
The application characteristics of traditional Chinese medical science treatment on vertigo based on data mining Apriori algorithm

Miao Wang

Shanghai University of Traditional Chinese Medicine,
No. 1200, Cailun Road, Shanghai 201210, China
Email: miawang2004@hotmail.com

Li Chen, Yanjun Huang, Lei Zhang
and Zihao Zhang

Department of Electronic Engineering,
Fudan University,
No. 220, Handan Road, Shanghai 200433, China
Email: li_chen12@fudan.edu.cn
Email: yjhuang12@fudan.edu.cn
Email: zhangl11@fudan.edu.cn
Email: zhangzh11@fudan.edu.cn

Jie Ding

Shanghai University of Traditional Chinese Medicine,
No. 1200, Cailun Road, Shanghai 201210, China
Email: dj_dingdang@sina.com

Huiliang Shang*

Department of Electronic Engineering,
Fudan University,
No. 220, Handan Road, Shanghai 200433, China
Email: shanghl@fudan.edu.cn

*Corresponding author

Abstract: As the mechanism of the effective traditional Chinese medical science treatment remains ambiguous, researchers are seeking to analyse the therapies using traditional Chinese medicine (TCM) with advanced techniques. In this paper, we take advantage of data mining techniques with the software SPSS (Statistical Product and Service Solutions), in particular Apriori algorithm for Association Rules Mining (ARM), which discovers relationships by iteratively analysing large data sets of TCM medical records. After analysing the database, we are likely to reveal the interrelationships among TCM syndromes, clinical symptoms, and curative Chinese herbal medicine. Through verification, Apriori proves effective in helping traditional Chinese doctors decide what to prescribe when facing various patients with vertigo. The data used in the analysis are collected from our specially designed mobile app where well-trained doctors can upload their recent diagnostic results. With the increasing amount of diagnosis data, the effectiveness of our diagnoses is constantly improving.

Keywords: data mining; Apriori algorithm; traditional Chinese medicine; vertigo; mobile app.

Reference to this paper should be made as follows: Wang, M., Chen, L., Huang, Y., Zhang, L., Zhang, Z., Ding, J. and Shang, H. (2015) 'The application characteristics of traditional Chinese medical science treatment on vertigo based on data mining Apriori algorithm', *Int. J. Wireless and Mobile Computing*, Vol. 9, No. 4, pp.349–354.

Biographical notes: Miao Wang, PhD, is an assistant researcher in Shanghai University of Traditional Chinese Medicine. Her research interest includes standardisation of traditional Chinese medicine, informatisation of traditional Chinese medicine, pathogenesis and aetiology of traditional Chinese medicine. She had undertaken five national or regional projects, including one special research foundation of Young Teacher in Shanghai. One of her ten publications is included in SCI index, and four of them are included in EI.

Li Chen is expected to receive his BSc degree from the Department of Electronic Engineering of Fudan University, Shanghai, China, in 2016. He is the author of many published articles. His research interests include image processing, computer vision, data-mining and traditional Chinese medicine informatics.

YanJun Huang is expected to receive his BSc degree from the Department of Electronic Engineering of Fudan University, Shanghai, China, in 2016. He is the author of many published articles. His research interests include image processing, computer vision, data-mining and traditional Chinese medicine informatics.

Lei Zhang received the BSc degree in Electronic Engineering from Fudan University, Shanghai, China, in 2015. He is currently working toward the MS degree in the School of Computer Science, University of Southern California. His research interests include data mining, traditional Chinese medicine and data analysis.

Zihao Zhang received the BSc degree in Electronic Engineering from Fudan University, Shanghai, China, in 2015. He is currently working toward the MSc degree in the Henry Samueli School of Engineering and Applied Science, University of California - Los Angeles. His research interests include data mining, traditional Chinese medicine and data analysis.

Jie Ding received doctor degree from Beijing University of Traditional Chinese Medicine in 2007. She works at Shanghai University of Traditional Chinese Medicine, where she is engaged in scientific research. Her major research direction lies on syndrome research. So far she has taken part in some projects and published seven academic papers and four translated monographs.

