

## Stellar spectra association rule mining method based on the weighted frequent pattern tree \*

Jiang-Hui Cai<sup>1</sup>, Xu-Jun Zhao<sup>1,†</sup>, Shi-Wei Sun<sup>2</sup>, Ji-Fu Zhang<sup>1</sup> and Hai-Feng Yang<sup>1</sup>

<sup>1</sup> School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China; [cjhjj@sohu.com](mailto:cjhjj@sohu.com)

<sup>2</sup> National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

Received 2012 July 2; accepted 2012 October 18

**Abstract** Effective extraction of data association rules can provide a reliable basis for classification of stellar spectra. The concept of stellar spectrum weighted itemsets and stellar spectrum weighted association rules are introduced, and the weight of a single property in the stellar spectrum is determined by information entropy. On that basis, a method is presented to mine the association rules of a stellar spectrum based on the weighted frequent pattern tree. Important properties of the spectral line are highlighted using this method. At the same time, the waveform of the whole spectrum is taken into account. The experimental results show that the data association rules of a stellar spectrum mined with this method are consistent with the main features of stellar spectral types.

**Key words:** methods: data analysis — stars: fundamental parameters — techniques: spectroscopic — astronomical data bases: miscellaneous

### 1 INTRODUCTION

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST, also called the Guo Shou Jing Telescope) is a special reflecting Schmidt telescope constructed so that its optical axis is fixed in the meridian plane. LAMOST's special design allows both a large aperture (effective aperture of 3.6m–4.9m) and a wide field of view (Cui et al. 2012). Compared with other large telescopes around the world, LAMOST is able to acquire celestial spectral data with the highest rate. Since the completion of LAMOST, there have been so many astronomical observations gathered by LAMOST that the associated “data avalanche” and “information explosion” have become urgent problems. For example, Wu et al. (2010) reports the discovery of eight new quasars by the LAMOST telescope in one extragalactic field. In order to effectively extract information from a large amount of complex, distributed and multi-band astronomical data, research on data fusion and data mining has gradually become one of the hot topics in astronomy. At the same time, it is challenging to analyze the massive amount of data collected by LAMOST; therefore, there is a pressing need to develop novel data analysis tools to be integrated into data processing systems located in observatory facilities. Association rule mining, as an important data mining method, can effectively discover relationships between properties derived from the spectra of large celestial bodies, and has great potential in studying the origin and evolution of the universe.

---

\* Supported by the National Natural Science Foundation of China.

† Corresponding author: Xu-Jun Zhao; [d\\_mining@126.com](mailto:d_mining@126.com)

In recent years, research on celestial spectral data processing has mainly addressed automatic spectral classification and recognition. Dimension reduction using the methods of principal component analysis and neural networks was applied to classifying simulated data of galactic spectra with low signal-to-noise ratio by Folkes et al. (1996), where classification accuracy reached more than 90%. By combining the Fisher information matrix with kernel techniques, the method of generalized discriminant analysis was used to classify stellar spectra by Xu et al. (2006). Based on the research of Xu et al. (2006), Yang et al. (2007) adopted the support vector method by integrating kernel techniques and with a covering algorithm to classify quasars. The multi-band data fusion method based on the Map-Reduce model in a distributed environment was proposed by Zhao et al. (2010).

