Combined Mining: Discovering Informative Knowledge in Complex Data

Longbing Cao, *Senior Member, IEEE*, Huaifeng Zhang, *Member, IEEE*, Yanchang Zhao, *Member, IEEE*, Dan Luo, and Chengqi Zhang, *Senior Member, IEEE*

***Abstract*—Enterprise data mining applications often involve complex data such as multiple large heterogeneous data sources, user preferences, and business impact. In such situations, a single method or one-step mining is often limited in discovering informative knowledge. It would also be very time and space consuming, if not impossible, to join relevant large data sources for mining patterns consisting of multiple aspects of information. It is crucial to develop effective approaches for mining patterns combining necessary information from multiple relevant business lines, catering for real business settings and decision-making actions rather than just providing a single line of patterns. The recent years have seen increasing efforts on mining more informative patterns, e.g., integrating frequent pattern mining with classifications to generate frequent pattern-based classifiers. Rather than presenting a specific algorithm, this paper builds on our existing works and proposes *combined mining* as a general approach to mining for informative patterns combining components from either multiple data sets or multiple features or by multiple methods on demand. We summarize general frameworks, paradigms, and basic processes for *multifeature combined mining*, *multisource combined mining*, and *multimethod combined mining*. Novel types of combined patterns, such as incremental cluster patterns, can result from such frameworks, which cannot be directly produced by the existing methods. A set of real-world case studies has been conducted to test the frameworks, with some of them briefed in this paper. They identify combined patterns for informing government debt prevention and improving government service objectives, which show the flexibility and instantiation capability of combined mining in discovering informative knowledge in complex data.**

***Index Terms*—Actionable knowledge discovery, combined mining, complex data, data mining, multiple source data mining, public service data mining.**

I. Introduction

ENTERPRISE data mining applications, such as mining public service data and telecom fraudulent activities, inevitably involve complex data sources, particularly multiple large scale, distributed, and heterogeneous data sources embedding information about business transactions, user preferences, and business impact. In these situations, business people certainly expect the discovered knowledge to present a full picture of business settings rather than one view based on a single source. Knowledge reflecting full business settings is more business friendly, comprehensive, and informative for business decision makers to accept the results and to take operable actions accordingly. With the accumulation of ubiquitous enterprise data, there is an increasing need to mine for such informative knowledge in complex data.

It is challenging to mine for comprehensive and informative knowledge in such complex data suited to real-life decision needs by using the existing methods. The challenges come from many aspects, for instance, the traditional methods usually discover homogeneous features from a single source of data while it is not effective to mine for patterns combining components from multiple data sources. It is often very costly and sometimes impossible to join multiple data sources into a single data set for pattern mining. In Section III, we discuss the challenges in more detail after introducing an example.

The aforementioned discussions show the need for developing effective techniques for involving multiple heterogeneous features, data sets, and methods in enterprise data mining. As we will discuss in Section II about related works, the existing works in handling the aforementioned challenges can be categorized into the following aspects: 1) data sampling; 2) joining multiple relational tables; 3) post analysis and mining; 4) involving multiple methods; and 5) mining multiple data sources. In real-life data mining, data sampling is often not acceptable since it may miss important data that are filtered out. Table joining may not be possible due to the time and space limit such as in dealing with hundreds of millions of transactions from multiple sources in our case studies. In addition, techniques for involving multiple methods and handling multiple data sources are often specifically developed for particular cases.

In [1], [22], and [25], we proposed the concepts of *combined association rules*, *combined rule pairs*, and *combined rule clusters* to mine for informative patterns in complex data by catering for the comprehensive aspects in multiple data sets. A combined association rule is composed of multiple heterogeneous itemsets from different data sets while combined rule pairs and combined rule clusters are built from combined association rules. Analysis shows that such combined rules cannot be directly produced by traditional algorithms such as the FPGrowth [15]. This paper builds on the existing works and proposes the approach of *combined mining* as a general method for directly identifying patterns enclosing constituents from multiple sources or with heterogeneous features such as covering demographics, behavior, and business impacts. Its deliverables are *combined patterns* such as the aforementioned combined association rules. Combined patterns consist of

Manuscript received January 16, 2010; revised August 11, 2010; accepted October 2, 2010. Date of publication November 9, 2010; date of current version May 18, 2011. This work was supported in part by the Australian Research Council under Discovery Grants DP1096218 and DP0988016 and Linkage Grants LP100200774 and LP0989721. This paper was recommended by Associate Editor N. Chawla.

L. Cao is with the University of Technology, Sydney (UTS), Sydney, NSW 2007, Australia, and also with the Australian Capital Markets Cooperative Research Centre, Sydney, NSW 2001, Australia.

H. Zhang and Y. Zhao are with Centrelink, Canberra, ACT 2612, Australia.

D. Luo and C. Zhang are with the University of Technology, Sydney (UTS), Sydney, NSW 2007, Australia.

Digital Object Identifier 10.1109/TSMCB.2010.2086060

multiple components, a pair or cluster of atomic patterns, identified in individual sources or based on individual methods.

The general ideas of combined mining are as follows.

1) By involving multiple heterogeneous features, combined patterns are generated which reflect multiple aspects of concerns and characteristics in businesses.
2) By mining multiple data sources, combined patterns are generated which reflect multiple aspects of nature across the business lines.
3) By applying multiple methods in pattern mining, combined patterns are generated which disclose a deep and comprehensive essence of data by taking advantage of different methods.
4) By applying multiple interestingness metrics in pattern mining, patterns are generated which reflect concerns and significance from multiple perspectives.

Rather than presenting a specific algorithm for mining a particular type of combined patterns, this paper focuses on abstracting several general and flexible frameworks from the architecture perspective, which can foster wide implications and particularly can be instantiated into many specific methods and algorithms to mine for various patterns in complex data.[1] The main contributions of this paper are as follows:

1) building on existing works, generalizing the concept of combined mining that can be expanded and instantiated into many specific approaches and models for mining complex data toward more informative knowledge;
2) discussing general frameworks and their paradigms and basic processes of *multifeature* and *multimethod combined mining* for supporting combined mining, which contribute to *multisource combined mining*—they are flexible to be instantiated into specific needs;
3) proposing various strategies for conducting pattern interaction when instantiating the aforementioned proposed frameworks—as a result, novel combined pattern types, such as incremental cluster patterns, can result from combined mining, which have not been investigated before;
4) illustrating the corresponding interestingness metrics for evaluating certain types of combined patterns;
5) demonstrating the use of combined mining in discovering combined patterns in real-world government service data for government debt prevention in an Australian Commonwealth Government agency.

This paper is organized as follows. Section III presents an example to illustrate the problem of mining combined patterns. In Section IV, we introduce the basic concepts, paradigms, and processes of combined mining. Section V presents the basic frameworks and procedures of the multifeature combined mining approach. Multimethod combined mining is introduced in Section VI in which we present two existing basic frameworks which are the parallel and serial multimethod combined mining. Five case studies are briefly discussed in Section VII, which conduct combined pattern mining in public service data. Related work is discussed in Section II. This paper is concluded in Section VIII.

[1]Readers who are interested in specific methods may refer to our relevant works [1], [22], [24], [25] about specific algorithms, examples, and performance evaluation in mining combined patterns.

II. Related Work

First, most of existing single-handed data mining methods do not target the discovery of informative patterns in complex data, as discussed in this paper. For instance, a combined association rule R is in the form of $R: A_1 \ldots \wedge A_i \wedge B_1 \wedge \ldots B_j \rightarrow T$, where $A_i \in D_i$ and $B_j \in D_j$ are itemsets in heterogeneous data sets D_i and D_j, respectively, $T \neq \varnothing$ is a target item or class, and $\forall i,\ j,\ A_i \neq \varnothing,\ B_j \neq \varnothing,\ A_i \neq B_j$. Analysis shows that such combined rules cannot be directly produced by traditional algorithms such as the FPGrowth [15].

Second, approaches to mining for more informative and actionable knowledge in complex data can be generally categorized as follows: 1) direct mining by inventing effective approaches; 2) postanalysis and postmining of learned patterns; 3) involving extra features from other data sets; 4) integrating multiple methods; and 5) joining multiple relational tables.

Direct mining for discriminative patterns has been highlighted, such as in Harmony [19], model-based search tree [6], and emerging contrast patterns [5]. Combined mining contributes to this category too. In the following paragraphs, we briefly discuss the work related to the other four approaches and explain what difference combined mining can make.

The postanalysis and postmining of learned patterns is a commonly used approach [28], for instance, to prune rules [12], reduce redundancy [10], and summarize learned rules [12]. Different from postanalysis-based methods, most of the combined patterns introduced in this paper can be generated directly. Some of them can be identified through the postanalysis. Aside from the direct mining of combined patterns, the postmining of the identified patterns can be conducted where necessary in order to make the patterns more actionable. For example, the multifeature combined mining approach considers features from multiple data sets during the direct generation of more informative patterns.

Further taking cluster patterns as an example, in this paper, they are not generated by pattern summarization. Cluster patterns are mined through the methods discussed in Section V-C. The patterns in a cluster have the same prefix or postfix, but the remaining items in the patterns make the results different. Our method can generate incremental and decremental combined clusters as well as pairs. However, the current methods mainly target contrast patterns, emerging patterns, etc., which are much simpler. Moreover, many methods, such as emerging pattern mining, cannot efficiently deal with a large scale of data.

The integration of multiple data mining methods is widely used to mine for more informative knowledge, such as associative classification [13], combining clustering and association rules for rarity mining [16], combining regression with association rule mining [14], the integration of boosting with associative classifiers [17], sequence classification [15], [23], and association-rule-mining-based classification [7]. A typical challenge is that a huge amount of sequential patterns is usually mined in the sequential mining procedure. Although pruning algorithms are used for postprocessing, there is still a large amount of sequential patterns constructing the feature space. Moreover, existing algorithms often do not tackle important problems such as how to efficiently and effectively select discriminative features from the large feature space. This issue is handled in our closed-loop sequence classification method.