Huiliang Shang received the BSc, MSc and PhD degrees from Fudan University, Shanghai, China, in 1999, 2002 and 2009, respectively. He is an Associate Professor at the Department of Electronic Engineering, Fudan University. He is the author of many published articles. His research has been concerned with circuit theory, network topology, industrial control, measurement and intelligent control and traditional Chinese medicine informatics.

This paper is a revised and expanded version of a paper entitled 'The application characteristics of traditional Chinese medical science treatment on headache based on data-mining Apriori algorithm' presented at the 'IEEE International Conference on Bioinformatics & Biomedicine', Belfast in Northern Ireland, UK, 2–5 November 2014.

1 Introduction

Vertigo is a subtype of dizziness, defined as an illusion of movement, usually a sense of rotation. Dizziness that is severe enough to bring an individual into a clinic is remarkably common, affecting approximately 30% of individuals aged 18–79 during their lifetime (Berlinger, 2011). Many primary care physicians find the devotion to dizzy patients challenging because of the vagueness of patients' specific symptoms, and the wide variety of possible diagnoses, ranging from benignancy to seriousness (Chan, 2009).

Medication remains the mainstay therapy for vertigo, but medicines usually have side effects that sometimes even endanger life security, which limits the drug efficacy.

Traditional Chinese Medicine (TCM) has been widely used in clinical practice for over 3000 years. Veteran TCM doctors classify vertigo syndromes into several types, and have summed up various TCM treatment options towards each category, which have definite curative effects and advantages. However, the mechanism of TCM on vertigo has not been worked out. Therefore, we use a data mining technique to search for the laws of TCM on vertigo, which may break the limitation of depending on personal experience (Fayyad et al., 1996).

Data mining, a computational process of discovering patterns, is usually used to find out potential and valuable information in a large scale of data (Geng and Du, 2014). Today, it is widely applied in various fields. Liu et al. (2012)

successfully discovered the association rules by Apriori algorithm between aetiology and TCM, syndromes and TCM, as well as symptoms and TCM in influenza. Zhou et al. (2012) built a disease risk model of osteoporosis based on TCM symptoms and western medicine risk factors with several data mining techniques. Luo et al. (2013) obtained association rules among diseases, TCM syndromes, formulae, and *Ligusticum Wallichii* by text mining techniques.

In this paper, we apply Apriori algorithm to the analysis of a series of medical records on vertigo from famous TCM doctors in order to find out the relationships between TCM syndromes, symptoms, and Chinese herbs. There is a certain novelty value in this approach as the approach, compared to traditional methods, is able to greatly enhance the efficiency of searching for the association rules. Specifically speaking, in Section 2, the details of Apriori algorithm are shown. In Section 3, we apply Apriori to deal with TCM medical records. Sections 4 and 5 are the verification and improvement of the results.

2 Materials and methods

2.1 Basic principles of Apriori algorithm

Frequent patterns, such as itemsets, subsequences, and substructures, are patterns that occur frequently in data.

Apriori is an iteration algorithm based on the prior knowledge of frequent itemsets, mining subsequent frequent itemsets for Boolean association rules. Frequent itemsets should be extended one item at a time, and the major purposes of Apriori algorithm are to generate candidate k -itemsets from frequent $(k-1)$ -itemsets, as well as to find frequent k -itemsets from candidate k -itemsets.

There are four parameters that can be used to describe the basic principles of Apriori algorithm.

- 1 *Support*: a description of the frequency and effectiveness of rules. Minsup represents the minimum support.
- 2 *Confidence*: a description of the intensity and credibility of rules. Minconf represents the minimum confidence.
- 3 *Lift*: a description of the extent of the mutual impact among items.
- 4 *Frequent itemset*: a description of the itemset that occurs frequently in data.

In short, Apriori algorithm is used to help to identify the association rules from the database, meeting the demands of the minimum support Minsup and the minimum confidence Minconf given by users (Xu, 2012).

