Pattern Recognition Letters 000 (2018) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Review on mining data from multiple data sources

Ruili Wang[a], Wanting Ji[a,*], Mingzhe Liu[b], Xun Wang[c], Jian Weng[d], Song Deng[e], Suying Gao[f], Chang-an Yuan[g]

[a] Institute of Natural and Mathematical Sciences, Massey University, Room 2.14, Mathematical Sciences Building, Auckland 0632, New Zealand
[b] School of Network Security, Chengdu University of Technology, Chengdu 610059, China
[c] School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China
[d] College of Cyber Security, Jian University, Guangzhou 519632, China
[e] School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210043, China
[f] School of Economics and Management, Hebei University of Technology, Tianjin 300401, China
[g] School of Computer and Information Engineering, Guangxi Teachers Education University, Nanning 530023, China

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Multiple data source mining
Pattern analysis
Data classification
Data clustering
Data fusion

A B S T R A C T

In this paper, we review recent progresses in the area of mining data from multiple data sources. The advancement of information communication technology has generated a large amount of data from different sources, which may be stored in different geological locations. Mining data from multiple data sources to extract useful information is considered to be a very challenging task in the field of data mining, especially in the current big data era. The methods of mining multiple data sources can be divided mainly into four groups: (i) pattern analysis, (ii) multiple data source classification, (iii) multiple data source clustering, and (iv) multiple data source fusion. The main purpose of this review is to systematically explore the ideas behind current multiple data source mining methods and to consolidate recent research results in this field.

1. Introduction

The advancement of information communication technology has generated a large amount of data from different sources, which may be stored in different geological locations. Each database may have its own structure to store data. Mining multiple data sources [1–3] distributed at different geological locations to discover useful patterns are critical important for decision making. In particular, the Internet can be seen as a large, distributed data repository consisting of a variety of data sources and formats, which can provide abundant information and knowledge.

Data from different sources may seem irrelevant to each other. Once information generated from different sources is integrated, new and useful knowledge may emerge. Here is an excellent example of how an organization to utilize mining data from different data sources to obtain profound information, which cannot obtain from an individual source.

The Australian Taxation Office (ATO) mines data from different data sources such as social media posts, private school records and immigration data to detect tax cheats. Mining data from different data sources become a sophisticated tool to crackdown tax cheats that yielded nearly $10 billion in 2016 [4]. For example, in a normal Australian family, the husband has a business and reported $80,000 of taxable income per year, putting him just inside the second-lowest tax bracket, and his wife reported earning $60,000 per year. But the data collected from different data sources revealed that the family had three children at private schools at an estimated cost of $75,000 per year, while immigration records and social media posts showed that the family had recently taken five business-class flights and a holiday in a Canadian ski resort, Whistler. It means their declared incomes did not match their lifestyle. This prompted ATO to contact them to confirm if they have unpaid taxes. From the above example, we can see that developing an effective data mining technique for mining from multiple data sources to discover useful information is crucially important for decision making.

However, how to efficiently mine quality information from multiple data sources is a challenging task for current research [5–9], especially in the current big data era, because in real world applications, data stored in multiple places often have conflictions [10]. The conflictions include: (i) data name conflictions: (a) the same object has different names in different data sources, or (b) two different objects from different data sources may have the same name; (ii) data format conflictions: the same object in different data sources has different data types, domains, scales, and preci-

* Corresponding author.
E-mail address: jwt@escience.cn (W. Ji).

https://doi.org/10.1016/j.patrec.2018.01.013
0167-8655/

sions; (iii) data value confliction: the same object in different data sources records different values; (iv) data sources confliction: different data sources have different database structures.

In order to overcome these conflictions, four effective approaches have been adopted to mine useful information and discover new knowledge from multiple data sources: (i) pattern analysis [11–14], which mining useful patterns and information from one data source or several data sources in accordance with changing conditions constraints or relationships; (ii) multiple data source classification, which labels data sources according to a certain standard, then classifies them by their labels; (iii) multiple data source clustering, which clusters data sources according to their similarities/distances; (iv) multiple data source fusion, which combines data from multiple data sources to achieve higher accuracy and more specific ratiocinations. Based on these three approaches, we can mine useful information from multiple data sources to discover new knowledge according to individual needs.

The main purpose of this review is to systematically overview current multiple source data mining methods and to consolidate recent research results in this field. In this paper, the pattern analysis approaches for multiple data sources are reviewed first, and the approach of multiple data source classification and clustering is reviewed after that. Then, mining multiple data sources using data fusion is reviewed. The rest of this paper is organized as follows. Section 2 shows typical methods of multiple data source mining using pattern analysis, while Section 3 is about multiple data source classification and clustering. Section 4 describes methods of multiple data source fusion. Finally, Section 5 provides conclusions and discussions.

2. Key methods for pattern analysis

In this section, we will describe different methods of pattern analysis on multiple data sources. Pattern analysis on multiple data sources is a process of mining valuable information or extracting hidden predictive information from different databases. Patterns existing in multiple data sources can be categorized as local patterns and global patterns. A pattern identified from a mono-database can be seen as a local pattern, while a global pattern has to be determined based on all the obtained local patterns from multiple data sources [15–17]. Correspondingly, two mainstream methodologies of pattern analysis are developed, named local pattern analysis and global pattern analysis, which will be discussed as follows.

2.1. Local pattern analysis

Local patterns analysis is a process of identifying patterns from a mono-database, which can efficiently extract knowledge from the mono-database [18]. Based on previous studies [19–25], we can summarize the local pattern analysis methods into three categories: association rule mining, sequential pattern mining, and others.

2.1.1. Association rule mining

Association rule mining is a process of discovering the probability of the co-occurrence of items in a database. Association rules express the relationships between co-occurring items, which are often used to analyze sales transactions. Association rule mining is also known as market basket analysis. There are four parameters used to describe an association rule: support, confidence, expected confidence, and lift.

For example, a transaction database of a supermarket records the number of customers on a given day is 1000. Among them, there are 100 customers bought both bread and milk, the support of bread → milk (bread and milk are two itemsets, bread → milk presents an association rule between bread and milk) is 10%. If 40% of customers who bought bread will also buy milk, then the confidence of bread → milk is 40%. If there are totally 200 customers bought milk in the given day, then the expected confidence of bread → milk is 20%. Thus, the lift of bread → milk is 40%/20% = 2.

While the support measures the importance of an association rule, the confidence measures the accuracy of an association rule. The support can reflect the representativeness of an association rule. Although some association rules have high confidences, their supports are very low, which means these association rules have little chance of being practical so that they are also unimportant.

The expected confidence of an association rule can be seen as the support of an itemset without the influence of other itemsets, while the lift of an association rule is the ratio of confidence to expected confidence, which is used to describe how one itemset affects another itemset. In general, the lift of a useful association rule should be greater than 1. Only when the confidence of an association rule is greater than its expected confidence, it shows a correlation between these two itemsets (i.e., one itemset has a promoting effect on another itemset), otherwise, this association rule is meaningless.