The currently available methods used for automated extraction of stellar spectral parameters are mainly based on minimum distance methods (MDMs), and a representative line of research is the ELODIE online stellar parameter estimation system by Katz et al. (1998) and Soubiran et al. (1998). They constructed a stellar spectral template library composed of 211 FGK type stellar spectral templates with a resolution of 0.1 Å. Recio-Blanco et al. (2006) designed the MATISSE algorithm to automatically derive the parameters and chemical abundances for the Gaia/RVS survey. Spectral classification by astronomers has mainly focused on stellar classification (von Hippel et al. 1994; Bailer-Jones et al. 1998; Bai et al. 2005) and galaxy classification (Connolly et al. 1995; Galaz & de Lapparent 1998; Zaritsky et al. 1995), where the latter usually needs to know redshifts of the spectra. The spectra with unknown redshifts will be roughly classified into three types: star, galaxy or quasar (QSO). Qin et al. (2003) and Zhang & Zhao (2003, 2004) have done some research on the rough classification of spectra using support vector machines (SVMs) and radial basis function (RBF) neural networks. General descriptions of the evolution of Young Stellar Objects and the process of planet formation can be found in the monographs of Hartmann (2001), Stahler & Palla (2005), Armitage (2010), Ward-Thompson & Whitworth (2011) and Garcia (2011), the proceedings of the Protostars and Planets V conference (Reipurth et al. 2007), as well as the recent reviews of Williams & Cieza (2011) and Armitage (2011). The spectral data correlation analysis system based on the constrained frequent pattern (FP) tree was presented by Zhao et al. (2008), but constraint conditions can only be artificially generated, and in the absence of expert information, a constrained FP-tree will give degraded performance.

Due to the lack of a priori knowledge from astronomical experts and users in stellar spectral data sets, the degree of importance for each item is difficult to set in advance, though each item has a different importance. According to the item's different importance in the stellar spectral data set, this paper introduces two concepts about the spectra of stellar weighted itemsets and a stellar spectrum of the weighted association rule. At the same time, information entropy is used to determine stellar spectra with single attribute weights, and a compromise between the geometric mean and maximum is used to determine the weights in a multi-layer itemset. Its purpose is to highlight the important properties of the spectrum, when the waveform of the whole spectrum is taken into account. On this basis, the stellar spectra of weighted frequent patterns are extracted by using the weighted frequent pattern tree of stellar spectra, and the method of using a weighted frequent pattern as a tree structure is given. In the end, the experimental results show that using the method of stellar spectral data mining association rules is consistent with the main features of stellar spectral types.

Section 2 includes the basic concepts. Section 3 defines the stellar spectral weighted frequent itemset. Section 4 describes a method of stellar spectra weighted FP-tree construction and rule extraction. The last section gives the summary and further prospects.

## 2 BASIC CONCEPTS

Let DB be a database of transactions, and  $I = I_1, I_2, \dots, I_m$  be a set of  $m$  transaction itemsets in DB. Each transaction  $T$  in DB is a subset in the set of transactions  $I$ , that is  $T \subseteq I$ .

**Definition 1.** We refer to pattern  $P$  as a subset of  $I$ . Thus, we have  $P = I_1 \cap I_2 \cap \dots \cap I_k$ ,  $I_i \in I$  ( $i = 1, 2, \dots, k$ ), and the length of pattern  $P$  is the number of items in  $P$ . For example, the length of the above pattern  $P$  is  $k$ , because there are  $k$  items in  $P$ .

**Definition 2.** We say pattern  $P$  has a support degree  $0 \leq \sigma \leq 1$  with respect to database DB if and only if at least  $\sigma \times |DB|$  number of transactions in DB satisfy pattern  $P$ .

Let  $|t \rightarrow P|$  be the number of transactions that satisfies pattern  $P$ . We define support  $\sigma$  as Equation (1), where  $\sigma(P/DB)$  is the fraction of transactions in DB that satisfies the given pattern. Thus, we have

$$\sigma(P/DB) = |t \rightarrow P| / |DB|. \quad (1)$$

We say two patterns A and B are disjoint patterns if and only if their intersection is a null set, i.e.  $\{A_i\} \cap \{B_j\} = \emptyset$ , where  $A = A_1 \cap A_2 \cap \dots \cap A_k$ ,  $B = B_1 \cap B_2 \cap \dots \cap B_m$ .

**Definition 3.** Given two disjoint patterns (A, B) and an association rule  $A \Rightarrow B$  in DB, we define the confidence of the association rule as below

$$\psi(A \Rightarrow B/DB) = \sigma(A \cap B/DB) / \sigma(A/DB), \quad (2)$$

where  $\sigma(A \cap B/DB)$  and  $\sigma(A/DB)$  can be derived from the two support factors defined in Equation (1).