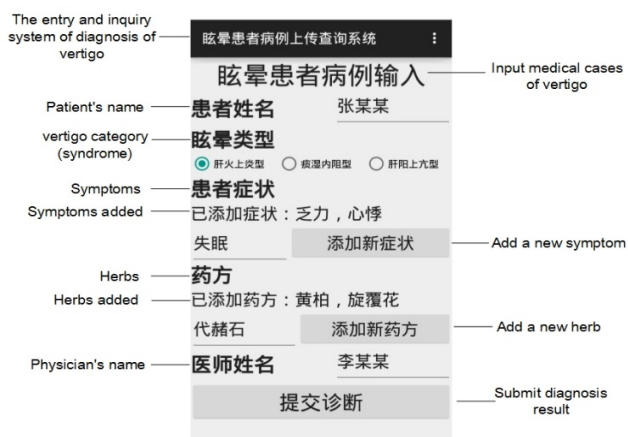
2.2 Application of Apriori algorithm for association rules mining (ARM)

2.2.1 Establish a database

2.2.1.1 Medical cases selection

With the prevalence of mobile phones, using a mobile phone app as an ancillary tool during diagnosis process has become a novel idea (Setyono et al., 2014; Ray and Biswas, 2014). We have created a mobile app to collect case data from various famous TCM doctors in different hospitals, who were instructed to use the app to upload their daily diagnostic results of vertigo. The data format is preset in the app so that we can get a clear table of all medical records with patients' and doctors' names, vertigo categories, symptoms, and prescriptions. Figure 1 shows the data entry interface of the specially designed app.

Figure 1 The data entry interface



2.2.1.2 Data standardisation

Each medical record is mainly divided into three parts: TCM syndrome, clinical symptoms, and Chinese herbs. Owing to the discrepancies in expressions of different medical records, which have an adverse impact on ARM, it is necessary to standardise all the medical expressions.

There are several rules for data standardisation.

- Replace all the synonyms of symptoms with one standard word corresponding to them.
- Split the polysemous word into several univocal words.
- Standardise and normalise the names of Chinese herbs to avoid duplication.

Then, we input the table into the statistical software SPSS to establish a database for Association Rules Mining (ARM) and analysis.

2.2.2 Application of Apriori algorithm

According to the database, we use Apriori algorithm to reveal the relationships among TCM syndromes, clinical symptoms, and Chinese herbs.

First of all, we set the minimum support and the minimum confidence (Minsup = 10% and Minconf = 20%) after adjustment for rational and accurate results. After that, we apply Apriori algorithm and find out the frequent itemsets which meet the requirements of the minimum support and confidence. In this process, we can decide the antecedents and consequents according to the demand of analysis. Next, by setting different antecedents and consequents, and comparing the parameters, we can get strong association rules of all these items.

With these rules, we are able to discover the differences among all kinds of vertigo such as vertigo due to Liver Fire Flaring Up (LFF), vertigo due to Upper Hyperactivity of Liver Yang (UHL), etc. We can also understand which kinds of symptoms are associated with vertigo, and which kinds of herbs are commonly used in vertigo treatment. It will not only bring profound guiding significance to the vertigo study, but also promote the TCM study if we apply this method in other disease fields (Purohit et al., 2012).

3 Results and discussion

3.1 Association rules between TCM syndromes (specific vertigo categories) and Chinese herbs

Figure 2 Association rules between TCM syndromes and Chinese herbs

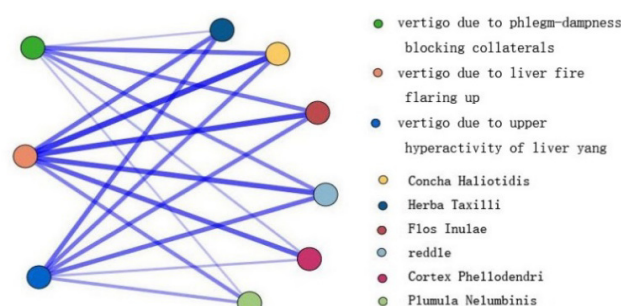


Table 1 Association rules between TCM syndromes and Chinese herbs

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
88.889	22.857/25.714	1.004	PBC → Concha Haliotidis
83.333	21.429/25.714	0.956	PBC → Flos Inulae
88.889	22.857/25.714	1.037	UHL → reddle
88.889	22.857/25.714	1.27	UHL → Herba Taxilli
80.952	24.286/30	1.232	LFF → Cortex Phellodendri
66.667	20/30	1.085	LFF → Plumula Nelumbinis