According to the definition above, any two itemsets in a transaction can have an association rule between them. In fact, users are only interested in the association rules that satisfy a certain support and confidence. Therefore, a support-confidence framework is used to mine the correlations between itemsets in the local databases. Two thresholds, the minimum support and the minimum confidence, specify the minimum requirements of a meaningful association rule.

Association rule mining consists of two phases: (i) In the first phase, all the high frequent itemsets are identified. (ii) In the second phase, association rules are generated from the high frequent itemsets. It has been extensively studied to identify the relation existing in databases between itemsets [26].

A classical algorithm of association rule mining named Apriori algorithm is presented by Agrawal and Srikant [27], which is used for mining frequent itemsets and relevant association rules. A key concept in the Apriori algorithm is the anti-monotonicity of the support measure. It assumes that: (i) All subsets of a frequent itemset must be frequent. (ii) For any infrequent itemset, all its supersets must be infrequent. The algorithm can be divided into two phases. In the first phase, it applies minimum support to find all the frequent sets with k items in a database. In the second phase, it uses the self-join rule to find the frequent sets of $k+1$ items with the help of frequent k-itemsets. Repeat this process from $k=1$ to the point when we are unable to apply the self-join rule. The Apriori algorithm can find the relationship between items in a database and it is the first efficient and practical algorithm for association rule mining. But it needs to generate a large number of candidate itemsets and repeatedly scan the entire database.

A Frequent-Pattern tree structure called FP-tree is presented by Han et al. in [28], which is constructed by (i) Scan a database to get all the frequent itemsets and their respective support in the database, sorting frequent items in frequent itemsets in a descending order of their support values. (ii) Create the root node and scan the database again. Then select the frequent items and sort them in the order of frequent itemsets. It is used for collecting compacted and important information about frequent patterns. According to FP-tree, a corresponding method called frequent-pattern growth algorithm is proposed for mining the complete set of frequent patterns [28]. The proposed method can efficiently avoid the process of candidate generation-and-test, and it can cut down the time cost of frequent pattern mining. However, the algorithm may have a challenge when encountering graph-based patterns or noise samples.

Different from traditional association rule mining and inspired by the paradigm named Ratio Rule (RR), which is used for acquiring the quantitative association knowledge, Yan et al. [29] proposed a method that uses a robust and adaptive one-pass method to mine ratio rules from a database independently. Then all the obtained ratio rules will be integrated into a probabilistic model. The method is called Robust and Adaptive Ratio Rule (RARR). The main notion of RARR is converting the ratio rule problem into the minimization problem of a non-negative energy function. The greatest advantage of RARR is robustness in dealing with outliers while traditional multiple-source ratio rule mining algorithms cannot solve. Besides, it is adaptive to mine the rules for a dynamic data source.

A least association rule means that the association rule consists of the least item, which is very important and critical when it is used to find the infrequent or abnormal itemsets. Abdullah et al. [30] proposed a method called Critical Relative Support (CRS) to mine the significant least association rules, which satisfies a certain predefined minimum critical threshold value. The experimental dataset is from the examination results of computer science students from University Malaysia Terengganu. CRS has been verified efficiently for discovering the significant least association rules and can substantially decrease the number of unwanted rules. But the performance of CRS has only tested on one educational dataset, which should be tested by a variety of different real datasets.

In general, association rule mining uses the support-confidence framework to find the rules of users' interests. However, the support-confidence framework cannot reflect the semantic measure among the items. Semantic measure is characterized by utility values, which are associated with transaction items typically. The items that cannot be measured directly by a decision maker such as stock, cost or profit, are called utility. For example, the confidence of an association rule represents the percentage that shows how frequently the rule occurs among all rules.

High utility itemsets mining can be seen as an extension of frequent itemset mining [31], which includes three requirements: (i) The item appears more than once. (ii) The item has its own profits. (iii) The mining goal is to find the items which have high profits. High utility itemsets mining will be more valuable than a single support-confidence framework and can permit with more accurate financial analysis and decisions.

A new utility-confidence framework based association rule mining method is proposed by Sahoo et al. in [31], which is also the first research on mining non-redundant association rules extracted from the high utility itemsets. It includes two-step process: (i) find all high utility itemsets and (ii) extract all valid rules. A compressed representation for association rules is generated with the help of high utility closed itemsets (HUCI) and their generators. The compressed representation is named as high utility generic basis (HGB), which contains rules with minimal antecedent and maximal consequent. After that, in order to extract the non-redundant rule set, HUCI and their generator are mined by the proposed algorithm HUCI-Miner that identifies HUCI and associates the high utility generators to the corresponding closed itemset. Moreover, an inference algorithm is proposed to derive all valid association rules under this framework. Experimental results show a significantly better achievement in compactness of the proposed rule mining algorithms.

2.1.2. Sequential pattern mining

As indicated in [32], a sequence is an ordered list of itemsets. For example, in customer transactions, an itemset corresponds to a transaction, and a sequence corresponds to a list of transactions sorted by transaction time. The goal of sequential patterns mining is to identify the maximal sequences satisfied a pre-specified minimum support, and the maximal sequence can be considered as a sequential pattern, which is a very important pattern used for identifying the sequential relationships between different itemsets. There is no doubt that studying sequential patterns is always an active area of data mining [32].

Agrawal and Srikant [32] first introduced the concept of sequential pattern mining and proposed two similar algorithms based on the Apriori algorithm for mining sequential patterns, called AprioriAll and AprioriSome. The only difference between them is that AprioriAll counts all the large sequences including non-maximal sequences while AprioriSome only counts the maximal sequences. Experimental results show that both AprioriAll and AprioriSome have the excellent scale-up properties because they scale linearly as the number of itemsets or transactions. However, in the process of candidate sequences generation, when the number of candidate sequences is too large or the length of a candidate sequence is too long, the memory space required for both proposed algorithms will become much huge.

A novel sequential pattern mining method, called PrefixSpan (i.e., Prefix-projected Sequential pattern mining) is proposed by Han et al. [33], which aims to reduce the cost of candidate subsequence generation and the size of projected databases. The proposed method only examines the prefix subsequences rather than the whole sequences. Then only the postfix subsequences of these prefix subsequences corresponding to will be projected into the projected databases. In each projected database, the proposed method only mines local frequent patterns to obtain sequential patterns. Moreover, in order to improve mining efficiency, two kinds of database projections, named level-by-level projection and bi-level projection, are explored in [33]. Besides, a main-memory-based pseudo-projection technique is developed for saving the cost of projection and speeding up computation processing.

By identifying potential trends in data, approximate sequence pattern can summarizes and represents the local database effectively. An alternative local pattern analysis method for sequential patterns mining based on a local database of a multi-database is proposed by Kum et al. [34]. In order to mine approximate sequential patterns, called consensus patterns, the proposed method presented a novel algorithm, called ApproxMAP (for Approximate Multiple Alignment Pattern mining). The processing of ApproxMAP includes two phases. In the first phase, all sequences are clustered according to their similarity/distance. In the second phase, consensus patterns are mined from each cluster through ApproxMAP directly. The experimental results show that ApproxMAP can efficiently summarize a local database and reduce the cost for the following global mining.