TABLE I
CUSTOMER DEMOGRAPHIC DATA (F—FEMALE, M—MALE)

Customer ID	Gender	...
1	F	
4	M	

TABLE II
TRANSACTIONAL DATA MIXING ORDERED AND UNORDERED DATA (Y—LEADING TO DEBT, N—NO DEBT IS GENERATED)

Customer ID	Policies	Activities	Debt
1	(c_1, c_2)	$(a_1 - a_2)$	Y
4	(c_2, c_4)	$(a_1 - a_2 - a_5)$	N
4	(c_1, c_2, c_4)	$(a_1 - a_3)$	N
4	(c_1, c_2, c_5)	$(a_1 - a_3 - a_4)$	Y
4	(c_2, c_3)	$(a_2 - a_4)$	N

TABLE III
TRADITIONAL ASSOCIATION RULES

Rules	Supp	Conf	Lift
$c_1 \rightarrow Y$	4/10	4/6	1.3
$c_1 \rightarrow N$	2/10	2/6	0.7
$c_2 \rightarrow Y$	4/10	4/8	1
$c_2 \rightarrow N$	4/10	4/8	1

TABLE IV
TRADITIONAL SEQUENTIAL PATTERNS

Rules	Supp	Conf	Lift
$a_1 \rightarrow Y$	3/10	3/7	0.9
$a_1 \rightarrow N$	4/10	4/7	1.1
$a_1 - a_2 \rightarrow Y$	1/10	1/2	1
$a_1 - a_3 \rightarrow N$	3/10	3/5	1.2

TABLE V
COMBINED ASSOCIATION RULES

Rules	Supp	Conf	Lift	$Cont$	I_{rule}
$F \wedge c_1 \rightarrow N$	2/10	1/2	1	1	1.4
$F \wedge c_2 \rightarrow Y$	2/10	2/3	1.3	1.3	1.3
$M \wedge c_2 \rightarrow N$	3/10	3/5	1.2	1.2	1.2
$M \wedge c_2 \rightarrow Y$	2/10	2/5	0.8	0.8	0.8

Table joining is widely used in order to mine patterns from multiple relational tables by putting relevant features from individual tables into a consolidated one. As a result, a pattern may consist of features from multiple tables. This method is suitable for mining multiple relational databases, particularly for small data sets. However, enterprise applications often involve multiple heterogeneous data sets consisting of large volumes of records. In the real world, it is too costly in terms of time and space, if not impossible, to join multiple sources of distributed data. Combined mining can identify such compound patterns in large data sets.

Multirelational data mining [4] and multidatabase mining [20] have been intensively studied. They are different from combined mining. Our method is not only aimed at multirelational data, but it is also a general approach for mining complex knowledge in complex data. As the multisource combined mining shows, combined mining does not rely on joining related tables. The resulting patterns of the multisource combined mining can consist of pairs or clusters of patterns with components from multiple data sets, which is new to multirelational mining, to the best of our knowledge.

A typical difference between the combined mining and other existing methods is that new pattern types can be produced such as incremental/decremental cluster patterns that are not previously identified.

III. EXAMPLE

A. Example: Combined Association Rules

Here, we explain the mining of combined patterns by illustrating a task that we did in detecting customer debts for the Australian Commonwealth Government Agency Centrelink [29]. The task involves a large scale of real-world complex public service data, including customer demographic information (see Table I), unordered government policies applied on customers, ordered customer activities, and the impact of customers on government service objectives, namely, whether a customer incurs debts or not. For simplicity, let us assume that they are small and can be merged into one table as shown in Table II (in the case studies in Section VII, we do not join these data sets simply because each is too large), indicating whether a customer has a debt or not (represented by "Y" for yes or "N" for no) under various policies or activities. For instance, Centrelink has policies stating that customers should report their income fortnightly or irregularly, depending on different allowances. Various activities are also conducted with customers, e.g., reviewing from Centrelink, reminder letters sent to customers, and to intervene in debts once identified. In the following paragraphs, we explain how traditional data mining methods and combined mining can be applied on the data and what difference of their findings could be.

Traditionally, such data are mined individually by frequent pattern mining or classification conducted on the unordered policy data and ordered activity data, respectively. For instance, when association mining is used to mine frequent rules, the rules shown in Table III can be discovered from the unordered transactional data set. Similarly, we can identify frequent sequential patterns as shown in Table IV.

However, such single frequent patterns are not informative because they do not reflect the full picture and reality of business and are simplified and separated from real business settings in which unordered policies and ordered activities are closely related. Their actionability is not strong enough to support business decision making and satisfy user needs.

We now explain the use of combined mining to produce more informative and actionable patterns.

1) We partition the whole population into male and female groups, based on the demographic data in Table I, and then mine the demographic and transactional data of the two groups separately, as partially shown in Table V, where $Cont$ denotes the contribution of the transactional data and I_{rule} reflects the interestingness of the combined rules. (The definitions of $Cont$ and I_{rule} will be given in Section V-A.) We can derive from Table V the following observations: 1) Rules are more informative than those in

TABLE VI
COMBINED ASSOCIATION RULE PAIRS

Pairs	Combined Rules	I_{pair}
$\mathcal{P}_1$	$M \wedge c_3 \to Y,\ M \wedge c_2 \to N$	0.55
$\mathcal{P}_2$	$F \wedge c_2 \to Y,\ M \wedge c_2 \to N$	0.63

TABLE VII
COMBINED FREQUENT PATTERNS WITH BOTH UNORDERED AND ORDERED ITEMSETS

Rules	Supp	Conf	Lift	*Cont*	I_{rule}
$M \wedge c_2 \wedge a_1 - a_3 \to Y$	2/10	2/3	1.3	1.6	2

TABLE VIII
CLASSIFICATION ON FREQUENT COMBINED PATTERNS

Customer ID	Policies	Activities	Prediction
1	(c_1, c_2)	$(a_1 - a_3)$	Y
2	$(c_2, c_4$	(a_1)	N
...	...	...	

Table III because they reflect multiple aspects of business, and 2) more discriminative rules with high confidence and lift can be found by combining rules from two separate data sets.

2) It is more interesting to organize the rules into contrast pairs as shown in Table VI, where I_{pair} is the interestingness of a rule pair. For instance, $\mathcal{P}_1$ is a rule pair for the male group, and it shows that c_3 is associated with debt but c_2 is not. $\mathcal{P}_1$ is actionable in that it suggests that c_2 is a preferred policy to replace c_3 and avoid the debt raised on male customers. Moreover, male customers should be excluded when initiating policy c_3. $\mathcal{P}_2$ is a rule pair with the same policy but different demographics. With action c_2, male customers have no debts while the females tend to have debts. It suggests that c_2 is a preferable policy for males but undesirable for females.

   A simple way to find the rules in Table V is to join Tables I and II and then apply traditional association mining to the derived table. Unfortunately, it is often infeasible from both the time and space perspectives to do so for enterprise applications involving multiple heterogeneous data sets with each of them consisting of hundreds of millions of records.

3) Frequent patterns combining unordered and ordered items can be identified as shown in Table VII. From Table V, we know that male customers under policy c_2 do not tend to have debts. However, in Table VII, we can see that, if activities $a_1 - a_3$ are taken, the male customers under policy c_2 are very likely to have a debt since its I_{pair} is as high as 2. Obviously, the ordered activity data set provides much richer information to allow a more reasonable decision.

4) The classification can be further conducted on the identified frequent pattern sets. Table VIII shows some examples of the frequent patterns which are further used for classification. Different from the features in conventional classification on the demographic data, both the ordered and unordered transactional data are used for prediction. The resulting classification is much more informative.

B. Discussions

The aforementioned example shows the challenge in mining for informative knowledge in complex data. The challenges come from many aspects, for instance, as follows.

1) Patterns identified by traditional methods usually only involve homogeneous features from a single source of data, e.g., frequent patterns of customer shopping habits. Such patterns consist of a single line of information and are not informative in business decision making. If attributes from multiple aspects can be included, the resulting patterns can then completely reflect the business situation and be workable in supporting business decision making. For instance, in basket analysis, the shopper personal information and prices of goods and items can complement frequent itemsets with additional information for a shop owner to take intervention actions.
2) It is often costly and space consuming, and sometimes impossible, to directly join a large amount of multiple, distributed, and heterogeneous data sets for centralized pattern analysis without strategic arrangement [30]. Alternatively, often, a single line of business data is mined through data summarization, sampling, and partitioning, although the resulting patterns are not informative enough and do not reflect the full business picture. As a result, their decision-support power is limited or weakened. For instance, demographic data show different characteristics from sequential behavioral data. They increase the complexity of pattern mining if they are merged into a single data set.
3) In mining multiple heterogeneous data sets, a single method is often not powerful to generate results sophisticated enough to match real-world comprehensive scenarios. For instance, sequence analysis is suitable for analyzing behaviors in the aforementioned example while demographic data can be inspected through associations or clustering.

In the following sections, we propose several frameworks for discovering informative knowledge in complex data. They are built on existing works and substantial experiments in real-life enterprise data mining projects sponsored by the Australian Research Council and our industry partners in areas such as government services, banking, health insurance, and capital markets (see more information about the projects from [31]). We want to show the analysis and discussions (see [1], [22], [24], and [25]) about the following advantages of combined mining in discovering informative knowledge in complex data, compared to a single use of existing methods.

1) Flexible frameworks for combining multifeatures, multisources, and multimethods covering various needs in mining complex data, which are customizable for specific cases. With combined mining, the advantage of specific algorithms can be well taken in handling particular tasks.
2) Effective in discovering patterns with constituents from multiple heterogeneous sources and a large scale of real-life data, which can provide patterns reflecting a full picture rather than a single line of business.
3) Novel combined patterns can be produced which cannot be identified by directly applying existing methods.

IV. Concept of Combined Mining

A. Basic Concepts

For a given business problem (Ψ), we suppose that the key entities associated with it in discovering interesting knowledge for business decision support are as follows: data set $\mathcal{D}$ collecting all data relevant to a business problem, feature set $\mathcal{F}$ including all features for data mining, method set $\mathcal{R}$ consisting of all data mining methods that can be used on the data $\mathcal{D}$, interestingness set $\mathcal{I}$ composed of all measures from all methods $\mathcal{R}$, impact set $\mathcal{T}$ referring to business impacts or outcomes such as fraud or nonfraud, and pattern set $\mathcal{P}$. They are described as follows.