**Definition 4.** Let  $\sigma_{\min}$  be the threshold of the minimum support factor, then the  $k$ -frequent pattern set  $L_k$  and the non- $k$ -frequent pattern set  $L'_k$  in database DB is defined as

$$L_k = \{A_1 \cap A_2 \cap \dots \cap A_k \mid A_i \in I, \sigma(A_1 \cap A_2 \cap \dots \cap A_k/DB) \geq \sigma_{\min}\}, \quad (3)$$

$$L'_k = \{A_1 \cap A_2 \cap \dots \cap A_k \mid A_i \in I, \sigma(A_1 \cap A_2 \cap \dots \cap A_k/DB) < \sigma_{\min}\}. \quad (4)$$

To extract association rules from database DB, one has to specify the minimum support threshold  $\sigma_{\min}$  and the minimum confidence threshold  $\psi_{\min}$ . Thus, we need to search for any association rule (e.g.  $A \Rightarrow B$ ) that satisfies the following two conditions

$$\sigma(A \cap B/DB) \geq \sigma_{\min} \quad \text{and} \quad \psi(A \Rightarrow B/DB) \geq \psi_{\min}. \quad (5)$$

**Definition 5.** An FP-tree is a tree data structure that meets the following three conditions:

- Condition 1: An the FP-tree contains a root node marked as “NULL” (denoted as root). The children of the root node are item-prefix subtree sets. The tree also consists of a head table containing frequent items.
- Condition 2: Each node in an item-prefix subtree is made up of three components, namely, item-name, count and node-link. Item-name represents the name of the item; count denotes the number of transactions that are on any path leading to the node; and node-link is a link pointing to the next node with the same item-name value as in the FP-tree. If the next node does not exist, then the value of the node-link is set to null.
- Condition 3: Each item in the head table of frequent items is comprised of two components - item-name and head-of-node-link. Head-of-node-link is a pointer pointing to the head node of a singly linked list for nodes that have the same item-name value in the FP-tree.

### 3 STELLAR SPECTRAL WEIGHTED FREQUENT ITEMSETS

#### 3.1 Determining a Single Attribute Weight for Stellar Spectral Data

Let DB be a stellar spectral database, and  $I = \{I_1, I_2, \dots, I_m\}$  be the itemset containing 200 wavelength attributes and six physical/chemical attributes.  $W = (W_{I_1}, W_{I_2}, \dots, W_{I_m})$  is the weight

vector of the stellar spectral attributes. Due to the lack of a priori knowledge from experts and users, weight vector  $W$  is difficult to predetermine. The uncertainty of  $W$  can be measured with a probability function. Information entropy based on this probability can reflect the uncertainty in discrete messages and has been successfully applied in many fields. The uncertainty implies there is information content, and it is feasible to use the information entropy to describe the importance of a stellar spectral data attribute.

According to Shannon’s definition of information entropy, let  $P(y)$  be the probability of property  $y$  in the stellar spectral data set, then  $I(y) = -\log_2 P(y)$ , in which  $I(y)$  represents the amount of information in  $Y$ . The information entropy of  $Y$  is denoted as  $H(Y)$  which represents mean information content in the stellar spectral property  $Y$ . A property of  $H(Y)$  is that when  $H(Y)$  gradually increases, the degree of uncertainty in  $Y$  also increases and more information is hidden in it. Therefore, the amount of information entropy can indicate the general characteristics of  $Y$ .

$$H(X) = E[lb(1/P(x_i))] = - \sum_{i=1}^n P(x_i)lbP(x_i). \tag{6}$$

In general, stellar spectra are independent of each other and the amount of information in all spectral data of the same attribute can be cumulative. Information entropy represents the importance of attributes, so it is reasonable to use information entropy as an attribute weight. Let  $T_i \in D$  be a stellar spectrum, and  $P(m/T_i)$  be the probability of  $m$  when the spectrum is  $T_i$ , then the mean amount of information is expressed by  $H(m)$ .

$$H(m) = E[I(m/T_i)] = - \sum_{i=1}^n P(m/T_i)lbP(m/T_i). \tag{7}$$

Let  $I = \{I_1, I_2, \dots, I_m\}$  be the attribute set of the stellar spectral data, then the weight of the attribute  $I_i$  equals  $H(I_i)$ , that is weight  $(I_i)=H(I_i)$ .

