, such that

1) $\mu(i)\neq\emptyset$ and $\mu(j)\neq\emptyset$, imply that each element in $\mu(i)$ is less than all the elements in $\mu(j)$,

2) for all $\mu(i)\neq\emptyset$ and $\mu(j)\neq\emptyset$, then all k, $i<k<j$ implies $\mu(k)\neq\emptyset$.

The $\mu(i)$ is called the ith granule of μ. Property 1) reveals that granules do not overlap and that the mapping must be monotonic. Property 2) disallows an empty set to be the value of a mapping for a certain index value if a lower index and a higher index are mapped to non-empty sets.

Definition 20. Let μ and ν be temporal types on (I, T), then there are some relationships as follow:

Finer-than: μ is said to be finer than ν, denoted $\mu\preceq\nu$, if for each $i\in I$, there exists $j\in I$ such that $\mu(i)\subseteq\nu(j)$.

Groups-into: μ is said to group into ν, denoted $\mu\trianglelefteq\nu$, if for each non-empty granule $\nu(j)$, there is a subset J of I such that $\nu(j)=\bigcup_{i\in J}\mu(i)$.

Shifting: μ and ν are said to be shifting equivalent, denoted $\mu\simeq\nu$, if for each $i\in I$, there exists $j\in I$ such that $\mu(i)\subseteq\nu(j)$ and $\nu(i)\subseteq\mu(j)$.

For shifting equivalent, we have a bijection function g: $I\rightarrow I$ such that $\mu(i)=\nu(g(i))$, for all $i\in I$. We disallow multiple types that are equivalent with respect to shifting of their indices, hereinafter.

When a temporal type μ is finer than a temporal type ν, we also say that ν is coarser than μ. The finer-than relationship formalizes the notion of finer partitions of the absolute time. By definition, this relation is obviously reflexive, transitive and antisymmetric, if it is not shifting equivalent, hence, it is a partial order. Therefore, there exists a unique least upper bound, denoted by $\mu\top$, and a unique greatest lower bound, denoted by $\mu\bot$.

Moreover, a temporal type system having an infinite index is a lattice with respect to the finer-than relationship. Concerning the groups-into relationship, it also satisfies the properties of a partial order.

Consider now the positive natural numbers is used as the index set and the absolute time set, namely the temporal type on ($\mathbb{N}$, $\mathbb{N}$). We impose to any temporal type μ the restrictions on Definition 19:

a) $\mu(i)=\emptyset$ implies $\mu(j)=\emptyset$

b) For each $j\in\mathbb{N}$, exists $j\in\mu(i)$

The set of granules which satisfy these conditions are denoted by $\mathbb{G}$. The condition a) enforces that the first non-empty granule must start with index 1. The condition b) reveals the granules cover all the absolute time. For $\mathbb{G}$, the relationships finer-than and groups-into is the equivalence and $\mu\perp(i)=\{i\}$ [9].

V. Formalism with Linear Granular Structure

If $K=(S, \sigma, V)$ is a first-order linear time structure, then the σ_1 is a complete state sequence, denoted $\overline{\sigma}$. The $\overline{\sigma}$ forms the absolute time A, by identifying the time moment i with the state s_i (on the ith position in the sequence). If μ is a temporal type from $\mathbb{G}$, then the temporal granule μ(i) consists of the set $\{s_j \in S, j \in \mu(i)\}$. Therefore, the temporal type μ induces a new state sequence $\overline{\sigma}^{\mu}$, defined as $\overline{\sigma}^{\mu}(i)=\mu(i)$. We assume the set μ($i$) will be considered either as a set of natural numbers, or as a set of states, hereinafter.

Consider now the linear time structure derived from K, $K_\mu=(S, \sigma^\mu, V^\mu)$, where $\sigma^\mu=\{\sigma_i^\mu \mid \sigma_i^\mu=(\mu(i),\mu(i+1),...), \mu(i) \subseteq S, i\in\mathbb{N}\}$. To be well defined, we must give the valuation $V^{\mu}_{\mu(i)}$ for each $i\in\mathbb{N}$. Because the set μ(i) is a finite sequence of states, it defines a session $\widetilde{K}_{\mu(i)}$ for K_μ. Therefore the meaning of truth of a formula p for a sequence S is defined as follow; here formula p is a temporal free formula, the 1 means true, the 0 means false.

$$V^{\mu}_{\mu(i)}(p)=V_{\widetilde{K}_{\mu(i)}}(p) = V_{\widetilde{K}_i}(p)=\begin{cases}1, (\widetilde{K}_i, S, i) \models p \\ 0, else\end{cases} \quad (1)$$

where $\widetilde{K}_i=(\widetilde{\sigma}_i, sl)$.