1) Data set $\mathcal{D}$: $\mathcal{D} = \{\mathcal{D}_k; k = 1, \ldots, K\}$ consists of all K subdata sets relevant to the underlying business problem, and X_k is the set of all items in the data set $\mathcal{D}_k \forall k \neq j$, $X_k \cap X_j = \emptyset$.
2) Feature set $\mathcal{F}$: $\mathcal{F} = \{\mathcal{F}_k; k = 1, \ldots, K\}$ refers to all features used for pattern mining on K subdata sets, where $\mathcal{F}_k$ is the feature set corresponding to the data set $\mathcal{D}_k$.
3) Method set $\mathcal{R}$: $\mathcal{R} = \{\mathcal{R}_l; l = 1, \ldots, L\}$, where $\mathcal{R}_l$ is a data mining method set deployed on the data set $\mathcal{D}_k$ involving the feature set $\mathcal{F}_k$.
4) Interestingness set $\mathcal{I}$: $\mathcal{I} = \{\mathcal{I}_{m,l}; m = 1, \ldots, M; l = 1, \ldots, L\}$, where $\mathcal{I}_{m,l}$ is an interestingness metric set corresponding to a particular data mining method $\mathcal{R}_l$, which is associated with m interestingness metrics. Suppose that $\mathcal{I}'_k \subset \mathcal{I}$, where $\mathcal{I}'_k$ is the interestingness used by methods $\mathcal{R}'_k$.
5) Impact set $\mathcal{T}$: $\mathcal{T} = \{\mathcal{T}_j; j = 1, \ldots, J\}$ consists of the categorized business impacts associated with certain patterns; in some cases, impacts can be categorized into *impact* (T) and *nonimpact* $(\bar{T})$, for instance, fraud or nonfraud. If a pattern is associated with an impact (T), represented by $X \rightarrow T$, then we call it an *impact-oriented pattern*. Similarly, if a pattern is mainly relevant to nonimpact, indicated by $X \rightarrow \bar{T}$, we call it a *nonimpact-oriented pattern*.
6) Pattern set $\mathcal{P}$: $\mathcal{P} = \{\mathcal{P}_{n,m,l}; n = 1, \ldots, N; m = 1, \ldots, M; l = 1, \ldots, L\}$, where $\mathcal{P}_{n,m,l}$ is an atomic pattern set resulting from data mining method $\mathcal{R}_l$ using interestingness $\mathcal{I}_{m,l}$; there are $n (n < N)$ atomic patterns in the set. Suppose that $\mathcal{P}'_k \subset \mathcal{P}$, where $\mathcal{P}'_k$ is the pattern set identified on $\mathcal{D}_k$ using methods $\mathcal{R}'_k$ and interestingness $\mathcal{I}'_k$.

Based on the aforementioned variables, a general pattern discovery process can be described as follows: Patterns $\mathcal{P}_{n,m,l}$ are identified through data mining method $\mathcal{R}_l$ deployed on features $\mathcal{F}_k$ from a data set $\mathcal{D}_k$ in terms of interestingness $\mathcal{I}_{m,l}$

$$\mathcal{P}_{n,m,l} : \mathcal{R}_l(\mathcal{F}_k) \rightarrow \mathcal{I}_{m,l} \tag{1}$$

where $n = 1, \ldots, N; m = 1, \ldots, M; l = 1, \ldots, L$.

Combined mining is a process defined as follows.

Definition 1 (Combined Mining): Combined mining is a two-to-multistep data mining procedure, consisting of the following:

1) Mining atomic patterns $\mathcal{P}_{n,m,l}$ as described in (1).
2) Merging atomic pattern sets into combined pattern set $\mathcal{P}'_k = \mathcal{G}_k(\mathcal{P}_{n,m,l})$ for each data set $\mathcal{D}_k$ by pattern merging method $\mathcal{G}_k; \mathcal{G}_k \in \mathcal{G}$, where $\mathcal{G}$ includes a set of pattern-merging methods suitable for a particular business problem.
3) If multiple data sets are involved, combined patterns identified in specific data sets are then further merged into the combined pattern set $\mathcal{P} = \mathcal{G}(\mathcal{P}'_k)$.

From a high-level perspective, combined mining represents a generic framework for mining complex patterns in complex data as follows:

$$\mathcal{P} := \mathcal{G}(\mathcal{P}_{n,m,l}) \tag{2}$$

in which atomic patterns $\mathcal{P}_{n,m,l}$ from either individual sources $\mathcal{D}_k$, individual methods $\mathcal{R}_l$, or particular feature sets $\mathcal{F}_k$ are combined into groups with the members closely related to each other in terms of pattern similarity or difference.

In combined mining, the word "combined" principally refers to either one or more of the following aspects on demand.

1) The combination of multiple data sources $(\mathcal{D})$: The combined pattern set $\mathcal{P}$ consists of multiple atomic patterns identified in several data sources, respectively, namely, $\mathcal{P} = \{\mathcal{P}'_k | \mathcal{P}'_k : \mathcal{I}'_k(X_j); X_j \in \mathcal{D}_k\}$; for example, demographic data and transactional data are two data sets involved in mining for demographic–transactional patterns.
2) The combination of multiple features $(\mathcal{F})$: The combined pattern set $\mathcal{P}$ involves multiple features, namely, $\mathcal{P} = \{\mathcal{F}_k | \mathcal{F}_k \subset \mathcal{F}, \mathcal{F}_k \in \mathcal{D}_k, \mathcal{F}_{j+k} \in \mathcal{D}_{j+k}; j, k \neq 0\}$, e.g., features of customer demographics and behavior.
3) The combination of multiple methods $(\mathcal{R})$: The patterns in the combined set reflect the results mined by multiple data mining methods, namely, $\mathcal{P} = \{\mathcal{P}'_k | \mathcal{R}'_k \rightarrow \mathcal{P}'_k\}$, for instance, association mining and classification.

B. Basic Paradigms

In this section, we briefly introduce some basic paradigms of combined mining. This involves combined pattern types, structures formed by atomic patterns, and relationships and time frames among atomic patterns.

From the pattern type perspective, combined patterns can be classified into *nonimpact-oriented combined patterns* (NICPs) and *impact-oriented combined patterns* (ICPs), depending on whether a pattern is associated with a certain target item or business impact. For a NICP, its itemsets are associated with each other under certain interestingness metrics while we do not bother about the impact of the pattern on the business outcome.

$$\mathcal{P}_n : R_l(X_1 \wedge \cdots \wedge X_i) \rightarrow I_m \tag{3}$$

$$\mathcal{P} := \mathcal{G}(P_1 \wedge \cdots \wedge P_n) \rightarrow \mathcal{I} \tag{4}$$

An ICP is associated with either a *target* itemset or *resulting impact* $(T_j; T_j \subset T$, where T is the target or impact set).

$$P_n : \{R_l(X_1 \wedge \cdots \wedge X_i) \rightarrow I_m\} \rightarrow T_1 \tag{5}$$

$$\mathcal{P} := \mathcal{G}(P_1, \cdots, P_n) \tag{6}$$

The *number* of the constituent atomic patterns in a combined pattern can vary. For example, the following list enumerates two kinds of general structures.

1) Pair patterns: $\mathcal{P} ::= \mathcal{G}(P_1, P_2)$, where two atomic patterns P_1 and P_2 are correlated to each other in terms of pattern merging method $\mathcal{G}$ into a pair. From such patterns, contrast and emerging patterns [5] can be further identified.
2) Cluster patterns: $\mathcal{P} ::= \mathcal{G}(P_1, \ldots, P_n)(n > 2)$, where more than two patterns are correlated to each other in terms of pattern merging method $\mathcal{G}$ into a cluster. A group of patterns, such as combined association clusters [25], can be further discovered.

Furthermore, the *structural relationships* governing the constituent patterns in a combined pattern set can be multiform, and we list a few examples as follows.

1) *Peer-to-peer* relation: As illustrated by $\mathcal{P} ::= P_1 \cup P_2$ in which P_1 and P_2 take equal positions in the pair, the pattern exists due to reasons such as similarity or difference from structural or semantic relationship perspectives.
2) *Master–slave* relation: This is also called the *underlying-derivative* relation. An example is $\{\mathcal{P} ::= P_1 \cup P_2, P_2 = f(P_1)\}$ in which the existence of pattern P_2 is subject to that of P_1 in terms of function f; an example is $P_2 = P_1 + \Delta P$, where ΔP is the additional part appending to P_1.
3) *Hierarchy* relation: This is illustrated by $\{\mathcal{P} ::= P_i \cup P_i' \cup P_j \cup P_j', P_j = \mathcal{G}(P_i), \ldots, P_j' = \mathcal{G}'(P_i)'\}$ in which some patterns are correlated in terms of relationship $\mathcal{G}$ while others with $\mathcal{G}'$ or something else.

From the *time frame* perspective, patterns may be correlated in terms of different temporal relationships, for instance, as follows:

1) *independent* relation: as illustrated by $\{P_1 : P_2\}$ in which P_1 and P_2 occur independently from the time perspective;
2) *concurrent* relation: as illustrated by $\{P_1 \| P_2\}$ in which P_1 and P_2 occur concurrently;
3) *sequential* relation: as illustrated by $\{P_1; P_2\}$ in which P_2 happens after the occurrence of P_1;
4) *hybrid* relation: as illustrated by $\{P_1 \otimes P_2 \cdots \otimes P_n; \otimes \in \{:, \|, ;\}\}$—there are more than two patterns existing in $\mathcal{P}$ in which some of them happen concurrently ($|$) or independently (:) while others occur sequentially (;).

In Sections V and VI, we will illustrate some of the aforementioned pattern types and relationships.

C. Basic Process

This section discusses a general process of combined mining. Real-world enterprise applications often involve multiple heterogeneous and distributed data sets that cannot or are too costly to be integrated. Another common situation is where the data volume is so large that it cannot be handled by scanning the whole data set. Such data have to be partitioned into either small and manageable sets or in terms of business categories such as billing, networking, and accounting data in telecommunication systems. Mining such complex data requires the handling of multidata sources implicitly or explicitly.

Fig. 1 illustrates a framework for combined mining [2]. It supports the discovery of combined patterns either in multiple data sets or subsets $(\mathcal{D}_1, \ldots, \mathcal{D}_K)$ through data partitioning in the following manner: 1) Based on domain knowledge, business understanding, and goal definition, one of the data sets or certain partial data (say $\mathcal{D}_1$) are selected for mining exploration $(\mathcal{R}_1)$; 2) the findings are used to guide either data partition or data set management through the data *coordinator* and to design strategies for managing and conducting serial or parallel pattern mining on relevant data sets or subsets or mining respective patterns on relevant remaining data sets; the deployment of method $\mathcal{R}_k$ $(k = 2, \ldots, L)$, which could be either in parallel or through combination, is informed by the understanding of the data/business and objectives, and if necessary, another step of pattern mining is conducted on data set $\mathcal{D}_k$ with the supervision of the results from step $k-1$; and 3) after finishing the mining of all data sets, patterns $(P^{\mathcal{R}_n})$ identified from individual data sets are merged $(\mathcal{G}\{P_n\})$ with the involvement of domain knowledge and further extracted into final deliverables $(\mathcal{P})$.