An efficient sequential pattern mining method based on item co-occurrences is proposed by Fournier-Viger et al. [35] for candidate sequences pruning. It proposed a data structure called Co-occurrence MAP (CMAP) for storing co-occurrence information. Then the CMAP is used to prune candidates in three vertical algorithms (SPADE (Sequential PAttern Discovery using Equivalence classes) [36], SPAM (Sequential PAttern Mining) [37] and ClaSP (Closed Sequential Patterns algorithm) [38]). Extensive experimental results based on real-world datasets show that the co-occurrence-based pruning proposed in [35] is effective and CMAP can reduce a substantial amount of time for candidate sequence selecting.

Recently, in order to deal with complex real-world data mining problems, how to mining frequent patterns with periodic wildcard gaps has gained much research interest. A frequent pattern with periodic wildcard gaps is to add gap constraints to normal frequent patterns. It requires that the user-specified gap constraints be satisfied between any two of frequent patterns. The contents between the gaps can be any value. Usually, it is a predetermined symbol, which is referred to as a wildcard collectively. Frequent patterns with periodic wildcard gaps can reflect the common characteristics of items in a certain gap. User-specified patterns will be mined directly without a large number of meaningless patterns generated at the same time.

Previous methods usually use matrices or other linear data structures for frequent patterns with periodic wildcard gaps mining, which consume a large amount of memory and running time. An Incomplete Nettree structure is proposed by Wu et al. [39], which is used to calculate the number of supports of its superpatterns in a one-way scan. After that, two novel algorithms, named MAPB (Mining sequentiAl Pattern using incomplete Nettree with Breadth first search) and MAPD (Mining sequentiAl Pattern using incomplete Nettree with Depth first search), are used to solve the problem above with low memory requirements. Moreover, a heuristic algorithm called MAPBOK (MAPB for top-K) is proposed to process the Top-K frequent patterns in each length [39]. Experimental results based on a real-world biological dataset demonstrate the superiority of the proposed algorithms in running time and memory consumption.

2.1.3. Other pattern mining

Yan and Han [40] proposed a method called graph-based Substructure pattern mining (gSpan) for frequent graph-based pattern mining. A lexicographic order is constructed by gSpan firstly. Then gSpan mines frequent substructure patterns without the process of candidate generation by using depth-first search strategy. gSpan is the first algorithm to apply depth-first search in the field of frequent sub-graph mining. It also can be applied to mining large frequent subgraphs.

For a multi-branch company, how to make decision based on a branch's patterns is an important problem. Suganthi and Kamalakannan [41] proposed a new clustering method for exact synthesized frequent items and a new method to deal with exceptional patterns. The first one is used to provide detailed information of the associations among different types of items in a market. The second one is used to process the transportation cost of a company, manpower and non-profitable items. Experimental results show that when the items in a database are greatly associated, the proposed two methods work well.

2.2. Global pattern analysis

Local pattern analysis is to obtain frequent patterns form local databases. Then these obtained frequent patterns will be synthesized for global pattern analysis. The key of global pattern analysis is how to set up synthesizing models.

In the process of global pattern analysis, in addition to synthesizing the support and confidence of an association rule, the frequency of an association rule is also an important pattern. The frequency of an association rule is the number of data sources which contain this association rule. According to a threshold setting, an association rule can be high-frequent, low-frequent, or neither high-frequent nor low-frequent in different data sources [45].

Heavy association rules, high-frequent association rules and exceptional association rules are three specific types of association rules for synthesizing global patterns from local patterns in different data sources [45]. A heavy association rule is an association rule with high-support and high-confidence in multiple data sources. A high-frequent association rule is an association rule with high extraction in multiple data sources. An exceptional association rule is a heavy association rule but low-frequent. A high-frequent association rule is not necessarily a heavy association rule. Based on these definitions, two algorithms are proposed by Adhikari and Rao [45]. The proposed methods synthesized heavy association rules in multiple databases and notified whether a heavy association rule is high-frequent or exceptional.

In practice, high-frequent association rules are important for decision making. A high-frequent association rule means the rule is supported by most of databases and has larger chance to become useful rules in the union of all databases. A weighting model is proposed by Wu and Zhang [19] for synthesizing high-frequent association rules from multiple databases. In order to extract high-frequent association rules from multiple databases, the proposed method constructed a new association rule selection method to enhance the proposed weighting model. In the rule selection method, high-frequent association rules will be taken as relevant rules, while low-frequent rules will be taken as irrelevant rules. Moreover, when mining association rules from unknown data sources, a relative synthesizing model is constructed to synthesize these association rules by clustering of local association rules. However, the proposed weighting model has a limitation, which needs to be built on the databases that have the same type of data sources. For example, if a company with several branches has three types of businesses: super markets, banks and investment trusts, all data sources need to be classified into three classes according to the business types, then the data sources in each class can be synthesized by using the proposed weighting model.

In order to overcome the limitation of the above proposed weighting model, an improved weighting model is proposed by Ramkumar and Srinivasan [43] for synthesizing high-frequency rules. A data source weight is calculated based on its transaction population (i.e., the number of itemsets) of this data source. Based on the weights of data sources and the supports of local databases, the proposed methods can calculate the global support and global confidence. Experimental results show that the global support and confidence calculated by the improved weighting method is much close to the mono-mining (i.e., putting all databases into one big database to mine patterns) results.

Global pattern analysis intends to find a synthesized rule from local patterns in multiple databases to match with the mono-mining results closely. Some models such as weighting model and minimum supporting principle are based on linear combination. However, when the extracted patterns are not satisfying a minimum support threshold value, they cannot be used for the pattern synthesizing process. In some cases, this requirement may ignore some useful information/patterns that are meaningful for a right decision making. Thus, an improved probabilistic model based on the maximum entropy method is proposed for the synthesizing process by Khiat et al. [44]. Both clique method (a grid-based spatial clustering algorithm) and bucket elimination technique [44] are used for reducing the complexity of the maximum entropy method. The experiments are based on the frequent itemsets which are discovered by the Apriori algorithm, and the results show the efficiency of the proposed synthesizing method [44].

Zhang et al. [42] propose a method for synthesizing local patterns in multiple databases, which is especially suitable for finding global exceptional patterns. It considers that different data sources should have different minimal supports due to they are comprised of different databases. Therefore, it adopts a method of computing a pattern's importance by measuring the distance between a pattern's support and the corresponding database's minimal support. However, the proposed method does not consider that different data sources may have different effects in the synthesizing process, which may exist in reality [42].

Different from many other approaches that do not consider the relationship between the data in one database, a nonlinear method is proposed by Zhang et al. [20] to find the nonlinear relations of data in a database. The proposed method is called Kernel Estimation for Mining Global Patterns (KEMGP). The proposed method uses the kernel estimation method (a mapping method used Gaussian kernel) to capture the patterns of nonlinear relation when synthesizing local patterns for global patterns. The KEMGP can better characterize that different databases have different contributions to the global pattern.