### 3.2 Determining Multi-attribute Weights in Stellar Spectral Data

Let  $I = \{I_1, I_2, \dots, I_m\}$  be an attribute set for the stellar spectral database. Weight  $W_j$  given by information entropy is assigned to attribute  $i_j$  to show its importance, that is  $W(i_j) = w_j$ . The weight vector is denoted as  $W = \{W_1, W_2, \dots, W_m\}$ .

Let  $a = W_1 + W_2 + \dots + W_m$ , then  $\{W_1/a, W_2/a, \dots, W_m/a\}$  is the normalized attribute weight vector, that is  $W' = \{W'_1, W'_2, \dots, W'_m\}$  in which  $W'_i = W_i/a$  and  $W'_1 + W'_2 + \dots + W'_m = 1$ . So if  $Y$  is an itemset of stellar spectra, its weight can be defined as follows

$$W(Y) = \frac{1}{2} \left( \sqrt[k]{\prod_{i_j \in Y} W'_j} + \max_{i_j \in Y} \{W'_j\} \right). \tag{8}$$

This definition has resolved the issue that the value of level weighted support may be greater than one, and has highlighted key aspects by using the maximum weights. At the same time, the difference between items is reduced by using the geometric mean.

### 3.3 Stellar Spectrum Weighted Frequent Itemset

**Definition 6.** Let  $Y = \{I_1, I_2, \dots, I_m\}$  be a stellar spectral data itemset, then the weighted support of itemset  $Y$ ,  $\sigma_{wsup}$ , can be defined as follows

$$\sigma_{wsup}(Y) = W(Y) \times \sigma(Y) = \frac{1}{2} \left( \sqrt[k]{\prod_{i_j \in Y} W'_j} + \max_{i_j \in Y} \{W'_j\} \right) \times \sigma(Y), \tag{9}$$

where  $\sigma(Y)$  is the general support of  $Y$ .

**Definition 7.** Let  $X \Rightarrow Y$  be an association rule for the stellar spectral data, then its weighted support can be defined as  $\sigma_{\text{wsup}}(X \Rightarrow Y) = W(X \cup Y) \times \sigma(X \cup Y)$ , in which  $W(X \cup Y)$  is the weight of a stellar spectral data itemset  $(X \cup Y)$ . In accordance with the above formula,  $\sigma(X \cup Y)$  is the general support of  $(X \cup Y)$ .

**Definition 8.** Let  $X \Rightarrow Y$  be an association rule of the stellar spectral data, then its weighted confidence can be defined as

$$\psi_{\text{wsup}}(X \Rightarrow Y/DB) = \sigma_{\text{wsup}}(X \cap Y/DB) / \sigma_{\text{wsup}}(X/DB). \quad (10)$$

**Definition 9.** For data itemset  $Y$ , if  $\sigma_{\text{wsup}}(Y) \geq \sigma_{\text{min}}$ , then  $Y$  is a weighted frequent itemset of the stellar spectrum, in which  $\sigma_{\text{min}}$  is the minimum support threshold set by the user.

**Definition 10.** The association rules simultaneously satisfying the minimum support and minimum confidence threshold are known as the weighted association rules of the stellar spectrum.

Property: Let  $X = \{i_1, i_2, \dots, i_k\}$  be the stellar spectrum's weighted frequent itemset, in which  $k > 1$ ,  $Y \subseteq X$ , then if  $W(Y) \geq W(X)$ ,  $Y$  is certainly a weighted frequent itemset of the stellar spectrum; if  $W(Y) < W(X)$ ,  $Y$  is not a weighted frequent itemset of the stellar spectrum.