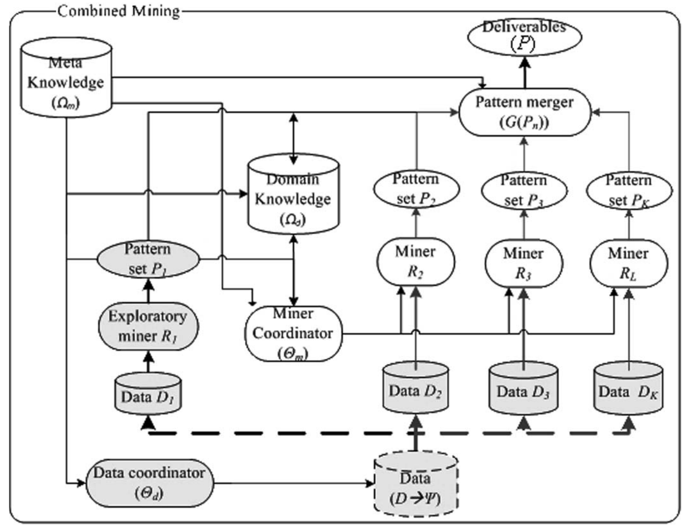


Fig. 1. Combined mining for actionable patterns.

The aforementioned process can be expressed as follows:

$$CM ::= \underbrace{\mathcal{D}_k[\mathcal{D} \xrightarrow{\otimes} \mathcal{D}_k]^{\mathcal{I}_k, \mathcal{R}_k, \Omega_m} \{\mathcal{P}_k\}}_{K} \longrightarrow^{\mathcal{G}^N \mathcal{P}_k, \Omega_d, \Omega_m} \mathcal{P} \tag{7}$$

where $\mathcal{I}_k$ is the interestingness of data mining method $\mathcal{R}_k$ on data set/subset $\mathcal{D}_k$ and $\otimes$ indicates data partition if the source data need to be split.

For instance, if multiple data sources are involved in combined mining, the process can be further expressed as follows:

PROCESS: Multisource Combined Mining
INPUT: target data sets $\mathcal{D}_k$ $(k = 1, \ldots, K)$, business problem Ψ
OUTPUT: combined patterns $\mathcal{P}$
Step 1: Identify a suitable data set or data part, for example, $\mathcal{D}_1$ for initial mining exploration.
Step 2: Identify the next suitable data set for pattern mining, or partition whole source data into K data sets supervised by the findings in Step 1.
Step 3: *Data set-kmining*: Extract atomic patterns $\mathcal{P}_k$ on data set/subset D_k.

FOR $k = 1$ to K

Develop modeling method $\mathcal{R}_k$ with interestingness $\mathcal{I}_k$.

Employ method $\mathcal{R}_k$ on the environment e and data $\mathcal{D}_k$ engaging metaknowledge Ω_m.

Extract the atomic pattern set $\mathcal{P}_k$.

ENDFOR

Step 4: *Pattern merger*: Merge atomic patterns into combined pattern set $\mathcal{P}$.

FOR $k = 1$ to K

Design the pattern merger functions $\mathcal{G}_k$ to merge all relevant atomic patterns into $\mathcal{P}_k$ by involving domain and metaknowledge Ω_d and Ω_m and interestingness $\mathcal{I}$.

Employ the method $\mathcal{G}(\mathcal{P}_k)$ on the pattern set $\mathcal{P}_k$.

Generate combined patterns into set $\mathcal{P} = \mathcal{G}_k(\mathcal{P}_k)$.

ENDFOR

Step 5: Enhance pattern actionability to generate deliverables $\mathcal{P}$.

Step 6: Output the deliverables $\mathcal{P}$.

The aforementioned framework can be instantiated into a number of mutations. For instance, for a large volume of data, combined mining can be instantiated into *data partition+unsupervised+supervised combined mining* by integrating data partition into combined mining. First, the whole data set is partitioned into several subsets based on the data/business understanding and domain knowledge jointly by data miners and domain experts, e.g., data sets 1 and 2. Second, *unsupervised learning* is developed to mine one of the preference data sets, for example, data set 1. Some of the mined results are then used to design new variables for processing the other data set. *Supervised learning* is further conducted on data set 2 to generate actionable patterns by checking both the technical and business performance. Finally, the individual patterns mined from both data sets are combined into pattern deliverables.

V. Multifeature Combined Mining

A. MFCPs

In multifeature combined pattern (MFCP) mining, a combined pattern is composed of heterogeneous features of different data types, such as binary, categorical, ordinal, and numerical, or of different data categories, such as customer demographics, transactions, and time series.

Definition 2 (MFCPs): Assuming that $\mathcal{F}_k$ denotes the set of features in data set $\mathcal{D}_k \forall i \neq j$, $\mathcal{F}_{k,i} \cap \mathcal{F}_{k,j} = \emptyset$, based on the variables defined in Section IV-A, an MFCP P is in the form of

$$\begin{aligned} &\mathcal{P}_k : \mathcal{R}_l(\mathcal{F}_1, \cdots, \mathcal{F}_k) \\ &\mathcal{P} := \mathcal{G}_F(\mathcal{P}_k) \end{aligned} \tag{8}$$

where $\exists i,\ j,\ i \neq j,\ \mathcal{F}_i \neq \emptyset,\ \mathcal{F}_j \neq \emptyset$, and $\mathcal{G}_F$ is the merging method for the feature combination.

As shown in Section III, an MFCP example is $F \wedge c_1 \wedge a_1 - a_2 \rightarrow N$. It combines one demographic component to many items from transactional data sets and business outcomes, e.g., whether it indicates significant impact of leading to government debt or not in Centrelink.

TABLE IX
Support, Confidence, and Lift of Pattern $X \rightarrow T$

Support	$Prob(X \wedge T)$
Confidence	$Prob(X \wedge T)/Prob(X)$
Lift	$Prob(X \wedge T)/(Prob(X) * Prob(T))$

New evaluation metrics may be necessary to measure the interestingness of the ICP. For instance, given a single combined pattern $\mathcal{P} : X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T$, the traditional $support$, $confidence$, and $lift$ are given in Table IX.

In selecting actionable combined patterns, the contribution of the aforementioned traditional interestingness measures is limited. Based on traditional $support$, $confidence$, and $lift$, two new metrics $contribution$ and I_{rule} are designed as follows for measuring the interestingness of a single combined pattern.

Definition 3 (Contribution): For an MFCP $P : X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T$, the $contribution$ of X_{e} to the occurrence of outcome T in rule P is

$$Cont_{\mathrm{e}}(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T) = \frac{Lift(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Lift(X_{\mathrm{p}} \rightarrow T)} \tag{9}$$

$$= \frac{Conf(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Conf(X_{\mathrm{p}} \rightarrow T)}. \tag{10}$$

$Cont_{\mathrm{e}}(P)$ is the lift of X_{e} with X_{p} as a precondition, which shows how much X_{e} contributes to the rule. $Contribution$ can be taken as the increase of $lift$ by appending additional items X_{e} to a rule. Its value falls in $[0, +\infty)$. A $contribution$ greater than one means that the additional items in the rule contribute to the occurrence of the outcome, and a $contribution$ less than one suggests that it incurs a reverse effect.

Based on the aforementioned definition of $contribution$, the interestingness of a single combined pattern is defined as follows:

$$I_{\mathrm{rule}}(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T) = \frac{Cont_{\mathrm{e}}(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Lift(X_{\mathrm{e}} \rightarrow T)}. \tag{11}$$

I_{rule} indicates whether the $contribution$ of X_{p} (or X_{e}) to the occurrence of T increases with X_{e} (or X_{p}) as a precondition. Therefore, "$I_{\mathrm{rule}} < 1$" suggests that $X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T$ is less interesting than $X_{\mathrm{p}} \rightarrow T$ and $X_{\mathrm{e}} \rightarrow T$. The value of I_{rule} falls in $[0, +\infty)$. When $I_{\mathrm{rule}} > 1$, the higher I_{rule} is, the more interesting the rule is.

B. Pair Pattern

Two atomic patterns or two patterns identified by combined mining may be merged into a pair, forming a *combined pair pattern* (or simply *pair pattern*), defined as follows.

Definition 4 (Pair Pattern): For impact-oriented combined mining, a pair pattern is in the form of

$$\mathcal{P} : \begin{cases} X_1 \rightarrow T_1 \\ X_2 \rightarrow T_2 \end{cases} \tag{12}$$

where $X_1 \cap X_2 = X_{\mathrm{p}}$ with X_{p} being the *prefix* of pair $\mathcal{P}$, $X_{1,e} = X_1 \setminus X_{\mathrm{p}}$, $X_{2,e} = X_2 \setminus X_{\mathrm{p}}$, X_1 and X_2 are different itemsets, and T_1 and T_2 are contrary to each other, or T_1 and T_2 are the same, but there is a big difference in the interestingness values of the two constituent patterns. $X_{1,e}$ and $X_{2,e}$ may show different impacts on business.

An example of a pair pattern in Section III is

$$\begin{aligned} &M \wedge c_3 \rightarrow Y \\ &M \wedge c_2 \rightarrow N. \end{aligned} \tag{13}$$

It shows that a group of male customers $(X_{\mathrm{p}} = M)$ may lead to different business outcomes. Males with c_3 are likely associated with debt occurrences while those males with c_2 likely incur no debt. Informed by such contrast findings, for males, we should encourage them to take c_2 rather than c_3 in order to convert business outcomes from debt to nondebt.

To illustrate the definition of the interestingness of a pair pattern $\mathcal{P}$, let us define the interestingness of a combined association rule pair $(I_{\mathrm{pair}}())$

$$I_{\mathrm{pair}}(\mathcal{P}) = \begin{cases} |Conf(P_1) - Conf(P_2)|, & \text{if } T_1 = T_2 \\ \sqrt{Conf(P_1)\, Conf(P_2)}, & \text{if } T_1 \text{ and } T_2 \text{ are contrary} \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where P_1 and P_2 are the two constituent patterns in the pair.