Please cite this article as: R. Wang et al., Review on mining data from multiple data sources, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.01.013

2.3. Others

Besides, other techniques such as compressed sensing [46] and sparse learning [47] also are widely used in pattern analysis. Since a signal can be expressed in Fourier series or Taylor series, a new input signal can be reconstructed by several known signals.

The compressed sensing theory [46] considers that, as long as an input signal is compressible or sparse in a transform domain, by using an observation matrix, which is not related to the transform base, the transformed high-dimensional input signal can be projected onto a low-dimensional space. Then the input signal can be reconstructed by solving an optimization problem from a small number of projections with a high probability.

The purpose of signal sparse representation is to represent an input signal with as few known signals (i.e., atoms in a given overcomplete dictionary) as possible. A more concise representation of the input signal will be obtained so that the information contained in the input signal can be used in the further processing like compression and encoding. In pattern analysis, the sparse representation is used as follows.

In the process of global and local learning, minimizing both global and local learning cost functions is much helpful to utilize both local and global discriminative information of a target domain [48]. Based on this premise, a novel Multi-source Adaptation learning method with Global and Local Regularization (MAGLR) is proposed by Tao et al. [48] for visual classification tasks by exploiting multi-source adaptation regularization joint kernel sparse representation (ARJKSR). ARJKSR uses target and source domains to reconstruct a dataset by sparse representation. Based on the mixed Laplacian regularization (i.e., global sparsity preserving Laplacian regularization and local learning Laplacian regularization) semi-supervised learning diagram, MAGLR can obtain a label prediction matrix for unlabeled instances from the target domain. The proposed methods considered correlations among different source domains and applies sparse representation to multiple source domains including local and global regularizations simultaneously. However, how to select the optimal parameters for the multi-source domain adaptation settings has not been solved, which is still an open problem.

A multiple measurement vector (MMV) model is used to learn the properties of the distributions of the sources in the recovery processing and improve the recovery of the sources in [49]. By using a mixture of Gaussians prior, which is inspired by existing Sparse Bayesian Learning approaches, the proposed method is used for simultaneous sparse approximation. The proposed method has been tested on a variety of synthetic, simulated and real data with good performance. The proposed method assumes that there is no prior knowledge about the sources besides the number of components in the Mixture of Gaussians distribution. In some cases, improved performance can be achieved by taking further knowledge of the prior into account.

In practice, global patterns in multiple databases are synthesized according to the independent characteristics of local patterns. The pipelined feedback technique (PFT) [50] is one of the most effective techniques in local pattern analysis (LPA) for this task. Before PFT is proposed, the Single Database Mining Technique (SDMT) was used to mine local databases in multiple databases. However, SDMT has a limitation related to sizes of databases: (i) when each of the local databases is small, SDMT can mine the union of all databases; (ii) when at least one of the local databases is large, SDMT can mine every local database, but fail to mine the union of all local databases; (iii) when at least one of the local databases is very large, SDMT fails to mine every local database [50]. In PFT, once a pattern is extracted from a local database, then it will be extracted from the remaining databases. PFT improves synthesized patterns as well as local patterns analysis significantly.

In order to analyze the effect of database grouping on multi-database mining, an approach of non-local pattern analysis (NLPA), which combined database grouping algorithm and PFT, is proposed in [51]. Non-local pattern analysis is more attractive than LPA when the frequency of data mining is lower, which is helpful to estimate a lesser number of patterns when synthesizing the global pattern. Experimental results show the disadvantage of non-local pattern analysis is that the pattern reported only from the last few groups which may result in the experimental error worse. This is due to the absence of a pattern reported which has been assumed to exist. Thus, the definition of the depth of a pattern is proposed to solve this drawback. It means a pattern will become useless if its depth is low.

3. Key methods for data source classification and clustering

In real world applications, a large amount of data from different sources normally is stored in different geological locations. In order to reduce data mining costs, these data can be classified or clustered before data mining. Based on the outcomes of clustering or classification, data sources relevant to a given data source mining application can be identified easily. This step is referred as data source selection.

3.1. Data source classification

Data source classification [52] is a preprocess for multiple data source mining, which is to classify given data sources by comparing their features and measuring the relevance between different data sources using a classifier [53]. Normally, the performance of the classifier is measured in three aspects: (i) accuracy, (ii) computational complexity, and (iii) the simplicity of the model descriptions.

Wu et al. [53] propose an application-independent multiple data source classification method, which is called BestClassification. This method has two phases. In the first phase, it divides all data sources by using the proposed GreedyClass algorithm. The data sources are divided into several smaller sets according to their similarity/distances (the similarity of two databases is measure by the number of the common itemsets that both database have divided by the total itemsets of both datasets) when a threshold is assigned. In the second phase, it finds a best classification for these data sources by using the proposed BestClassification algorithm. If there is no such the best classification for these data sources, a trivial best classification is produced. The best classification in this paper means that all data sources can be classified properly. However, since the time complexity of GreedyClass is $O(n^4)$, when n (the number of databases) is very large, the time complexity will become very large. At the same time, BestClassification does not always obtain the best classification from the multiple data sources.

An improved data classification algorithm for multiple data source classification is proposed by Li et al. [54]. The new algorithm, named as CompleteClass, has less time complexity comparing with GreedyClass in [53] since CompleteClass sets a smaller similarity threshold and decreases the number of data sources in a class. The complexity of CompleteClass is $O(n^3)$, where n is the number of databases. BestCompleteClass, an improve version of BestClassification, can always achieve the complete classification (i.e., the best classification) for multiple data sources. Theoretical analysis and experimental results verify the effectiveness of the algorithm. However, the running time of this method is still long when dealing with a large amount of data.

Based on these two classification algorithms, a new multiple data source classification method is presented by Dingrong et al.

[55], which is called application-independent database classification. It can achieve higher intra-class cohesion and lower inter-class coupling compared with BestClassification and BestCompleteClass. In order to search for the best complete classification, there are three phases in this new algorithm. In the first phase, it computes the similarity between each data source. In the second phase, it chooses several databases from all data sources and classifies them into two classes based on their similarities. In the final phase, it classifies other databases from all data sources into the given classes in the second phase. Then it will inspect whether the classification satisfy the given condition, if it does then end the classification, otherwise recursive calls for further classification. To prove the efficiency of the new algorithm, Dingrong et al. compares the new method with the methods in [54] and [53]. The experimental results show the new method uses less time cost to achieve the best complete classification for any kind of multiple data sources.

3.2. Data source clustering

Data source clustering is another data preprocess for multiple data source mining. Different from data source classification, it is a kind of unguided learning. In other words, data clustering is to cluster data according to the similarity between data without knowing the classes they belong to in advance.