## 4 STELLAR SPECTRA WEIGHTED FREQUENT PATTERN TREE CONSTRUCTION AND RULE EXTRACTION

### 4.1 Weighted Frequent Pattern Tree Construction Algorithm

Similar to FP-tree construction, all information associated with the weighted frequent itemset should be stored in the Weighted Frequent Pattern tree (WFP-tree) of stellar spectral data. In WFP-tree construction, all 1-frequent and non-1-frequent pattern sets need to be gathered when the database is scanned for the first time, which is different from construction of the FP-tree. The WFP-tree of stellar spectral data can be built through traversing database  $D$  twice according to the following steps.

- (1) Scan through stellar spectral database  $D$  for the first time, gathering frequent length-1 patterns of sets, non-frequent length-1 patterns of sets, and their weighted supports, sort frequent length-1 patterns in descending order of the weighted supports and generate a frequent item table  $L$ ;
- (2) Create the root node of the stellar spectral data and denote it by "NULL;"
- (3) For each transaction  $T$  in  $D$ , sort frequent items of  $T$  in order of  $L$  and generate a new list of frequent items named  $T'$ , then update the WFP-tree according to the followings three steps:
  - (i) Search for a path that has the longest prefix matching  $T'$  in the WFP-tree;
  - (ii) The count of the node that is in the matching path is increased by one;
  - (iii) Search for the mismatching suffix in  $T'$ , and determine the node to which the last frequent item in the longest matching prefix is corresponding as the root node, then create child nodes successively in the WFP-tree and set the count value to 1.

### 4.2 Experimental Analysis

We implement a data-mining tool for the correlation analysis of stellar spectral data sets. The WFP-tree-construction algorithm described in Section 4.1 is incorporated in our correlation analysis module for stellar spectral data. The experiments are performed on a PC with an Intel Pentium IV 3.0 GHz processor and 512 MB of main memory; the operating system is Windows XP professional. We have fully implemented our correlation analysis module on top of Oracle 9i DBMS. Our stellar spectral data-mining tool and the correlation analysis module, in which the WFP-tree-construction algorithm is implemented, are developed with Visual C++ 6.0. 8315 SDSS stellar spectral data are selected as the data set (SDSS is one of the largest astronomical survey projects to date), the data which were publicly released by SDSS in June of 2007, including images and the associated database. The

**Table 1** Construction Time of WFP-Tree(s)

Data set	$\sigma_{\min} = 5\%$	$\sigma_{\min} = 3\%$	$\sigma_{\min} = 2\%$	$\sigma_{\min} = 1\%$
1000	143	175	201	287
2000	188	230	255	330
4000	198	257	289	395
8315	390	550	638	820

**Table 2** Numbers of Weighted Frequent Patterns

Data set	$\sigma_{\min} = 5\%$	$\sigma_{\min} = 3\%$	$\sigma_{\min} = 2\%$	$\sigma_{\min} = 1\%$
1000	435	530	850	1562
2000	489	750	1008	2017
4000	956	1218	1356	1980
8315	1007	1236	1478	2030

amount of data for figures is about 15TB. The SDSS is currently the largest astronomical data set that has been publicly released. The data are provided by National Astronomical Observatories, Chinese Academy of Sciences. We select 200 wavelengths and five physical/chemical properties including chemical composition, surface temperature, diameter, quality, luminosity and the like as attributes to represent stellar spectral data. The experimental data set is obtained after pretreatment on the source data set. The pretreatment is briefly summarized as follows: 1. Selecting six physical/chemical properties and 200 wavelengths which are sampled from 3810 Å to 7790 Å with an interval of 20 Å, which represent a set of attributes for each spectral data. 2. Discretizing spectral data according to the flow, peak width and shape of each wavelength. Experiments are divided into four groups with different data set sizes and minimum support thresholds. The data sets contain 1000, 2000, 4000 and 8315 objects and the minimum support threshold is respectively set to 1%, 2%, 3% and 5%.