I_{pair} measures the contribution of the two different parts in the antecedents to the occurrence of different classes in a group of customers with the same patterns. The value of I_{pair} falls in [0, 1]. The larger I_{pair} is, the more interesting and actionable a pair of rules is. This kind of knowledge can help to design business campaigns and intervention strategies and to improve the business process.

C. Cluster Pattern

With combined mining, atomic or combined patterns can be further organized into *clusters* by placing similar or related patterns together. Such patterns can be more informative than their constituent patterns. A *cluster pattern* is defined as follows.

Definition 5 (Cluster Pattern): If there are k atomic patterns $X_i \rightarrow T_i, (i = 1, \ldots, k)$, $k \geq 3$, and $X_1 \cap X_2 \cap \cdots \cap X_k = X_{\mathrm{p}}$, a *cluster pattern* $(\mathcal{P})$ is in the form of

$$\mathcal{P} : \begin{cases} X_1 \rightarrow T_1 \\ \cdots \\ X_k \rightarrow T_k \end{cases} \tag{15}$$

where $k > 2$ and X_{p} is the *prefix* of cluster $\mathcal{P}$.

Section VII-A1 shows the examples of cluster patterns.

With regard to the interestingness I_{pair} of a combined pattern cluster, let us illustrate it in terms of association rule clusters. Based on the interestingness of a pair pattern [(14)], for a cluster rule $\mathcal{P}$ with k constituent patterns $P_1, P_2, \ldots, P_k$, its interestingness $(I_{\mathrm{cluster}}())$ is as follows:

$$I_{\mathrm{cluster}}(\mathcal{P}) = \max_{P_i, P_j \in \mathcal{C}, i \neq j} I_{\mathrm{pair}}(P_i, P_j). \tag{16}$$

The aforementioned definition of I_{cluster} indicates that interesting clusters are those rules including interesting rule pairs, and the other rules in the cluster provide additional information. Similar to I_{pair}, the value of I_{cluster} also falls in [0, 1].

D. Incremental Pair Pattern

In some of the pair patterns, there is a certain relationship between items X_1 and X_2. One situation is $X_2 = X_1 \cup X_p, T_1 \neq T_2$, where we then have incremental pair patterns.

Definition 6 (Incremental Pair Pattern): An incremental pair pattern is a special pair of combined patterns as follows:

$$\mathcal{P} : \begin{cases} X_{\mathrm{p}} \rightarrow T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T_2 \end{cases} \tag{17}$$

where $X_{\mathrm{p}} \neq \emptyset$, $X_{\mathrm{e}} \neq \emptyset$, and $X_{\mathrm{p}} \cap X_{\mathrm{e}} = \emptyset$.

The second constituent pattern is an extension of the first, and by appending items X_{e} to it, X_{e} leads to the difference between the outcomes of the constituent patterns. The relationship between X_p and X_e can be unordered or ordered.

In Section VII-A2, we introduce examples of incremental pair patterns identified in social security data associated with government debt. Another example of incremental pair sequences is the *impact-reversed activity patterns* [1]. An impact-reversed activity pattern consists of an underlying activity pattern and a derivative pattern with an incremental activity sequence X_{e}. In the reversal from one pattern's impact (T_1) to the other's (T_2), the extra itemset X_{e} plays an important role. This phenomenon is of great interest to business. For instance, it can be used for improving a business process, recommending to avoid activities or government–customer contacts that may lead to or be associated with debts.

To measure the interestingness of incremental pair patterns, we define the *conditional Piatetsky–Shapiro's ratio* Cps as follows.

Definition 7 (Conditional Piatetsky–Shapiro's (P–S) Ratio): Cps measures the difference led by the occurrence of X_{e} in an incremental pair pattern, which is defined as follows:

$$\begin{aligned} Cps(X_{\mathrm{e}} \rightarrow T | X_{\mathrm{p}}) &= Prob(X_{\mathrm{e}} \rightarrow T | X_{\mathrm{p}}) - Prob(X_{\mathrm{e}} | X_{\mathrm{p}}) \\ &\quad \times Prob(T | X_{\mathrm{p}}) \\ &= \frac{Prob(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Prob(X_{\mathrm{p}})} - \frac{Prob(X_{\mathrm{p}} \wedge X_{\mathrm{e}})}{Prob(X_{\mathrm{p}})} \\ &\quad \times \frac{Prob(X_{\mathrm{p}} \rightarrow T)}{Prob(X_{\mathrm{p}})}. \end{aligned}$$

Cps measures the statistical or proportional significance of incremental sequence X_{e} leading to the impact reversal from T_1 to T_2.

E. Incremental Cluster Pattern

Similar to incremental pair patterns, for cluster patterns, we have *incremental cluster patterns*. We illustrate here the incremental cluster sequences.

Definition 8 (Incremental Cluster Sequences): An incremental cluster sequence is a special cluster of combined patterns with additional items appending to every previously adjacent constituent patterns. An example is

$$\mathcal{P} : \begin{cases} X_{\mathrm{p}} \rightarrow T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \rightarrow T_2 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \rightarrow T_3 \\ \cdots \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \wedge \cdots \wedge X_{\mathrm{e},k-1} \rightarrow T_k \end{cases} \tag{18}$$

where $\forall i$, $1 \leq i \leq k-1$, $X_{i+1} \cap X_i = X_i$, and $X_{i+1} \setminus X_i = X_{e,i} \neq \emptyset$, i.e., X_{i+1} is an *increment* of X_i. The aforementioned cluster of rules shows the impact of the pattern increment on their outcomes.

In Section VII-A3, we illustrate incremental cluster patterns in social security data associated with government debt.

Note that, in extracting frequent pattern-based incremental cluster patterns, it is not necessary for all constituent patterns to have high interestingness values. For instance, combined association rule clusters do not need a high $confidences$. In fact, a pattern with a low confidence is also useful because it helps to judge the extent of the negative impact of the incremental part on the pattern and the extent of the positive impact of the incremental part on the next pattern.

A new metric, *impact*, is designed as follows to measure the interestingness of incremental cluster sequences.

Definition 9 (Impact): The *impact* of X_{e} on the outcome in a cluster pattern is

$$impact_{\mathrm{e}}(P) = \begin{cases} cont_{\mathrm{e}}(P) - 1 & : \text{if } cont_{\mathrm{e}}(P) \geq 1 \\ \frac{1}{cont_{\mathrm{e}}(P)} - 1 & : \text{otherwise.} \end{cases} \quad (19)$$

Impact measures how much the incremental items change the outcomes, and its values fall in $[0, +\infty)$. To select interesting incremental cluster sequences, one may want to set a threshold for the minimum or the average *impact* in a cluster.

F. Procedure for Generating MFCPs

Based on the different expectations on combined pattern types, MFCPs may be instantiated into pairs, clusters, incremental pairs, and incremental clusters. Correspondingly, the discovery of such types of patterns can be segmented into six steps on demand. The process is as follows. First, atomic patterns P_1 are discovered in one data set and then are used to partition another data set. Then, in a derived subdata set, atomic patterns P_2 are discovered. After that, P_1 and P_2 are merged into a combined pattern. Through finding common prefixes or postfixes in these patterns, interesting pair patterns are discovered by putting contrast patterns together. In addition, patterns with the same prefixes or postfixes form cluster patterns. Finally, incremental pair and cluster patterns can be built upon the identified pairs/clusters, respectively.

METHOD: Mining MFCPs
INPUT: target data sets $\mathcal{D}_k$ $(k = 1, \ldots, K)$, business problem Ψ
OUTPUT: combined patterns $\mathcal{P}$
Step 1. Mining atomic patterns: For each data set or partitioned data set $\mathcal{D}_k$, mine for interesting atomic patterns $\mathcal{P}_k$ on the data set.
Step 2. Combining atomic patterns: Merge relevant atomic patterns identified in the aforementioned step as per pattern merging method $\mathcal{G}_k$.
Step 3. Generating pair patterns: Generate pair patterns from the resulting combined patterns. For instance, those patterns with common prefixes but contrary outcomes (or the same outcomes but having a big difference in interestingness) form pair patterns.
Step 4. Generating cluster patterns: For each pair pattern, add other related patterns to it to form a cluster pattern.
Step 5. Generating incremental pair patterns: For those pair patterns, if one pattern is an extension of the other, then output it as an incremental pair pattern.
Step 6. Generating incremental cluster patterns: In a cluster pattern, if there is an ordinal relation between the relevant adjacent patterns and the latter patterns consist of additional information on top of its former ones, output them as incremental cluster patterns.

For instance, as shown in [25], interesting combined rules and rule clusters can be extracted on atomic association rules with interestingness metrics such as $support$, $confidence$, $lift$, $Cont_{\mathrm{e}}$, and I_{rule}. The learned rules with high $support$ and $confidence$ are also organized into clusters, and then, the clusters are ranked by I_{cluster} to find actionable cluster patterns. Section VII-A further illustrates these techniques.

VI. Multimethod Combined Mining

A. Basic Frameworks

Multimethod combined mining is another approach to discover more informative knowledge in complex data. The focus of multimethod combined mining is on combining multiple data mining algorithms as needed in order to generate more informative knowledge. In fact, the combination of multiple data mining methods has been recognized as an essential and effective strategy in dealing with complex applications.

Definition 10 (Multimethod Combined Mining): Assuming that there are l data mining methods $\mathcal{R}_l$ $(l = 1, \ldots, L)$, their respective interestingness metrics are in the set $\mathcal{I}_m$ $(m = 1, \ldots, M)$. The features available for mining the data set are denoted by $\mathcal{F}$, and *multimethod combined mining* is in the form of

$$\begin{aligned} &\mathcal{P}_l : \mathcal{R}_l(\mathcal{F}) \rightarrow \mathcal{I}_{m,l} \\ &\mathcal{P} := \mathcal{G}_M(\mathcal{P}_l) \end{aligned} \quad (20)$$

where $\mathcal{G}_M$ is the merging method integrating the patterns identified by multiple methods.

In dealing with complex real-world applications, the general process of multimethod combined mining is as follows.

1) First, based on the domain knowledge, business understanding, data analysis, and goal definition, a user determines which methods should be used in the framework.
2) Second, the patterns discovered by each method are combined with the patterns by the other methods in terms of merging method $\mathcal{G}$. In reality, the merger could be through either *serial* or *parallel* combined mining.
3) Finally, after mining by all methods, the combined patterns are further reshaped into more workable patterns.

In the following sections, we introduce three general frameworks of multimethod combined mining. They are *parallel multimethod combined mining*, *serial multimethod combined mining*, and *closed-loop multimethod combined mining*.