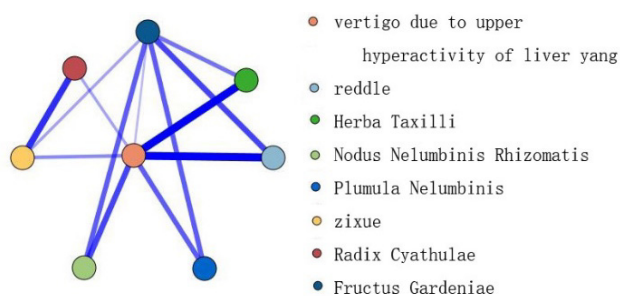
The results in Table 1 indicate that there are close associations between specific vertigo categories, such as vertigo due to LFF, vertigo due to UHL, vertigo due to Phlegm-Dampness Blocking Collaterals (PBC), and several kinds of Chinese herbs.

Each category has its own fundamental remedy or remedies, which is or are conducive to the selection and decision of relatively appropriate herbs.

Figure 2 shows the association network between three specific vertigo categories and some types of Chinese herbs listed in Table 1. The points with three different colours on the left side stand for the vertigo categories, while the points on the right side stand for different Chinese herbs. The lines signify their connections, reflecting the connection strength through their thickness, which means the thicker the line, the stronger the connection. (This property is applicable to all the following figures.) According to Figure 2, it is proved that each specific vertigo category has corresponding and effective Chinese herbs commonly used in the field of TCM.

3.2 Association rules between UHL and Chinese herbs

Figure 3 Association rules between UHL and Chinese herbs



Considering the confidence and support degree, we take cases related to UHL as examples (see Table 2). And Figure 3 reveals the correlation of the chosen vertigo category with some Chinese herbs. Obviously, UHL, represented by the central point in Figure 3, has strong connections with a few Chinese herbs (with the average of the confidence degree exceeding 60%) of which both reddle and Herba Taxilli enjoy the same highest confidence degree (88.889%). In other words, the two Chinese herbs mentioned above have an extraordinary effect on treating this vertigo category. In

view of various statistical results, especially the confidence and support degree, Herba Taxilli, one of the main herbs for UHL, is selected for further analysis. Correlations between adjuvants and Herba Taxilli are shown in Table 3.

Table 2 Association rules between UHL and Chinese herbs

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
88.889	22.857/25.714	1.037	UHL → reddle
88.889	22.857/25.714	1.27	UHL → Herba Taxilli
72.222	18.571/25.714	1.149	UHL → Nodus Nelumbinis Rhizomatis
66.667	17.143/25.714	1.085	UHL → Plumula Nelumbinis
55.556	14.286/25.714	1.341	UHL → zixue
50	12.857/25.714	1.094	UHL → Radix Cyathulae
44.444	11.429/25.714	2.074	UHL → Fructus Gardeniae

3.3 Association rules between Herba Taxilli (HT) and its adjuvants

Figure 4 Association rules between HT and its adjuvants

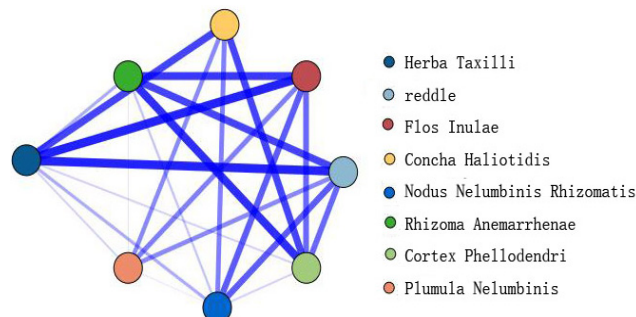


Table 3 Association rules between HT and its adjuvants

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
95.918	67.143/70	1.119	HT → reddle
95.918	67.143/70	1.101	HT → Flos Inulae
89.796	62.857/70	1.014	HT → Concha Haliotidis
71.429	50/70	1.136	HT → Nodus Nelumbinis Rhizomatis
71.429	50/70	0.962	HT → Rhizoma Anemarrhenae
65.306	45.714/70	0.994	HT → Cortex Phellodendri
65.306	45.714/70	1.063	HT → Plumula Nelumbinis

Figure 4 corresponds to Table 3, and gives a vivid description of the table. It is easy to recognise that reddle, Flos Inulae, and Concha Haliotidis are closely related to HT (with 95.918%, 95.918%, and 89.796% for the confidence degree,

respectively), which then demonstrates their feasibility of being the adjuvants and the possibility of utilisation when HT is used to treat a certain kind of vertigo.