According to the semantics of formula on the state sequence $\widetilde{\sigma}_i$, this valuation is extended to multi-formula, if each formula p_i is independent.

$$V^{\mu}_{\mu(i)}(\Delta_{k_1}p_1 \wedge ... \wedge \Delta_{k_n}p_n)=\prod_{j=1}^{n} V^{\mu}_{\mu(i+k_j)}(p_j) \quad (2)$$

where p_i are temporal free formulae and $k_i\in Z$, i=1 ... n.

Definition 21. If $K=(S, \sigma, V)$ is a first-order linear time structure and it is a temporal type from $\mathbb{G}$, then the linear temporal structure induced by it on K is the triple $K_\mu=(S, \sigma^\mu, V^\mu)$, where $\sigma^\mu=\{\sigma_i^\mu \mid \sigma_i^\mu=(\mu(i),\mu(i+1),...), \mu(i) \subseteq S, i\in\mathbb{N}\}$ and V^μ: $Form(L)\times\sigma^\mu\rightarrow$\{true, false\} is a function that associates with the sequence σ_i^μ a valuation $V^{\mu}_{\sigma_i}$ of all formulae of L, according to the rules (1) and (2).

For simplicity, we assume the state set μ(i) means the sequence σ_i^μ beginning with μ(i), viz. $V^{\mu}_{\sigma_i}=V^{\mu}_{\mu(i)}$.

For $\mathbb{G}$ and any accessible states, as $\mu\perp(i)=\{i\}$, we have $\widetilde{K}_i=(\{i\},1)$, $V^{\mu\perp}_{\mu(i)}(p)=V_{\widetilde{K}_{\mu(i)}}(p)=V_{\widetilde{K}_i}(p)=V_{s_i}(p)$, namely the $V^{\mu\perp}(p)$ is equivalent to the $V(p)$. Hence, in this case, $K_{\mu\perp}=(S, \sigma^{\mu\perp}, V^{\mu\perp})$ is same as the initial linear structure K at the valuation level.

Furthermore, we may establish the relation linking the measure supports, from two linear temporal structures induced by μ and ν, when exists a relationship finer-than between these two temporal types. When μ, ν are from $\mathbb{G}$, for $\forall j\in\mathbb{N}$, there is a subset $N_i \subset \mathbb{N}$ such that $\nu(j)=\bigcup_{i\in N_i}\mu(i)$. If p is a temporal free formula in L, then the measure $supp(p)$ for p at ν(j) is the maximum value within the set of $supp(p)$ for p at μ(i), where $i\in N_i$. We formalize this result in the following theorem:

Theorem 4. If μ, ν are temporal types from $\mathbb{G}$, such that $\mu\preceq\nu$, and $Supp^\mu(p)$, $Supp^\nu(p)$ are the supports of the p from the induced linear temporal structures K_μ and K_ν on K, then for each $j\in\mathbb{N}$,

$$Supp^{\nu}_{\nu(j)}(p)=\underset{i\in N_j}{Max}(Supp^{\mu}_{\mu(i)}(p)) \quad (3)$$

where N_j is the subset of $\mathbb{N}$ which satisfies $\nu(j)=\bigcup_{i\in N_j}\mu(i)$ and p is a temporal free formula in L.

The theorem 4 is only applied to temporal free formulae. For the unspanned-granule case, we can prove that the measure support, in the coarser world, of a temporal formula with a given temporal types ν is linked with the support, in the finer world, of a similar formula but having a temporal types μ.

Lemma 1. If μ, ν are temporal types from $\mathbb{G}$ such that $\mu\preceq\nu$ and $Supp^\mu(p)$, $Supp^\nu(p)$ are the supports of the p from the induced linear temporal structures K_μ and K_ν on K, then for each $i, j\in\mathbb{N}$,

$$Supp^{\nu}_{\nu(j)}(p\wedge\Delta_1 q)= \underset{i\in N_j}{Max}(Supp^{\mu}_{\mu(i)}(p))\times\underset{i\in N_{j+1}}{Max}(Supp^{\mu}_{\mu(i)}(p)) \quad (4)$$

where $\nu(j)=\bigcup_{i\in N_j}\mu(i)$, $\nu(j+1)=\bigcup_{i\in N_{j+1}}\mu(i)$ and p,q are temporal free formulae in L.

Given the set of temporal types $\mathbb{G}_1=\{\mu\in\mathbb{G} \mid i\in\mathbb{N}, |\mu(i)|=c_\mu\}$, then the following lemma and theorem hold.