B. Parallel Multimethod Combined Mining

One approach to involving multiple methods for combined mining is the parallel multimethod combined mining.

Definition 11 (Parallel Multimethod Combined Mining): Given K data sets $\mathcal{D}_k\,(k = 1, \ldots, K)$, L data mining methods $\mathcal{R}_l\,(l = 1, \ldots, L)$ are used to mine them, respectively, and the *parallel multimethod combined mining* is a process as follows.

1) Parallel data mining is conducted on each data set using different data mining methods to find respective atomic pattern sets.

$$\begin{cases} \mathcal{D}_1 \overset{e,\mathcal{I}_1,\mathcal{R}_1,\Omega_m}{\longrightarrow} \mathcal{P}_1 \\ \mathcal{D}_2 \overset{e,\mathcal{I}_2,\mathcal{R}_2,\Omega_m}{\longrightarrow} \mathcal{P}_2 \\ \cdots \\ \mathcal{D}_K \overset{e,\mathcal{I}_l,\mathcal{R}_l,\Omega_m}{\longrightarrow} \mathcal{P}_n \end{cases} \tag{21}$$

2) The atomic patterns identified by individual methods are merged into combined patterns by a merging method $\mathcal{G}$

$$\mathcal{P} := \mathcal{G}(\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n). \tag{22}$$

In parallel multimethod combined mining, multiple methods are implemented on multiple data sources or partitioned data sets. The resulting patterns are the combination of the outputs of individual methods on particular data sources.

An example of parallel multimethod combined mining is to mine for demographic patterns on customer demographic data using association rule mining and, at the same time, to discover event classes on transactional data sets by a decision tree. The identified results from the association rule mining and decision-tree-based classification are then merged to form combined patterns: *frequent demographic pattern–event class*. The combined patterns may show that customers with certain frequent demographic characteristics are likely to be further associated with the occurrences of particular types of events.

C. Serial Multimethod Combined Mining

The second type of approach to involving multiple methods into combined mining is the *serial multimethod combined mining*, which is described as follows.

Definition 12 (Serial Multimethod Combined Mining): Supposing that we have L methods $\mathcal{R}_l\,(l = 1, \ldots, L)$, the *serial multimethod combined mining* is a gradual process as follows.

1) Based on the understanding of domain knowledge, data, business environment, and metaknowledge, select a suitable method (for example, $\mathcal{R}_1$) on the data set $\mathcal{D}$. Consequently, we obtain the resulting pattern set $\mathcal{P}_1$

$$\mathcal{D} \overset{e,\mathcal{R}_l,\mathcal{F}_l,\mathcal{I}_l,\Omega_m}{\longrightarrow} \mathcal{P}_1,\ or \tag{23}$$

$$\{\mathcal{R}_1, \mathcal{F}_1, \mathcal{I}_1\} \overset{e,\mathcal{D},\Omega_m}{\longrightarrow} \mathcal{P}_1. \tag{24}$$

2) Supervised by the resulting patterns $\mathcal{P}_1$ and deeper understanding of the business and data during mining P_1, select the second appropriate data mining methods $\mathcal{R}_2$ to mine $\mathcal{D}$ for pattern set P_2

$$\{\mathcal{R}_2, \mathcal{F}_2, \mathcal{I}_2\} \overset{e,\mathcal{D},\Omega_m,\mathcal{P}_1}{\longrightarrow} \mathcal{P}_2. \tag{25}$$

$\mathcal{P}_1$ involves and contributes to the discovery of $\mathcal{P}_2$.

3) Similarly, select the next method to mine the data with the supervision of the corresponding patterns from the previous stages; repeat this process until the data mining objective is met, and we get eventual pattern set P

$$\{\mathcal{R}_L, \mathcal{F}_L, \mathcal{I}_L\} \rightarrow \mathcal{P}. \tag{26}$$

In serial multimethod combined mining, the data mining methods are used one by one according to specific arrangements. That is, a method is selected and used based on the output of the previous methods. Such serial combination of data mining methods is often very useful for mining complex data sets.

An example is associative classification through the gradual deployment of association rule mining and classification [13]. In other cases, association mining can be deployed on top of the results of clustering for rarity mining [16] or vice versa to discover more interesting patterns as in [8], [22], and [24]. More examples include the combination of sequential pattern mining and classification [9], classification and clustering [21], and regression and association rule mining [14].

D. Closed-Loop Multimethod Combined Mining

Most of the current combinations of multiple data mining methods are either in parallel or serial. In these two approaches, we generally do not bother about the impact of one method on the other. For example, in serial multimethod combined mining, a previously applied method R_j, in general, has no impact on another method's (R_i) resulting patterns and performance, even though R_j follows R_i. This is actually a common issue in open-loop combination.

In practice, the feedback from a latter method's results to its previous methods may assist with the pattern refinement in combination and enhance the deliverable performance and the efficiency of the data mining process. To this end, we propose the concept of *closed-loop multimethod combined mining*. Its general idea is as follows.

Definition 13 (Closed-Loop Multimethod Combined Mining): Supposing that we have data set $\mathcal{D}$, L data mining methods $\mathcal{R}_l\,(l = 1, \ldots, L)$ are used to mine $\mathcal{D}$, and if multiple data mining methods are serially applied, we then conduct *closed-loop multimethod combined mining* through multiple loops of pattern discovery processes as follows.

1) Loop 1: Follow the process of serial multimethod combined mining to generate pattern set $\mathcal{P}$ through a progressive pattern formation process as shown in (27). During each step of extracting pattern set $\mathcal{P}^1$, there are some samples that cannot be properly identified. This may not indicate that such samples are not identifiable but, rather, that it is due to the constraints and conditions applied on the respective methods.

$$\{\mathcal{R}_1, \mathcal{F}_1, \mathcal{I}_1\} \rightarrow \ldots \{\mathcal{R}_2, \mathcal{F}_2, \mathcal{I}_2\} \rightarrow \{\mathcal{R}_L, \mathcal{F}_L, \mathcal{I}_L\} \rightarrow \mathcal{P} \tag{27}$$

2) Loop 2: The patterns identified by data mining methods $\mathcal{R}_l(l = 1, \ldots, L)$ are further checked to see whether the identified patterns are valid to all samples in the data set $\mathcal{D}$. Those samples on which the patterns are not valid form a data set $\mathcal{D}^1$ from the data set $\mathcal{D}$. They are called *exceptional itemsets*. The exceptional itemsets are further fed back to another loop of mining by reusing the methods from $\mathcal{R}_1$ through $\mathcal{R}_L$ as needed, with the

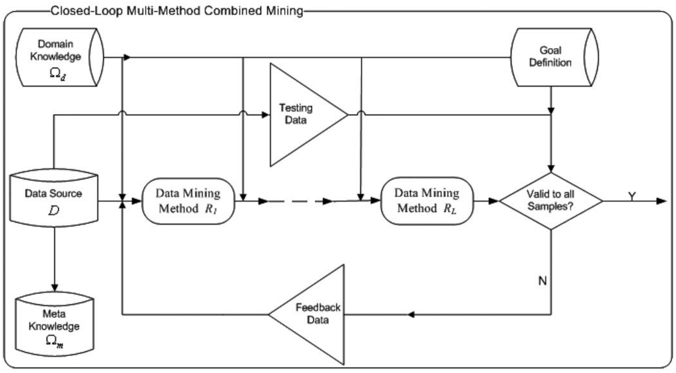


Fig. 2. Closed-loop multimethod combined mining.

refinement of parameters, etc. We then get another resulting set $\mathcal{P}^2$.

3) Repeat the process of loop 2 as needed. Supposing that Z loops are needed, in order that the final remaining exceptional itemsets $\mathcal{D}^Z$ that cannot be covered by patterns are within an acceptable level, we correspondingly obtain Z pattern sets in the whole process, namely, $\{\mathcal{P}^1, \ldots, \mathcal{P}^Z\}$.
4) Merge the identified Z pattern sets to generate final combined patterns

$$\mathcal{P} := \mathcal{G}_C(\mathcal{P}^1, \mathcal{P}^2, \ldots, \mathcal{P}^Z) \tag{28}$$

where $\mathcal{G}_C$ represents the merging methods for closed-loop multimethod combined mining.

Fig. 2 further illustrates the process of the closed-loop multimethod combined mining. In the closed-loop combination, whether a pattern is interesting or not does not only depend on a particular method that extracts the pattern but also depend on the other methods used in the system. Hence, the performance and efficiency of the system could be much improved by using the same interestingness measures.

E. Closed-Loop Sequence Classification

In recent years, sequence classification has been recognized as a challenging data mining issue. It has a wide range of applications such as bioinformatics and customer behavior predictions. Most of existing sequence classification algorithms follow the theory of serial multimethod combined mining in which the classification follows sequence pattern mining. It is known that efficiency is a key problem in sequential pattern mining even though many algorithms have been proposed to improve the efficiency. In sequential pattern mining, the time order has to be taken into account to find frequent subsequences. Hence, a huge number of candidates have to be checked in the algorithm. The sequential pattern mining may take weeks or even months if all the candidates are generated and processed. On the other hand, in order to build sequential classifiers, a number of processes, such as the *significance test* and the *coverage test*, have to be conducted on the sequential pattern set. If the sequential pattern set contains huge amounts of sequential patterns, the classifier building can also be extremely time consuming. Therefore, in sequence classification, the efficiency problem exists not only in sequential pattern mining but also in classifier building.

In fact, in rule-based classification, the most important task is not to find the complete rule set but the most discriminative rules. In [3], experimental results show that "redundant and nondiscriminating patterns often overfit the model and deteriorate the classification accuracy." To solve such issues, we propose a novel closed-loop sequence classification method as follows. First, a small set of the most discriminating sequential patterns is mined. These patterns are then used for the coverage test on the training data set. If the sequential pattern set is small enough, there must be some samples that have not been covered by the mined patterns. These uncovered samples are further fed back to the next loop of sequential pattern mining. Again, a coverage test is implemented on the newly mined patterns. The remaining samples that cannot be covered are fed back for sequential pattern mining until the predefined thresholds are reached or all samples are covered.

TABLE X
2×2 Feature–Class Contingency Table

	x	$\neg x$	$\sum rows$
T	a	b	$a+b$
$\bar{T}$	c	d	$c+d$
$\sum cols$	$a+c$	$b+d$	$n=a+b+c+d$

Section VII-B1 introduces an example of the *closed-loop sequence classification*.