3.4 Association rules between TCM syndromes and symptoms

Figure 5 Association rules between TCM syndromes and symptoms

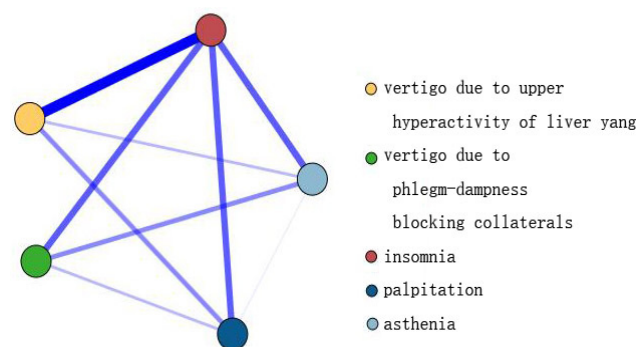


Table 4 Association rules between TCM syndromes and symptoms

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
38.889	10/25.714	1.237	UHL → insomnia
22.222	5.714/25.714	1.556	UHL → palpitation
27.778	7.143/25.714	0.884	PBC → insomnia
22.222	5.714/25.714	1.197	PBC → asthenia

Patients with diverse vertigo categories usually show distinct clinical symptoms through observation. By data mining, two symptoms are listed for UHL and PBC separately (see Table 4). The statistics are not as ideal as anticipated because the confidence degrees are relatively lower in comparison with those in previous tables. However, in the light of its reference value, Table 4 still assists us to figure out that insomnia is a key symptom to both vertigo categories. In Figure 5, two relevant lines are shown as the thickest, making the above conclusion reasonable.

3.5 Association rules between symptoms and Chinese herbs

Figure 6 Association rules between symptoms and Chinese herbs

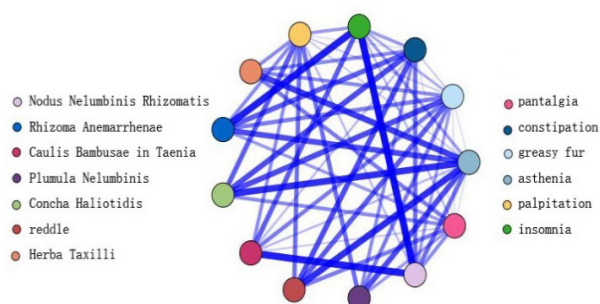


Table 5 Association rules between symptoms and Chinese herbs

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
85.714	8.571/10	1.364	Pantalgia → Nodus Nelumbinis Rhizomatis
81.818	12.857/15.714	1.101	Constipation → Rhizoma Anemarrhenae
70	10/14.286	2.042	Greasy fur → Caulis Bambusae in Taenia
76.923	14.286/18.571	1.252	Asthenia → Plumula Nelumbinis
90	12.857/14.286	1.016	Palpitation → Concha Haliotidis
81.818	25.714/31.429	0.955	Insomnia → reddle
81.818	25.714/31.429	1.169	Insomnia → Herba Taxilli

Table 5 records a variety of statistics mirroring the connections between symptoms and Chinese herbs. To make the table more representative, we pick six symptoms completely different from each other, namely pantalgia, constipation, greasy fur, asthenia, palpitation, and insomnia. The table indicates that each symptom has at least one type of highly related Chinese herb, which is witnessed by Figure 6 since the thicker the line, the stronger the connection.

4 Verification and conclusion

By data mining from modern clinical literature on treatment for vertigo, we draw basic rules between commonly used traditional Chinese herbs, symptoms, and TCM syndromes. Meanwhile, these rules are in accord with diagnostic results and principles of the clinical practice. With the view of verifying the feasibility of Apriori algorithm as well as the validity of the data, we pick the data in Table 5 to conduct the verification test.

In Table 5, both reddle and Herba Taxilli have intimate connections with insomnia, indicating their probability of being used to treat insomnia. Then, it can be easily figured out that insomnia is one of cardinal symptoms for UHL in Table 4. Tables 1 and 2 tell us reddle and Herba Taxilli, of which the strong correlation is witnessed in Table 3, play extremely important roles in the therapeutic process of this kind of vertigo.