A simple method for multiple data source clustering to achieve high cohesion and low coupling is proposed by Tang and Zhang [56]. Similar to the proposed method in [53], it clusters data sources based on the similarity between each data source. The proposed method firstly constructs a multi-objective optimization problem, then uses a hierarchical clustering algorithm, which clusters multiple data sources based on the similarity between each data sources. The hierarchical clustering algorithm can avoid instability and reduce time complexity, to find the optimal cluster structure. The multi-objective optimization problem is to find a set of decision variables, which satisfy constraints as well as all objective functions and function values for each target (Pareto optimal solutions), and provide the decision variables to a decision maker. The best cluster is evaluated by the cophenetic correlation coefficient. Comparing with the BestClassification algorithm, the experimental results show that this clustering method has stronger stability, lower time complexity and stronger generalization ability.

For big data, Xin et al. [57] provide two novel algorithms for multiple data source clustering: a multi-source cross-domain classification (MSCC) algorithm, which is based on the existing maximum consistency function model and Newton gradient descent method, and the MSCC Dual coordinate descent method, which is suitable to deal with large dataset.

The MSCC algorithm is proposed to implement efficient cross-domain classification of target domains using a logistic regression model and a proposed consensus measure. After that, in order to use MSCC to deal with big data in the field of multiple data source mining, a MSCC's fast version MSCC-CDdual is derived and theoretically analyzed. MSCC-CDdual is based on the algorithm CDdual (Dual coordinate descent method) [58] as the recent advance on large-scale logistic regression which can achieve a fast speed, high classification accuracy and good domain adaption for multiple big data source mining.

In [57], it proposed a new coherence function model, which makes it easy to evolve the MSCC algorithm into a fast algorithm for large samples. MSCC can reduce the error rate of the classifier trained from a single source domain and improve the accuracy of comprehensive classification. A new fast algorithm MSCC-CDdual shows its operating speed advantage from the perspective of high-dimensional data in [57]. Unlike the existing cross-domain algorithms, MSCC is oriented to multiple data source domain learning rather than single data source domains. Thus, it has a higher data mining accuracy and data mining speed. However, for multiple high-dimensional big data sources, the classifying and mining efficiency of this method still need to be improved.

Table 1
Multiple source data fusion methods in different fusion levels.

Pixel-level	Feature-level	Decision-level
Algebraic method	Bayesian	Knowledge-based fusion
HIS transform	Dempster_shafer	Dempster_shafer
High – pass filtering	Entropy method	Fuzzy set theory
Regression model	Weighted average	Reliability theory
Best variable substitution	Neural network	Bayesian
Kalman filter	Clustering	Neural network
Wavelet transform	Voting	Logical template

4. Key methods for data source fusion

Multiple data source fusion is a process of combining data from multiple data sources to achieve higher accuracy and more specific ratiocinations. Multiple data source fusion has been used in many areas, including image processing (i.e., multi-sensor data fusion).

Multi-sensor data fusion is an intelligent synthesis process of remote sensing image data from multiple sources, and to generate accurate, complete and reliable estimation and judgment [59]. The process is to collect the information from different sensors firstly, and then to eliminate the data redundancy and data conflictions that may exist among the multi-sensor information. Multi-sensor data fusion can improve the timeliness and reliability of remote sensing information. This is very useful for decision making, planning, and verification [60]. The main application areas of data fusion include multiple data source image composite [61], robot and intelligent instrument system [62], battlefield and unmanned aircraft [63], image analysis and understanding, target detection and tracking [64], and automatic target recognition [65].

There are three types of data fusion methods for remote sensing images: (i) at the pixel level, (ii) at the feature level, and (iii) at decision level. The level of fusion is in an order from low to high. The process of multi-sensor data fusion has four steps. In the first step, remote sensing image data from multiple data sources. The second step is to do feature extraction. Pixel, feature or decision level fusion is used in the third step. The last step it to do fusion attribute description. Table 1 lists several multiple source data fusion methods in different fusion levels.

4.1. Pixel-level multi-sensor data fusion

Pixel-level fusion is a low level of data fusion. The advantages of pixel-level fusion are to retain as much information as possible. This fusion approach has the highest precision. The main pixel-level fusion methods include Principal Component Transformation (PCT) [66], HIS transform [68], and wavelet transform [68].

Principal Component Transformation (PCT), which is also called Karhunen–Loeve Transform (K–L transform), is a multidimensional linear transformation based on statistical features of an image. It has the effect of variance information enrichment and data compression, which can more accurately reveal the remote sensing information within the multi-band data structure [66]. The process of remote sensing data fusion by using PCT has four steps [67]. The first step uses the input of multi-band remote sensing data for spatial registration and principal component transformation. The second step uses the results of spatial registration and principal component transformation for histogram matching. The third step uses the results of histogram matching to replace the fusion process components in the first principal multi-source remote sensing data. The last step generates a multi-band fused image with high spatial resolution by using principal component inverse conversion.

Due to the physical constraints between spatial and spectral resolutions, poor-resolution multispectral (MS) data is desirable. Thus, a high-resolution panchromatic observation needs to be processed before multispectral image fusion. Two general and formal solutions, which are based on undecimated discrete wavelet transform and generalized Laplacian pyramid, respectively, are proposed by Aiazzi et al. [69]. Undecimated discrete wavelet transform is an octave band pass representation, which is to omit all decimators and upsample the wavelet filter bank from a conventional discrete wavelet transform. Generalized Laplacian pyramid is another oversampled structure, which is obtained by recursively subtracting an expanded decimated lowpass version from an image. The proposed solutions based on these two methods above are to selectively perform spatial-frequency spectrum substitution from one image to another. Different from other multiscale fusion methods, the advantages of the proposed methods include: (i) Context dependency in the proposed solutions is achieved by thresholding local correlation coefficient between the images to be merged. This can avoid some spatial details which are not likely occurring to be injected into the target image. (ii) The proposed solutions use uncritically subsampling to avoid possible impairments in the fused images due to missing cancellation of aliasing terms. The results presented and discussed on the SPOT data [71] show its effectiveness.

In recent years, wavelet decomposition has become an attractive wavelet-based fusion technique to fuse multisensor images. Normally, the input images are decomposed using an orthogonal wavelet to extract features, which are combined through an appropriate fusion rules. Then the fused image is reconstructed by applying inverse wavelet transform. A nonorthogonal (or redundant) wavelet is proposed in [70] as an alternative method for feature extraction. Compared to the normal orthogonal wavelet decomposition, the redundant wavelet decomposition is superior to the standard orthogonal wavelet decomposition, regardless of the fusion rules.

After that, a new alternative method for multispectral and panchromatic image fusion based on a high-pass filtering procedure is proposed by González-Audícana et al. [68]. Multi-resolution wavelet decomposition is used to perform detail feature extraction. In this paper, Intensity Hue-Saturation (IHS) and Principal Component Analysis (PCA) are used to inject the spatial details of a full-color image into a multispectral one. A higher quality image can be obtained due to the selective inclusion of the spatial details of the full-color image, which is missed in the multispectral image. The method proposed improves the quality of synthetic images by standard IHS, PCA and standard wavelet based fusion methods. For the two proposed fusion methods in [68], better results are obtained when an undecimated algorithm is used to perform the multi-resolution wavelet decomposition.