1) Discriminating Measures: In order to discover a small set of discriminating patterns in each loop, we use the chi-square test and the class correlation ratio (CCR) [18] as the principal interestingness measures. The CCR can be defined given a contingency table shown in Table X.

CCR, defined as follows, measures how a correlated sequence X is with impact T compared to nonimpact $\bar{T}$:

$$CCR(X \rightarrow T) = \frac{\hat{corr}(X \rightarrow T)}{\hat{corr}(X \rightarrow \bar{T})} = \frac{a \cdot (c+d)}{c \cdot (a+b)}. \tag{29}$$

Here, $\hat{corr}$ is the correlation between X and T

$$\hat{corr}(X \rightarrow T) = \frac{\sup(X \cup T)}{\sup(X) \cdot \sup(T)} = \frac{a \cdot n}{(a+c) \cdot (a+b)}. \tag{30}$$

CCR measures to what extent the antecedent is correlated with the class (e.g., impact T or nonimpact $\bar{T}$) that it predicts. CCR falls in $[0, +\infty)$. $CCR = 1$ means that the antecedent is independent of the class. $CCR < 1$ means that the antecedent is negatively correlated with the class. $CCR > 1$ means that the antecedent is positively correlated with the class.

2) Algorithm Outline: The closed-loop sequence classification algorithm is outlined as follows. Since an aggressive pattern mining strategy is used in the closed-loop sequence classification algorithm, only a very small set of sequential patterns, rather than the complete sequential patterns, can be mined in each loop. The number of sequential patterns increases in each loop so that the total sequential pattern number is much less than the complete sequential pattern set.

Algorithm: Mining Closed-Loop Sequential Classifiers
INPUT: transactional data
OUTPUT: sequential classifiers
Step 1. Calculate the *frequency* of each one-event sequence and the corresponding CCR. Only the events with $CCR > 1 + m_1$ or $CCR < 1 - m_2$ (m_1 and m_2 are the margins) are extracted

into a sequential pattern set. The pattern growth is also based on this sequential pattern set. With this greedy strategy, only a small set of sequential patterns is mined.

Step 2. Calculate the *frequency*, *chi-square* value (which measures how significantly a rule is positively correlated to a class), and CCR of each sequence, and only those where the sequences meet the $support$, $significance$, and CCR criteria are output into the resulting sequential pattern set.

Step 3. After all the sequential patterns are extracted in the aforementioned steps, pattern pruning is implemented on the mined sequences. We follow the pattern pruning algorithm in [11]. The only difference is that, in our algorithm, CCR instead of $confidence$ is used as the measure for pruning.

Step 4. Conduct the coverage test following the ideas in [11] and [13]. Since the greedy pattern mining strategy is used in this algorithm, a large number of training samples cannot be covered by the mined sequential patterns.

Step 5. These training samples are fed back to Step 1. With updated parameters, sequential patterns are mined again. After the pattern pruning and coverage test, those uncovered training samples are fed back to Step 1 for further sequential pattern mining by updating parameters. The process iterates until the predefined thresholds are reached or all samples are covered.

We use the following two strategies to build sequence classifiers.

1) Highest weighted score (CCR_{Highest}). Given a sequence instance s, the class label corresponding to the classifiable sequential pattern with the highest weighted score is assigned to s.
2) Multiple weighted scores (CCR_{Multi}). Given one sequence instance s, all the classifiable sequential patterns on a one-level-covered s are extracted. It is not difficult to compute the sum of the weighted scores corresponding to each target class. The class label corresponding to the largest weighted score sum is assigned to s.

VII. Case Study: Mining Combined Patterns in Government Service Data

In this section, we briefly introduce several real-life case studies by instantiating some of the proposed frameworks in mining combined patterns in our relevant projects [31] of public service data mining. Since this paper focuses only on formalizing the frameworks and due to the space limit for fitting in more details of the case studies, we only explain the instantiation and illustrate some results in mining combined patterns in Australian Commonwealth Government Agency Centrelink [29] social security data for government customer debt prevention. Readers who are interested in more case studies can refer to our related works for more details.

A. Mining MFCPs

This section illustrates some case studies of identifying single combined patterns, cluster patterns, and incremental pair and cluster patterns, which combine demographics, arrangement, and repayment activities. More relevant information can be found in [1], [22], [24], and [25].

1) Mining Single Combined Patterns and Cluster Patterns: In [22] and [25], we report details and results in identifying combined association rules, combined rule pairs, and combined rule clusters, which demonstrate the instantiation of the pair and cluster patterns discussed in Section V-A–C. Zhao *et al.* [25] show selected combined rules. Compared with traditional association rules, the combined association rules are much more informative and can improve policy or design campaigns to recover government debts.

For instance, pattern r_9 in the pattern cluster $\mathcal{R}_2$ in Table VIII is as follows:

$$\{\mathcal{R}_2 = \{r_9, r_{10}, r_{11}, r_{12}, r_{13}, r_{14}\} \tag{31}$$

$$r_i = \{U_i(\text{demographic}) + V_i(\text{arrangement}) + V_i(\text{repayment}) \rightarrow T_i\} \tag{32}$$

$$U_9(\text{demographic}) = U_{10} = U_{11} = U_{12} = U_{13} = U_{14} = \text{marital : single, gender : F, benefit : N} \tag{33}$$

$$V_9(\text{arrangement}) = \text{irregular} \tag{34}$$

$$V_9(\text{repayment}) = \text{cash or post} \tag{35}$$

$$T_9 = A \tag{36}$$

$$\ldots\}. \tag{37}$$

It shows that, for single women on benefit "N," the best way to get their debts repayed as quickly as possible is through repayment methods "cash or post" with "irregular" or "withholding" arrangements (r_9 or r_{10}). An actionable policy is to push them to pay through these arrangements, instead of those given in patterns $r_{11}r_{14}$ so as to shift their debt risk (T) from higher level B or C down to lower A.

2) Mining Incremental Pair Patterns: Cao *et al.* [1] report details and results identifying sequential impact-contrasted patterns and sequential impact-reversed patterns in imbalanced data. They illustrate the instantiation of incremental pair patterns in Section V-B. Cao *et al.* [1] shows examples of impact-reversed sequential activity patterns of Centrelink customers. For example, the first row shows an incremental pair pattern

$$\begin{cases} a_{14} \rightarrow \bar{T} \\ a_{14}, a_4 \rightarrow T \end{cases}. \tag{38}$$

The local supports of $a_{14} \rightarrow T$ and $a_{14} \rightarrow \bar{T}$ are 0.903 and 0.684, respectively, so the ratio of the two values is $0.903/0.684 = 1.3$. The local supports of a_{14}, $a_4 \rightarrow T$ and a_{14}, $a_4 \rightarrow \bar{T}$ are 0.428 and 0.119, respectively, so the ratio of the two values is $0.428/0.119 = 3.6$. The aforementioned two ratios indicate that the appearance of a_4 tends to change the impact from $\bar{T}$ (no debt) to T (debt) when a_{14} happens first. These kinds of pair patterns help show what effect an additional activity may have on the impact of the patterns.

3) Mining Incremental Cluster Patterns: A case study of incremental cluster patterns is given in [26], which shows a set of incremental cluster patterns in customer activity data. A dynamic chart, as shown in Fig. 3, is designed to show the dynamics of a cluster of patterns' interestingness and pattern evolution. An example of discovered incremental cluster patterns is

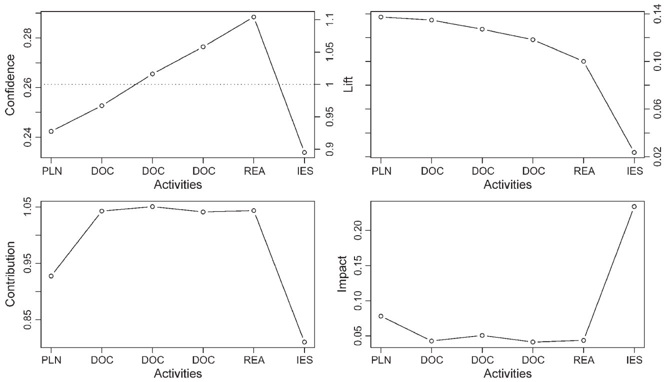


Fig. 3. Dynamic charts showing the dynamics of incremental cluster patterns.

(T stands for a debt, and the codes PLN, DOC, REA, and IES represent a series of activities)

$$\begin{cases} PLN \rightarrow T \\ PLN, DOC \rightarrow T \\ PLN, DOC, DOC \rightarrow T \\ PLN, DOC, DOC, DOC \rightarrow T \\ PLN, DOC, DOC, DOC, REA \rightarrow T \\ PLN, DOC, DOC, DOC, REA, IES \rightarrow T. \end{cases} \quad (39)$$

The upper-left chart shows that the pattern "PLN, DOC, DOC, DOC, REA" is associated with debt occurrence and the lift is 1.1. However, an additional IES following the sequence will dramatically reduce the likelihood of debt occurrence. The bottom-left chart shows that PLN and IES are negatively associated with debt occurrence and DOC is slightly positively associated with debt. The bottom-right chart suggests that IES has a big impact on the outcome, PLN also has some impact on debt, and the impact of DOC is relatively minor.

B. Mining Multimethod-Based Combined Patterns

1) Mining Closed-Loop Sequential Classifiers: Zhang *et al.* [23] illustrate the proposed closed-loop sequence classification method discussed in Section VI-E on Centrelink customer activity data. It proposes a novel hierarchical algorithm to build sequential classifiers using discriminative sequential patterns. We build a three-level hierarchical algorithm to predict and further prevent debt occurrence based on the customer transactional activity data. Four subclassifiers CCR_{CMAR}, CCR_{Highest}, CCR_{Multi}, and CCR_{SPAM} have been built to predict the debt occurrence of an identified frequent activity sequence. As shown in [23], our algorithms CCR_{Highest} and CCR_{Multi} outperform two improved algorithms CCR_{CMAR} and CCR_{SPAM} in terms of both the efficiency and the accuracy.

2) Sequential Classification Using Both Positive and Negative Sequences: Zhao *et al.* [27] further conduct sequence classification based on identified frequent positive and negative sequences. We analyze the relationship between the transactional activity patterns and the debt occurrences and build sequence classifiers for debt detection. Negative sequential rules have been used to find both the positive and negative sequences in the Centrelink customer debt-related activity data. Classifiers are built on one type of positive sequences and three types of negative sequences to predict their correlation to debt occurrence.