The experimental steps are as follows. Firstly, the minimum weighted support is calculated according to the above Equation (8). Then the weighted frequent pattern tree is constructed using the WFP-tree-construction algorithm. The construction time is shown in Table 1.

Table 1 shows that for the same data set, the construction time of the WFP-tree gradually becomes longer with the decrement of the minimum support threshold  $\sigma_{\min}$ . When the minimum support decreases, its corresponding minimum weighted support decreases which leads to an increase in the number of weighted frequent patterns satisfying the conditions. This is why the time for constructing the tree has increased. It can also be seen from Table 1 that, for the same support threshold, the construction time of the WFP-tree gradually increases with the increase of data set size. If the data set increases in size, and the number of stellar spectra increases, the time to scan the database must also increase.

After traversing a constructed WFP-tree, one can extract frequent patterns from the stellar spectral data. To efficiently traverse a WFP-tree, we create an item-head table in which each item has a node link pointing to its location in the WFP-tree. Note that a conditional pattern base is a stellar spectral sub-database composed of the prefix path sets appearing in both the WFP-tree and the suffix pattern. Next, a conditional WFP-tree is built. The association rule mining process is performed recursively on the WFP-tree. Finally, the pattern growth is realized through the combination of frequent patterns generated by the suffix and the conditions associated with the WFP-tree. The numbers of weighted frequent patterns are described in Table 2.

Table 2 shows that, with the same data sets, the number of WFP-trees gradually increases, when the minimum support threshold  $\sigma_{\min}$  decreases. It can also be further explained that the construction time of the WFP-tree gradually increases when the minimum support is smaller. With the same minimum support, the number of weighted frequent patterns between the different data sets did not

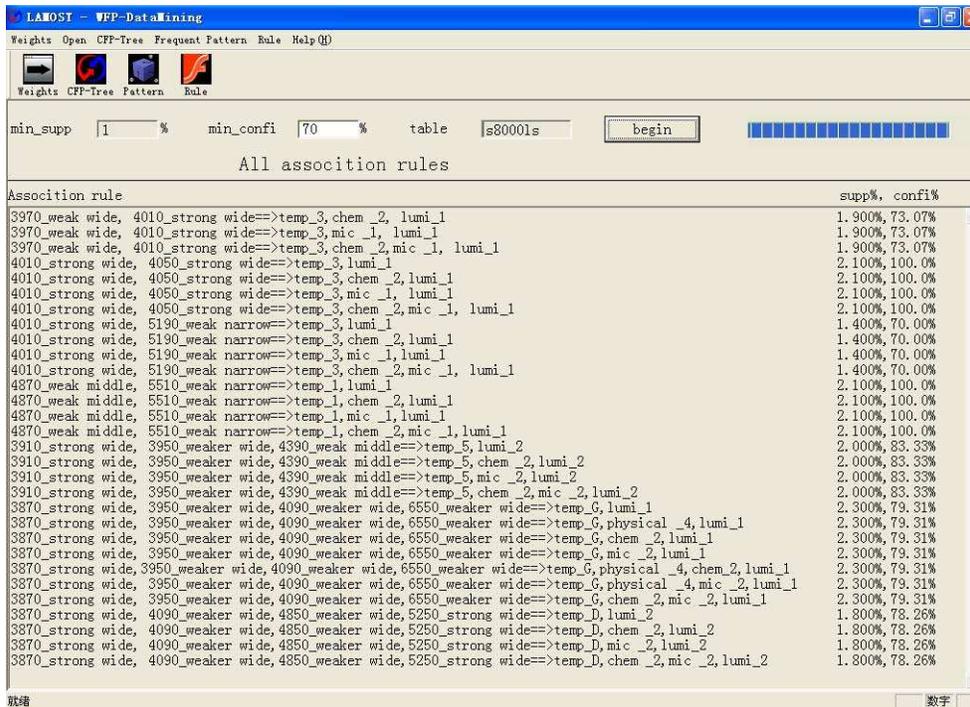


Fig. 1 Results of the association rule for a stellar spectrum.