We define the operator Z_d over the formulae in L as

$$Z_d(\Delta_{k_1}p_1\wedge...\wedge\Delta_{k_m}p_m)=\Delta_{d.k_1}p_1\wedge...\wedge\Delta_{d.k_m}p_m \quad (5)$$

where $|\mu(i)|=c_\mu$, $|\nu(j)|=c_\nu$, $d=c_\nu/c_\mu$ and $k_i\in\mathbb{Z}$, i=1 ... n.

The operator Z_d changes temporal type from the coarser world to the finer world. For Δ_{k_i} time instant, we have $N_j=\{d\times k_i+1,...,d\times k_i+d\}$, after Z_d transformation.

Theorem 5. If μ, ν are temporal types from $\mathbb{G}_1$ such that $\mu\preceq\nu$ and $Supp^\mu(p)$, $Supp^\nu(p)$ are the supports of the p from the induced linear temporal structures K_μ and K_ν on K, then for each $i, j\in\mathbb{N}$,

$$\begin{aligned}&Supp^{\nu}_{\nu(j)}(\Delta_{k_1}p_1\wedge...\wedge\Delta_{k_m}p_m)\\ &=Supp^{\mu}_{\mu(i)}(\Delta_{d.k_1}p_1\wedge...\wedge\Delta_{d.k_m}p_m)\\ &=\prod_{l=1}^{m}\underset{i\in N_j}{Max}(Supp^{\mu}_{\mu(i)}(p_l))\end{aligned} \quad (6)$$

where $|\mu(i)|=c_\mu$, $|\nu(j)|=c_\nu$, $d=c_\nu/c_\mu$, $\nu(j)=\bigcup_{i\in N_j}\mu(i)$, $N_j=\{d\times k_i+1,...,d\times k_i+d\}$ and p,q are temporal free formulae in L.

Consequently, if we have three world, W_1, W_2 and W_3 corresponding to three temporal type, and defined in turn $\preceq$ relationship, then the support information in each granule is transferred from W_1 to W_2. When W_1 is "lost", it is possible to transfer the support information from W_2 to W_3 and to obtain the same result as the transfer from W_1 to W_3, according to the theorem 4 and 5. From above definitions and theorems, if we know the degree of support of a temporal rule in the W_1, then we can expect larger degree of support of the same rule in the world W_2, coarser than W_1.

For a spanned-granule case, a formula may span several temporal granules. The transition between temporal types may produce that some kinds of rules disappear or new kinds of rules appear. We need transfer the state sequence σ_i^{μ} into σ_i^{ν}, and $|N_j|$=1 for (5), namely state aggregation firstly, using some fitting methods, as we want to know the meaning of truth of the formula in coarser world.

As a state sequence**,** $\overline{\sigma}$ has three types of basic state event, viz. fluctuant, ascend and descend. We can solve the parameter estimation problem of the sequence, based on the principle of information diffusion, in [11].

The principle of information diffusion: Let S={s_1, s_2,...,s_m} be a sample, U be the universe of discourse, and u_j be an observation of s_j. Let x=φ(u-u_j), if S is incomplete, there is a reasonable information diffusion function θ(x) which can lead to the information obtained from u_j, value as 1, diffuse to u according to θ(x), and the diffusion estimate is nearer to the real distribution than non-diffusion estimate, as diffused primary distribution $Q(x)=\sum_{j=1}^{n}\theta(\varphi(u-u_j))$.

For fluctuant, let C be a time instant subset, used to take information diffusion of $\widetilde{\sigma}_i$ on F. So,

$S = \{s_1, s_2,...,s_m\}$

$\widetilde{\sigma}_i = \{s_{i_1}, s_{i_2},..., s_{i_m}\}$

$U = [0,1]$,

$C=\{c_1,c_2,\ldots,c_k\}$,

The c_i is called a control point. It receives total information diffused from each state s_{i_j}, deduces the expectation estimation of state aggregation, using a Borel measurable function θ, as follow:

$$Q(c_i)=\hat{f}_n(c_i)=\frac{1}{nd}\sum_{j=1}^{n}\theta[\frac{c_i-s_{i_j}}{d}]$$

$$=\frac{1}{\sqrt{2\pi}nh}\sum_{j=1}^{n}\exp(-\frac{(c_i-s_{i_j})^2}{2h^2}) \qquad (7)$$

where h=σd, σ is standard error, d is the diffusion window width. Generally, h=α(S_{max}-S_{min})/(n-1). If n≥10, α=1.4208.

$$E=\sum_{i=1}^{m}c_iP_i=\sum_{i=1}^{m}c_i(Q(c_i)/\sum_{i=1}^{m}Q(c_i)) \qquad (8)$$

For ascend and descend, the expectation $E(s_{i_j})$ may be regarded as the local excursion along with state sequence. Hence, we find the center points of states using clustering method firstly, and then fit these center points using linear equation.