4.2. Feature-level multi-sensor data fusion

Feature-level fusion is a medium level of data fusion, which includes three steps. The first step extracts features from all remote sensing image data, the extracted features should be a sufficient representation of the original information or statistically sufficient. According to the feature information, all data from different sources is classified, clustered and synthesized to generate eigenvectors in the second step. In the third step, some feature-level fusion methods are used to fuse these generated eigenvectors to make a feature description based on fusion eigenvectors.

A feature-level image fusion method based on segmented regions and neural networks is proposed by Wang et al. [72] to process the images from CCD (Charge-coupled Device) devices. Firstly, a source image is divided into a set of common areas as a preprocessing of the entire image. After that, the proposed method will select the corresponding segmentation areas from the source image, respectively, and extracts features which represent the clarity of the corresponding regions. Then the extracted features will be feed into a neural network to determine a clear region to reconstruct the final fusion image. The experimental results show that the effectiveness of this new fusion method was superior to previous methods.

A new scalable feature-level sensor fusion architecture is proposed by Kaempchen et al. [73], which combines the data from a multi-layer laser scanner with a monocular video to increase the functionality of driver assistance systems. In order to enhance the precision of object tracking, the proposed method established a new geometric object model for a large diversity of object shapes recognition. Besides, a new free form object tracking approach is used for a precise velocity estimation of trucks on highways. The proposed method is designed to maximize synergetic by combining feature-level measurement features and keeping the fusion architecture as general as possible. Experimental results in an actual traffic scene show that the proposed method outperforms the other known reference algorithms.

Canonical Correlation Analysis (CCA) is a dimensionality reduction algorithm commonly used in image processing. When extracting various features from an image, each feature can form a linear space, and CCA can be used to analyze the correlation between these spaces. Based on CCA, a new feature extraction method based on feature fusion is proposed by Sun et al. [74]. It is the first time to apply CCA to feature fusion and image recognition. The proposed method extracts two groups of feature vectors with the same pattern firstly, then uses the correlation feature of two groups of feature to form effective discriminant vector for recognition. The proposed method not only is suitable for information fusion, but also eliminates the redundant information within the features, which offers a new way of applying two groups of feature fusion to discrimination. Experimental results on the Concordia University CENPARMI database of handwritten Arabic numerals and Yale face database show that the recognition rate is much higher than the use of a single feature or existing fusion algorithms.

4.3. Decision-level multi-sensor data fusion

Decision-level fusion is cognitive-based fusion, which is the highest level of data fusion and the highest level of abstraction. Aiming at the specific requirements of the proposed problem, decision-level fusion utilizes several sub-decisions or features to yield a final or higher decision.

Decision-level fusion includes two phases. The first phase is to make feature description for each data. The second phase fuses the results in the last phase to get the fusion feature description for the target or environment. The advantages of the decision-level fusion include: strong fault tolerance, good openness, short processing time, low data requirement and strong analytical ability. Among the three fusion levels, the computational complexity of decision-level fusion is the smallest. However, due to the higher requirements for feature extraction, the cost of the decision-level fusion is higher than the other two levels of fusion. In some cases, decision-level fusion has a strong dependence on the other two levels fusion.

In practice, a new method for automatic ovarian tumor classification based on the decision level fusion is proposed by Khazendar et al. [75]. Firstly, the proposed method extracts two different types of features (i.e., histograms and local binary patterns) from ultrasound images of an ovary. Then ovarian tumor is classified according to the features of an ultrasound image using Support Vector Machine (SVM). At this stage, the proposed method uses the new decision-level fusion method, which classifies SVM-based decision scores into a confidence measure to help the final diagnosis decision making. Experimental results show that such confidence-based prediction outcomes are much closer to the reality in diagnosis of ovarian cancers.

The Facial Action Coding System (FACS) is the taxonomy of human facial expressions designed to facilitate human annotation of facial expressions. The atomic facial muscle actions in this system are called Action Units (AUs). Based on a new region-based face representation and a mid-level region-specific decision layer, a new method for AU detection is proposed in [76]. The proposed method uses domain knowledge (i.e., expert knowledge of how FACS is defined) regarding AU-specific facial muscle contractions to define a set of facial regions covering the entire face. Other approaches, however, typically represent a face as a regular grid based on the locations on a face only (holistic representation), which may not utilize the full facial appearance. The proposed method then uses an AU-specific weighted sum model as a decision-level fusion layer for region-specific probabilistic information combination. It allows each classifier to learn the typical appearance changes for a specific face part and reduces the dimensionality of the problem. Experimental results based on the DISFA [77] and GEMEP-FERA [78] datasets show superior performance of both domain-specific region definition and decision-level fusion.

More recent, a novel multi-sensor fusion framework is proposed by Kumar et al. [79] using Coupled Hidden Markov Model (CHMM) for Sign Language Recognition (SLR). This framework calibrates two different sensors (named Leap Motion [80] and Kinect [81]) with different depth information and frames per second (fps) to form a multi-sensor data framework for capturing dynamic sign gestures. Leap Motion is kept below a signer's hand to sense the movements of signer's hand and fingers in horizontal. Kinect is kept in front of the signer to acquire the movements of signer's hand and fingers in vertical. Experimental results prove that new CHMM fusion framework is better than the single sensor-based SLR camera system.

5. Conclusion and future work

With the continuous development of data mining technology, research in multiple data sources mining is becoming more imperative and important. It has a wide range of applications in the fields such as robotics, automation and intelligent system design. This has been and will continue to be a growing interest in the research community to develop more advanced data mining methods and architectures. This paper critically reviewed many useful methods to mine meaningful information and discover new knowledge from multiple data sources: (i) pattern analysis, (ii) multiple data source classification, (iii) multiple data source clustering, and (iv) multiple data source fusion. There are still several challenges in these three effective approaches, which need further research.

Acknowledgments

This work is in part supported by the Marsden Fund, New Zealand, the National Natural Science Foundation of China (U1435214, U1609215, 51507084, 61363037), the Natural Science Foundation of Zhejiang Province, China (LY18F020008), the Youth Innovation Research Team of Sichuan Province, China (2015TD0020), CDUT (KYTD201404), the Primary R&D Plan of Jiangsu Province, China (BE2015213), the China Postdoctoral Science Foundation (2016M591890), the National Social Science Foundation of China (17BJY033), and the Chinese Scholarship Council PhD scholarships.