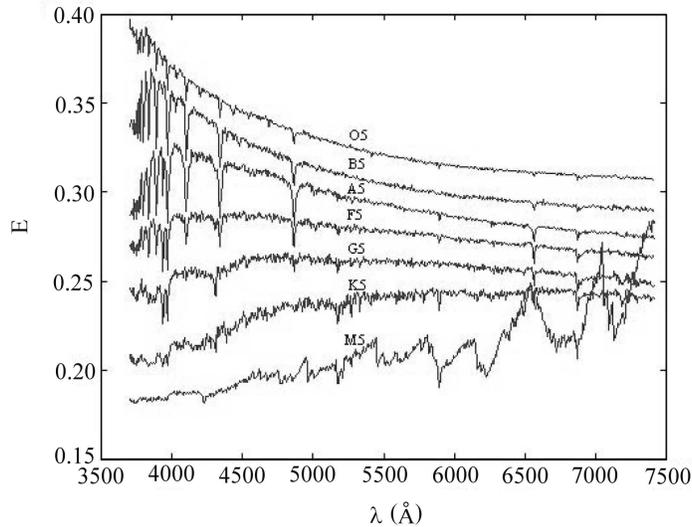
significantly change and did not have an obvious trend in terms of increase. So the above can show that the relation between the number of weighted frequent patterns and the size of the data sets is not obvious.

### 4.3 Weighed Association Rule Extraction

After the weighed frequent pattern is obtained, the weighed association rules are extracted through the following two steps. Firstly, the confidence of every spectrum's frequent pattern is calculated by following Definition 3, and some of the spectrum's frequent patterns are filtered since their confidence is lower than the minimum confidence set by users. Secondly, every frequent pattern that remains is divided into two parts: the feature set of attributes after discretizing the stellar spectra and the attribute set of physical/chemical properties after discretizing. Let  $L$  be a frequent pattern,  $s$  and  $s'$  be two parts obtained through all non-empty subsets of  $L$ , and  $\psi_{\min}$  be the threshold of minimum confidence, for instance, the association rule " $s \Rightarrow s'$ " is output under the condition of  $\sigma((s + s')/DB)/\sigma(s/DB) \geq \psi_{\min}$ . Its results are shown in Figure 1.

Figure 1 shows the results of the correlation analysis of the stellar spectral data. In this experiment, the minimum support threshold  $\sigma_{\min}$  is set to 1% and the minimum confidence threshold  $\psi_{\min}$  is set to 70%. Interesting conclusions can be drawn from the association rules. For example, one of the association rules is 3870\_ strong wide, 4090\_ weaker wide, 4850\_ weaker wide, 5250\_ strong wide  $\Rightarrow$  temperature\_ D, chemical\_ 2, microturbulence\_ 2, luminosity\_ 2 (1.800%, 78.26%). This association rule can be explained as follows:

- (1) There exist very strong and very wide peaks at the wavelength of 3870;
- (2) There exist weaker and very wide peaks at the wavelength of 4090;



**Fig. 2** Stellar spectral classification.

- (3) The peak at the wavelength of 4850 is weaker and very wide;
- (4) If the peak at the wavelength of 5250 is very strong and very wide, then the temperature range of this spectrum is from 7500 to 8300, the chemical abundance range is from  $-3$  to  $-0.5$ , the microturbulence value is 2, and the luminosity range is from 0 to 1.1. The support of the association rule is 1.8%, and the confidence is 78.26%.

Comparing the rules generated by our celestial-spectral data-mining tool with the properties of stellar spectral wavelengths in Figure 2, one can conclude that it is basically similar to the properties of spectral type A. Thus, it shows that the weighted association rule mining used to extract knowledge was successful. From these association rules, we can see that some of them are known by astronomy experts, which validates the correctness, and others are unknown which can help astronomy experts to discover new properties.

## 5 CONCLUSIONS

According to the item's different importance in the stellar spectral data set, the concepts of stellar spectral weighted itemsets and weighted association rules are introduced, and the weight of a single property in the stellar spectrum is determined by information entropy; at the same time a compromise between the geometric mean and maximum is used to determine the multi-layer itemset weights. On this basis, the stellar spectra of a weighted frequent pattern tree are constructed, and a method is presented to mine the weighted frequent patterns and association rules for stellar spectra.