The results in [27] show that, if built with the same number of rules, in terms of recall, our classifiers built with both the positive and negative rules outperform the traditional classifiers with only positive rules under most conditions. It means that, with negative rules involved, our classifiers can predict more debt occurrences. Another finding is that, compared with that of positive patterns, we only need a small number of negative patterns in building good classifiers. For instance, the following positive and negative sequential activity patterns are identified to build classifiers for debt prediction

$$\{\neg(STM, REA, DOC) \rightarrow \neg T(CCR = 1.86, conf = 0.84) \quad (40)$$

$$REA, CCO, EAD \rightarrow T(CCR = 17.8)\}. \quad (41)$$

VIII. Conclusion and Future Work

Typical enterprise applications, such as telecom fraud detection and cross-market surveillance in stock markets, often involve multiple distributed and heterogeneous features as well as data sources with large quantities and expect to cater for user demographics, preferences, behavior, business appearance, service usage, and business impact. There is an increasing need to mine for patterns consisting of multiple aspects of the aforementioned information so as to reflect comprehensive business scenarios and present patterns that can inform decision-making actions. This challenges existing data mining methods such as postanalysis and table joining based analysis.

Building on existing works, this paper has presented a comprehensive and general approach named *combined mining* for discovering informative knowledge in complex data. We focus on discussing the frameworks for handling multifeature-, multisource-, and multimethod-related issues. We have addressed challenging problems in combined mining and summarized and proposed effective pattern merging and interaction paradigms, combined pattern types, such as pair patterns and cluster patterns, interestingness measures, and an effective tool—dynamic chart for presenting complex patterns in a business-friendly manner.

The frameworks are extracted from our relevant business projects conducted and currently under investigation from the domains of government service, banking, insurance, and capital markets. Several real-life cases studies have been briefed which instantiate some of the proposed frameworks in identifying combined patterns in multiple sources of governmental service data. They have shown that the proposed frameworks are flexible and customizable for handling a large amount of complex data involving multiple features, sources, and methods as needed, for which data sampling and table joining may not be acceptable. They have also shown that the identified combined patterns are more informative and actionable than any single patterns identified in the traditional way.

We are further developing effective paradigms, combined pattern types, combined mining methods, pattern merging methods, and interestingness measures for handling large and multiple sources of data available in our industry projects for government, stock market, insurance, and banking.

REFERENCES

[1] L. Cao, Y. Zhao, and C. Zhang, "Mining impact-targeted activity patterns in imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 8, pp. 1053–1066, Aug. 2008.

[2] L. Cao, Y. Zhao, H. Zhang, D. Luo, and C. Zhang, "Flexible frameworks for actionable knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 9, pp. 1299–1312, Sep. 2010.

[3] H. Cheng, X. Yan, J. Han, and C.-W. Hsu, "Discriminative frequent pattern analysis for effective classification," in *Proc. ICDE*, 2007, pp. 716–725.

[4] S. Dzeroski, "Multirelational data mining: An introduction," *ACM SIGKDD Explor. Newslett.*, vol. 5, no. 1, pp. 1–16, Jul. 2003.

[5] G. Dong and J. Li, "Efficient mining of emerging patterns: Discovering trends and differences," in *Proc. KDD*, 1999, pp. 43–52.

[6] W. Fan, K. Zhang, J. Gao, X. Yan, J. Han, P. Yu, and O. Verscheure, "Direct mining of discriminative and essential graphical and itemset features via model-based search tree," in *Proc. KDD*, 2008, pp. 230–238.

[7] K. K. R. Hewawasam, K. Premaratne, and M.-L. Shyu, "Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 6, pp. 1446–1459, Dec. 2007.

[8] A. Jorge, "Hierarchical clustering for thematic browsing and summarization of large sets of association rules," in *Proc. SDM*, 2004, pp. 178–187.

[9] N. Lesh, M. J. Zaki, and M. Ogihara, "Mining features for sequence classification," in *Proc. KDD*, 1999, pp. 342–346.

[10] B. Lent, A. N. Swami, and J. Widom, "Clustering association rules," in *Proc. ICDE*, 1997, pp. 220–231.

[11] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules," in *Proc. ICDM*, 2001, pp. 369–376.

[12] B. Liu, W. Hsu, and Y. Ma, "Pruning and summarizing the discovered associations," in *Proc. KDD*, 1999, pp. 125–134.

[13] B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," in *Proc. 4th Int. Conf. Knowl. Discov. Data Mining (KDD)*, 1998, pp. 80–86.

[14] A. Ozgur, P.-N. Tan, and V. Kumar, "RBA: An integrated framework for regression based on association rules," in *Proc. SDM*, 2004, pp. 210–221.

[15] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu, "Mining sequential patterns by pattern-growth: The PrefixSpan approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1424–1440, Nov. 2004.

[16] M. Plasse, N. Niang, G. Saporta, A. Villeminot, and L. Leblond, "Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 596–613, Sep. 2007.

[17] Y. Sun, Y. Wang, and A. Wong, "Boosting an associative classifier," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 7, pp. 988–992, Jul. 2006.

[18] F. Verhein and S. Chawla, "Using significant, positively associated and relatively class correlated rules for associative classification of imbalanced datasets," in *Proc. ICDM*, 2007, pp. 679–684.

[19] J. Wang and G. Karypis, "HARMONY: Efficiently mining the best rules for classification," in *Proc. SDM*, 2005, pp. 205–216.

[20] X. Yin, J. Han, J. Yang, and P. S. Yu, "Efficient classification across multiple database relations: A CrossMine approach," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 6, pp. 770–783, Jun. 2006.

[21] H. Yu, J. Yang, and J. Han, "Classifying large data sets using SVM with hierarchical clusters," in *Proc. KDD*, 2003, pp. 306–315.

[22] H. Zhang, Y. Zhao, L. Cao, and C. Zhang, "Combined association rule mining," in *Proc. PAKDD*, 2008, pp. 1069–1074.

[23] H. Zhang, Y. Zhao, L. Cao, C. Zhang, and H. Bohlscheid, "Customer activity sequence classification for debt prevention in social security," *J. Comput. Sci. Technol.*, vol. 24, no. 6, pp. 1000–1009, Nov. 2009.

[24] Y. Zhao, H. Zhang, F. Figueiredo, L. Cao, and C. Zhang, "Mining for combined association rules on multiple datasets," in *Proc. DDDM*, 2007, pp. 18–23.

[25] Y. Zhao, H. Zhang, L. Cao, C. Zhang, and H. Bohlscheid, "Combined pattern mining: From learned rules to actionable knowledge," in *Proc. AI*, 2008, pp. 393–403.

[26] Y. Zhao, H. Zhang, L. Cao, C. Zhang, and H. Bohlscheid, "Efficient mining of event-oriented negative sequential rules," in *Proc. WI*, 2008, pp. 336–342.

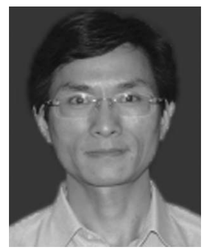
[27] Y. Zhao, H. Zhang, S. Wu, J. Pei, L. Cao, C. Zhang, and H. Bohlscheid, "Debt detection in social security by sequence classification using both positive and negative patterns," in *Proc. ECML-PKDD*, vol. 5782, *LNAI*, 2009, pp. 648–663.

[28] Y. Zhao, C. Zhang, and L. Cao, Eds., *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*. Hershey, PA: Inf. Sci. Ref., 2009.

[29] Centrelink. [Online]. Available: www.centrelink.gov.au

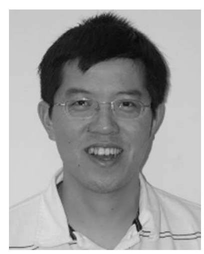
[30] Top 10 Best Practices for Building a Large Scale Relational Data Warehouse. [Online]. Available: http://sqlcat.com/top10lists/archive/2008/02/06/top-10-best-practices-for-building-a-large-scale-relational-data-warehouse.aspx

[31] [Online]. Available: http://datamining.it.uts.edu.au/group/projects.php



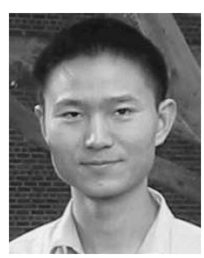
Longbing Cao (M'05–SM'06) received the Ph.D. degree in intelligent sciences from the Chinese Academy of Sciences, Beijing, China, in 2002, and the Ph.D. degree in computing sciences from the University of Technology, Sydney (UTS), Sydney, Australia, in 2005.

He is currently a Professor with the UTS. He is also currently a Data Mining Research Leader with the Australian Capital Markets Cooperative Research Centre, Sydney, Australia. His research interests include data mining and machine learning and their applications, behavior informatics, multiagent technology, and agent mining.



Huaifeng Zhang (M'06) received the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2004.

He was a Postdoctoral Research Fellow and a Lecturer with the University of Technology Sydney, Sydney, Australia. He is currently a Senior Data Mining Specialist with Centrelink, Canberra, Australia. He has published one book and more than 40 refereed papers. His research interests include combined pattern mining, sequence analysis, and behavior analysis.



Yanchang Zhao (M'01) received the Ph.D. degree in computing science from the University of Technology Sydney, Sydney, Australia, in 2007.

From 2007 to 2009, he was an Australian Postdoctoral Fellow (Industry) with the University of Technology Sydney. He is currently a Senior Data Mining Specialist with Centrelink, Canberra, Australia. His research interests include clustering, association rules, time series, outlier detection, and data mining applications. He has over 40 papers published in journals and conference proceedings.

Dan Luo received the Ph.D. degree in computing sciences from the University of Technology, Sydney, Sydney, Australia, in 2009.

She is currently a Research Fellow with the University of Technology Sydney, Sydney, Australia. Her research interests include data mining and its applications and business intelligence.

Chengqi Zhang (M'90–SM'95) received the Ph.D. degree in computer science from The University of Queensland, Brisbane, Australia, in 1991, and the Dr.Sc. degree from Deakin University, Geelong, Australia, in 2002.

He is currently with the University of Technology, Sydney (UTS), Sydney, Australia, where he is a Research Professor of information technology and the Director of the UTS Priority Investment Research Centre for Quantum Computation and Intelligent Systems. He has published more than 200 refereed research papers. His main research interests include data mining and its applications.

Prof. Zhang is a Fellow of the Australian Computer Society.