References

[1] X. Zhu, L. Zhang, Z. Huang, A sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process. 23 (9) (2014) 3737–3750.
[2] X. Zhu, Z. Huang, H. Cheng, J. Cui, H.T. Shen, Sparse hashing for fast multimedia search, ACM Trans. Inf. Syst. (TOIS) 31 (2) (2013) 9.
[3] A. Adhikari, P. Ramachandrarao, W. Pedrycz, Mining multiple large databases, in: Developing Multi-Database Mining Applications, 2010, pp. 37–50.
[4] http://www.theaustralian.com.au/national-affairs/treasury/australian-taxation-office-turns-to-facebook-to-catch-cheats/news-story/6d4bca2d0223dcb3924c061fa28bc3f9.
[5] S. Zhang, C. Zhang, Q. Yang, Data preparation for data mining, Appl. Artif. Intell. 17 (5–6) (2003) 375–381.
[6] S. Zhang, Z. Qin, C.X. Ling, S. Sheng, "Missing is useful": missing values in cost-sensitive decision trees, IEEE Trans. Knowl. Data Eng. 17 (12) (2005) 1689–1693.
[7] R. Hu, X. Zhu, D. Cheng, W. He, Y. Yan, J. Song, S. Zhang, Graph self-representation method for unsupervised feature selection, Neurocomputing 220 (2017) 130–137.
[8] X. Zhu, H.I. Suk, S.W. Lee, D. Shen, Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification, IEEE Trans. Biomed. Eng. 63 (3) (2016) 607–618.
[9] X. Zhu, X. Li, S. Zhang, Block-row sparse multiview multilabel learning for image classification, IEEE Trans. Cybern. 46 (2) (2016) 450–461.
[10] M.W. Bright, A.R. Hurson, S.H. Pakzad, A taxonomy and current issues in multidatabase systems, Computer 25 (3) (1992) 50–60.
[11] S. Zhang, X. Li, M. Zong, X. Zhu, R. Wang, Efficient kNN classification with different numbers of nearest neighbours, IEEE Trans. Neural Networks Learn. Syst. (2017), doi:10.1109/TNNLS.2017.2673241.
[12] S. Zhang, X. Li, M. Zong, X. Zhu, D. Cheng, Learning k for knn classification, ACM Trans. Intell. Syst. Technol. (TIST) 8 (3) (2017) 43.
[13] X. Zhu, X. Li, S. Zhang, C. Ju, X. Wu, Robust joint graph sparse coding for unsupervised spectral feature selection, IEEE Trans. Neural Networks Learn. Syst. 26 (6) (2016) 1263–1275.
[14] S. Zhang, Z. Jin, X. Zhu, Missing data imputation by utilizing information within incomplete instances, J. Syst. Softw. 84 (3) (2011) 452–459.
[15] A. Turinsky, R. Grossman, A framework for finding distributed data mining strategies that are intermediate between centralized strategies and in-place strategies, in: Proceedings of Workshop on Distributed and Parallel Knowledge Discovery at KDD-2000, 2000, pp. 1–7.
[16] X. Wu, C. Zhang, S. Zhang, Database classification for multi-database mining, Inf. Syst. 30 (1) (2005) 71–88.
[17] S. Zhang, C. Zhang, X. Wu, Data mining and multi-database mining, in: Knowledge Discovery in Multiple Databases, 2004, pp. 27–61.
[18] S. Zhang, M.J. Zaki, Mining multiple data sources: local pattern analysis, Data Mining Knowl. Discov. 12 (2) (2006) 121–125.
[19] X. Wu, S. Zhang, Synthesizing high-frequency rules from different data sources, IEEE Trans. Knowl. Data Eng. 15 (2) (2003) 353–367.
[20] S. Zhang, X. You, Z. Jin, X. Wu, Mining globally interesting patterns from multiple databases using kernel estimation, Expert Syst. Appl. 36 (8) (2009) 10863–10869.
[21] X. Yan, C. Zhang, S. Zhang, Toward databases mining: preprocessing collected data, Appl. Artif. Intell. 17 (5–6) (2003) 545–561.
[22] S. Zhang, Q. Chen, Q. Yang, Acquiring knowledge from inconsistent data sources through weighting, Data Knowl. Eng. 69 (8) (2010) 779–799.
[23] S. Zhang, M. Zong, K. Sun, Y. Liu, D. Cheng, Efficient kNN algorithm based on graph sparse reconstruction, in: International Conference on Advanced Data Mining and Applications, 2014, pp. 356–369.
[24] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, Z. Xu, Missing value estimation for mixed-attribute data sets, IEEE Trans. Knowl. Data Eng. 23 (1) (2011) 110–121.
[25] S. Zhang, X. Wu, C. Zhang, Multi-database mining, IEEE Comput. Intell. Bull. 2 (1) (2003) 5–13.
[26] Y. Chen, F. Li, J. Fan, Mining association rules in big data with NGEP, Cluster Comput. 18 (2) (2015) 577–585.
[27] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: 20th International Conference Very Large Data Bases, VLDB, 1215, 1994, pp. 487–499.
[28] J. Han, J. Pei, Y. Yin, R. Mao, Mining frequent patterns without candidate generation: a frequent-pattern tree approach, Data Mining Knowl. Discov. 8 (1) (2004) 53–87.
[29] J. Yan, N. Liu, Q. Yang, B. Zhang, Q. Cheng, Z. Chen, Mining adaptive ratio rules from distributed data sources, Data Mining Knowl. Discov. 12 (2–3) (2006) 249–273.
[30] Z. Abdullah, T. Herawan, N. Ahmad, M.M. Deris, Mining significant association rules from educational data using critical relative support approach, Proc. Soc. Behav. Sci. 28 (2011) 97–101.
[31] J. Sahoo, A.K. Das, A. Goswami, An efficient approach for mining association rules from high utility itemsets, Expert Syst. Appl. 42 (13) (2015) 5754–5778.
[32] R. Agrawal, R. Srikant, Mining sequential patterns, in: Proceedings of the Eleventh International Conference on Data Engineering, 1995, pp. 3–14.
[33] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, M.C. Hsu, Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth, in: Proceedings of the 17th International Conference on Data Engineering, 2001, pp. 215–224.
[34] H.C. Kum, J.H. Chang, W. Wang, Sequential pattern mining in multi-databases via multiple alignment, Data Mining Knowl. Discov. 12 (2-3) (2006) 151–180.