VI. Conclusion

Implementation of a continuous KDD process has been attracting more and more interests from various industries. We have proposed a formalism of continuous knowledge discovery process and temporal granularity operation, based on first-order linear temporal logic. The process model endeavors to separate autonomous discovery process, and forms session mining and merge mining, based on a session model, and then it is extended to include the concept of temporal granularity. We defined the main notions of event, sequence pattern, temporal rule, etc. in a formal way. We established also the measures of support and confidence about ranged degree of truth of a formula on whole data set, considering some data are inaccessible or missing. By defining transition operation between temporal types satisfied finer-than relationship, we proved that only the independent information for unspanned-granule may be transferred without loss among different temporal granularities. Otherwise, a state aggregation method was proposed based on the principle of information diffusion.

Acknowledgment

This work is supported by the National Natural Science Foundation of China under grant No. 70771044 and 70872020, and the Joint Project in Industry-university-research of Guangdong Province and Ministry of Education of China under grant No. 2010B090400169.

References

[1] R. Agrawal and G.Psaila, “Active data mining,” in Proc. of the KDD’95. California, 1995, pp.3-8.

[2] D. Cheung, J. Han and V. Ng, “Maintenance of discovered association rules in large databases,” in Proc. of the ICDE’96. Los Alamitos, 1996, pp. 106-114.

[3] H. Cheng, X. Yan and J. Han, “IncSpan: incremental mining of sequential patterns in large database,” in Proc. of ACM SIGKDD’04, New York, 2004, pp. 527-532.

[4] M. Spiliopoulou and J. Roddick, “Higher order mining,” in Proc. of 2nd Data Mining Methods and Databases, Southampton, 2000, pp. 309-320.

[5] S. K. Gupta, V. Bhatnagar and S. K. Wasan, “Intension mining,” TR No. IITD/CSE/TR2000/001, Delhi: Indian Institute of Technology, 2000.

[6] S. C. Yoon, L. J. Henschen and E. K. Park, “Using domain knowledge in knowledge discovery,” in Proc. of the CIKM’99. New York, 1999, pp. 243–250.

[7] G. R. Livingston, J. M. Rosenberg and B. G. Buchanan, “An agenda- and justification-based framework for discovery systems,” *Knowledge and Information Systems*, vol.5, no.2, pp.133-161, 2003.

[8] P. Cotofrei and K. Stoffel, “Higher Order Temporal Rules,” in Proc. of the ICCS 2003, LNCS 2657. Melbourne, 2003, pp.323-332.

[9] C. Bettini, X. S.Wang, and S. Jajodia, “A general framework for time granularity and its application to temporal reasoning.” *Ann. Math. Artif. Intell.* vol. 22, no. 1-2, pp. 29-58, 1998.

[10] D. Pan, “A Conceptual Infrastructure for Knowledge Discovery Process with Granularity,” in Proc. of the WISM'10-AICI'10, Sanya, Oct., 2010, pp. 99-103.

[11] D. Pan and Y. Pan, "The estimation of rule measure based on principle of information diffusion," in Proc. of the 5th ICMLC 2006, Dalian. Aug. 2006, pp. 1025-1029.

Ding Pan was born in Fuzhou, China. He received his Ph.D. in computer science and technology from the Xi'an Jiaotong University, China; received his M.Sc. in accounting from the Xiamen University and B.Sc. in computer software from the Huazhong University of Science and Technology, China.

His research interests include data mining, business intelligence, and accounting information system, which are funded by the National Natural Science Foundation of China under grant #70372024, #70771044 and #70872020. He has published over 40 papers in journals or academic conferences, such as Journal of Software, Pattern Recognition and Artificial Intelligence, Journal of Computational Information Systems and Control and Decision. He is the author or co-author of numerous monographs and books, such as Continuous Temporal Data Mining and Implementation Mechanism, published in Chinese by Economic Science Press, 2008.

He is a professor of information system and deputy director of the Center for Business Intelligence Research at the Jinan University, Guangzhou, China. He has over 28 years experience developing information systems in China and Japan, and about 10 years experience as a consultant for business intelligence. He was certified as a Chinese System Analyst in 1995.

He has also won the awards in the association: one Chinese National Science and Technology Progressive Award, two Chinese National Science and Technology Key Project Major Awards, one Chinese Book Award, three Provincial or Ministerial Science and Technology Progressive Awards and one Provincial Social Science Award.