**Acknowledgements** We are very grateful to an anonymous referee for many useful comments and suggestions, which allowed us to substantially improve the manuscript. This work is supported by the National Natural Science Foundation of China (Grant Nos. 61073145, 41140027 and 41210104028), the Shanxi Province Natural Science Foundation (No. 2012011011-4), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi, China (No. 20121011), and the Shanxi Province Science Foundation for Youths (No. 2012021015-4).

## References

- Armitage, P. J. 2010, *Astrophysics of Planet Formation* (Cambridge: Cambridge Univ. Press)
- Armitage, P. J. 2011, *ARA&A*, 49, 195
- Bai, L., Guo, P., & Hu, Z.-Y. 2005, *ChJAA* (Chin. J. Astron. Astrophys.), 5, 203
- Bailer-Jones, C. A. L., Irwin, M., & von Hippel, T. 1998, *MNRAS*, 298, 361
- Connolly, A. J., Szalay, A. S., Bershad, M. A., Kinney, A. L., & Calzetti, D. 1995, *AJ*, 110, 1071
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA* (Research in Astronomy and Astrophysics), 12, 1197
- Folkes, S. R., Lahav, O., & Maddox, S. J. 1996, *MNRAS*, 283, 651
- Galaz, G., & de Lapparent, V. 1998, *A&A*, 332, 459
- Garcia, P. J. V. 2011, *Physical Processes in Circumstellar Disks around Young Stars* (Chicago, IL: University of Chicago Press)
- Hartmann, L. 2001, *Accretion Processes in Star Formation* (Cambridge: Cambridge Univ. Press)
- Katz, D., Soubiran, C., Cayrel, R., Adda, M., & Cautain, R. 1998, *A&A*, 338, 151
- Qin, D.-M., Guo, P., Hu, Z.-Y., & Zhao, Y.-H. 2003, *ChJAA* (Chin. J. Astron. Astrophys.), 3, 277
- Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, *MNRAS*, 370, 141
- Reipurth, B., Jewitt, D., & Keil, K. 2007, *Protostars and Planets V* (Tucson, AZ: University of Arizona Press)
- Soubiran, C., Katz, D., & Cayrel, R. 1998, *A&AS*, 133, 221
- Stahler, S. W., & Palla, F. 2005, *The Formation of Stars* (Weinheim: Wiley-VCH)
- von Hippel, T., Storrie-Lombardi, L. J., Storrie-Lombardi, M. C., & Irwin, M. J. 1994, *MNRAS*, 269, 97
- Ward-Thompson, D., & Whitworth, A. P. 2011, *An Introduction to Star Formation* (Cambridge: Cambridge Univ. Press)
- Williams, J. P., & Cieza, L. A. 2011, *ARA&A*, 49, 67
- Wu, X.-B., Jia, Z.-D., Chen, Z.-Y., et al. 2010, *RAA* (Research in Astronomy and Astrophysics), 10, 745
- Xu, X., Yang, J. F., Wu, F. C., & Zhao, Y. H. 2006, *Spectroscopy and Spectral Analysis*, 26, 1960
- Yang, J. F., Xu, X., Wu, F. C., & Zhao, Y. H. 2007, *Spectroscopy and Spectral Analysis*, 27, 602
- Zaritsky, D., Zabludoff, A. I., & Willick, J. A. 1995, *AJ*, 110, 1602
- Zhang, Y., & Zhao, Y. 2003, *PASP*, 115, 1006
- Zhang, Y., & Zhao, Y. 2004, *A&A*, 422, 1113
- Zhao, Q., Sun, J. Z., Xiao, J., et al. 2010, *Application Research of Computers*, 27, 3322
- Zhao, X. J., Zhang, J. F., & Cai, J. H. 2008, *Spectroscopy and Spectral Analysis*, 28, 2996