[35] P. Fournier-Viger, A. Gomariz, M. Campos, R. Thomas, Fast vertical mining of sequential patterns using co-occurrence information, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2014, pp. 40–52.
[36] M.J. Zaki, SPADE: an efficient algorithm for mining frequent sequences, Mach. Learn. 42 (1) (2001) 31–60.
[37] J. Ayres, J. Flannick, J. Gehrke, T. Yiu, Sequential pattern mining using a bitmap representation, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002, pp. 429–435.
[38] A. Gomariz, M. Campos, R. Marin, B. Goethals, ClaSP: an efficient algorithm for mining frequent closed sequences, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2013, pp. 50–61.
[39] Y. Wu, L. Wang, J. Ren, W. Ding, X. Wu, Mining sequential patterns with periodic wildcard gaps, Appl. Intell. 41 (1) (2014) 99–116.
[40] X. Yan, J. Han, gspan: graph-based substructure pattern mining, in: Proceedings of IEEE International Conference on Data Mining, 2002, pp. 721–724.
[41] R. Suganthi, P. Kamalakannan, Exceptional patterns with clustering items in multiple databases, Indian J. Sci. Technol. 8 (31) (2015) 1–10.
[42] C. Zhang, M. Liu, W. Nie, S. Zhang, Identifying global exceptional patterns in multi-database mining, IEEE Intell. Inf. Bull. 3 (1) (2004) 19–24.
[43] T. Ramkumar, R. Srinivasan, Modified algorithms for synthesizing high-frequency rules from different data sources, Knowl. Inf. Syst. 17 (3) (2008) 313–334.
[44] S. Khiat, H. Belbachir, S.A. Rahal, Probabilistic models for local patterns analysis, University of Sciences and Technology of Oran, 2015.
[45] A. Adhikari, P.R. Rao, Synthesizing heavy association rules from different real data sources, Pattern Recognit. Lett. 29 (1) (2008) 59–71.
[46] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[47] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892.
[48] J. Tao, S. Wen, W. Hu, Multi-source adaptation learning with global and local regularization by exploiting joint kernel sparse representation, Knowl. Based Syst. 98 (2016) 76–94.
[49] R. Porter, V. Tadic, A. Achim, Sparse Bayesian learning for non-Gaussian sources, Digit. Signal Process. 45 (2015) 2–12.
[50] A. Adhikari, P. Ramachandrarao, B. Prasad, J. Adhikari, Mining multiple large data sources, Int. Arab J. Inf. Technol. 7 (3) (2010) 241–249.
[51] A. Adhikari, L.C. Jain, S. Ramanna, Analysing effect of database grouping on multi-database mining, IEEE Intell. Inf. Bull. 12 (1) (2011) 25–32.
[52] X. Zhu, Q. Xie, Y. Zhu, X. Liu, S. Zhang, Multi-view multi-sparsity kernel reconstruction for multi-class image classification, Neurocomputing 169 (2015) 43–49.
[53] X. Wu, C. Zhang, S. Zhang, Database classification for multi-database mining, Inf. Syst. 30 (1) (2005) 71–88.
[54] H. Li, X. Hu, Y. Zhang, An improved database classification algorithm for multi-database mining, in: International Workshop on Frontiers in Algorithmics, 2009, pp. 346–357.
[55] Y.U.A.N. Dingrong, F.U. Huiwen, L.I. Zhixin, W.U. Hongyi, An application-independent database classification method based on high cohesion and low coupling, J. Inf. Comput. Sci. 9 (15) (2012) 4337–4344.
[56] H. Tang, M. Zhang, A simple methodology for database clustering, in: Proceedings of Science (CENet), 2015, pp. 1–7.
[57] G. Xin, S. Wang, M. Xu, A new cross-multidomain classification algorithm and its fast version for large datasets, Acta Autom. Sin. 40 (3) (2014) 531–547.
[58] C.J. Hsieh, K.W. Chang, C.J. Lin, S.S. Keerthi, S. Sundararajan, A dual coordinate descent method for large-scale linear SVM, in: Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 408–415.
[59] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (1) (2013) 28–44.
[60] R.C. Luo, M.H. Lin, R.S. Scherp, Dynamic multi-sensor data fusion system for intelligent robots, IEEE J. Robot. Autom. 4 (4) (1988) 386–396.
[61] Y. Zhang, S. Prasad, Locality preserving composite kernel feature extraction for multi-source geospatial image analysis, IEEE J. Selected Top. Appl. Earth Observ. Remote Sens. 8 (3) (2015) 1385–1392.
[62] R.C. Luo, M.G. Kay, Multisensor integration and fusion in intelligent systems, IEEE Trans. Syst. Man Cybern. 19 (5) (1989) 901–931.
[63] M.P. Đurišić, Z. Tafa, G. Dimić, V. Milutinović, A survey of military applications of wireless sensor networks, in: 2012 Mediterranean Conference on Embedded Computing (MECO), 2012, pp. 196–199.
[64] Y. Bar-Shalom, P.K. Willett, X. Tian, Tracking and Data Fusion, YBS Publishing, 2011.
[65] J. Llinas, D.L. Hall, An introduction to multi-sensor data fusion, in: Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, 1998. ISCAS'98, 6, 1998, pp. 537–540.
[66] A. Dang, X. Wang, X. Chen, in: ERDAS IMAGINE Remote Sensing Image Processing Method, Tsinghua University Press, 2003, pp. 1–6.
[67] G. Wei, S. Yang, Multisource RS data fusion based on principal component analysis, J. Lanzhou Jiaotong Univ. (Nat. Sci.) 24 (3) (2005) 57–60.
[68] M. González-Audícana, J.L. Saleta, R.G. Catalán, R. García, Fusion of multispectral and panchromatic images using improved IHS and PCA mergers based on wavelet decomposition, IEEE Trans. Geosci. Remote Sens. 42 (6) (2004) 1291–1299.
[69] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis, IEEE Trans. Geosci. Remote Sens. 40 (10) (2002) 2300–2312.
[70] Y. Chibani, A. Houacine, Redundant versus orthogonal wavelet decomposition for multisensor image fusion, Pattern Recognit. 36 (4) (2003) 879–887.
[71] P. Terretaz, Comparison of different methods to merge SPOT P and XS data: evaluation in an urban area, in: Proceedings of the 17th Symposium of EARSeL, Future Trends in Remote Sensing, Lyngby, Denmark, 1997, pp. 17–20.
[72] R. Wang, F. Bu, H. Jin, L. Li, A feature-level image fusion algorithm based on neural networks, in: Proceedings of the First International Conference on Bioinformatics and Biomedical Engineering, 2007, pp. 821–824.
[73] N. Kaempchen, M. Buehler, K. Dietmayer, Feature-level fusion for free-form object tracking using laserscanner and video, in: Proceedings IEEE Intelligent Vehicles Symposium, 2005, pp. 453–458.
[74] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng, D.S. Xia, A new method of feature fusion and its application in image recognition, Pattern Recognit. 38 (12) (2005) 2437–2448.
[75] S. Khazendar, H. Al-Assam, et al., Automated classification of static ultrasound images of ovarian tumours based on decision level fusion, in: Computer Science and Electronic Engineering Conference (CEEC), 2014, pp. 148–153.
[76] B. Jiang, B. Martinez, M.F. Valstar, M. Pantic, Decision level fusion of domain specific regions for facial action recognition, in: 2014 22nd International Conference on Pattern Recognition (ICPR), 2014, pp. 1776–1781.
[77] S.M. Mavadati, M.H. Mahoor, K. Bartlett, P. Trinh, Automatic detection of non-posed facial action units, in: 2012 19th IEEE International Conference on Image Processing (ICIP), 2012, pp. 1817–1820.
[78] M.F. Valstar, M. Mehu, B. Jiang, M. Pantic, K. Scherer, Meta-analysis of the first facial expression recognition challenge, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 42 (4) (2012) 966–979.
[79] P. Kumar, H. Gauba, P.P. Roy, D.P. Dogra, Coupled hmm-based multi-sensor data fusion for sign language recognition, Pattern Recognit. Lett. 86 (2017) 1–8.
[80] https://www.leapmotion.com/.
[81] https://developer.microsoft.com/en-us/windows/kinect.

Please cite this article as: R. Wang et al., Review on mining data from multiple data sources, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.01.013