Table 6 explains if UHL and insomnia exist simultaneously, the above-mentioned two kinds of Chinese herbs are able to be utilised for treatment owing to their effectiveness and pertinence. Therefore, the results are valid in theory and have been confirmed to be accurate by several TCM doctors.

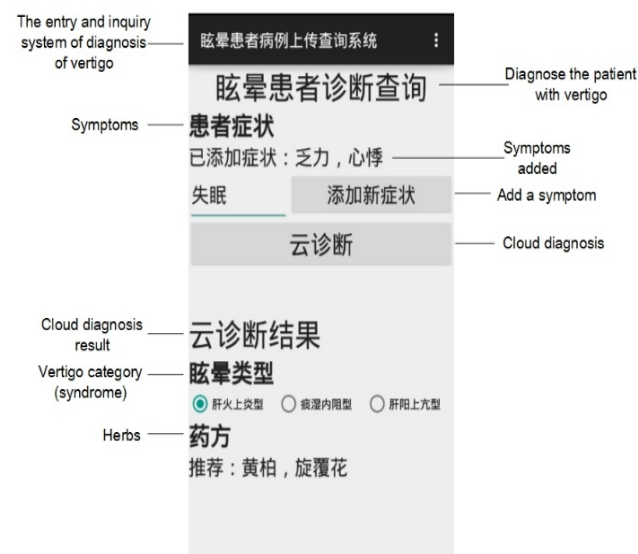
In addition, we make deeper ARM with two antecedents and one consequent (see Table 6). One kind of vertigo category (or certain symptoms) and one kind of Chinese herb are the antecedents. These two kinds of association rules are attempts at constructing a platform to help inexperienced TCM doctors to make diagnoses since if the vertigo category (or the symptom) and a type of Chinese herb are confirmed,

and another reasonable Chinese herb can be considered for treatment due to the strong associations. Also, the association rules about vertigo are available to every app user. Figure 7 shows the relevant inquiry interface.

Table 6 Association rules between TCM syndromes & symptoms and Chinese herbs (Rows 1–2), TCM syndromes & adjuvants and Chinese herbs (Rows 3–4), symptoms & adjuvants and Chinese herbs (Rows 5–6)

Confidence (%)	Rule support (%) / support (%)	Lift	Rules
85.714	8.571/10	1	UHL & insomnia → reddle
85.714	8.571/10	1.224	UHL & insomnia → Herba Taxilli
87.5	10/11.429	1.801	LFF & fresh lotus leaf → Flos Inulae
85.714	8.571/10	2.069	UHL & Radix Gentianae → zixue
77.778	10/12.857	2.865	Insomnia & Talcum → Poria
71.429	7.143/10	1.724	Greasy fur & Caulis Bambusae in Taenia → zixue

Figure 7 The inquiry interface



The database located in our server is constantly updated with an increasing amount of data uploaded daily by various famous and well-trained TCM doctors. Therefore, the accuracy of the data mining can be improved with time.

5 Improvements of Apriori algorithm

Apriori algorithm has the disadvantage of inadequate efficiency, which is caused by a great number of frequent itemsets and repeated scanning in database. In our research, the disadvantage is not obvious as the data scale is relatively small. However, when applied to a larger database, the time consumed will increase sharply. Thus, optimising the

efficiency of Apriori algorithm will contribute to larger-scale and more complex medical cases analysis.

We also tested FP-growth algorithm in finding frequent itemsets. With the similar correctness, the improved algorithm largely saved the calculation time. In the future research where database is too large to be calculated by Apriori algorithm, FP-growth algorithm may be a better solution.

Acknowledgements

The authors would like to thank the handling editor and all the reviewers for their constructive comments. This work was supported by National Natural Science Foundation of China (Grant No. 61301028), Natural Science Foundation of Shanghai (Grant No. 13ZR1402900), and Specialised Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20120071120016).

References

Berlinger, N.T. (2011) ‘Meniere’s disease: new concepts, new treatments’, *Minnesota Medicine*, Vol. 94, No. 11, pp.33–36.

Chan, Y. (2009) ‘Differential diagnosis of dizziness’, *Current Opinion in Otolaryngology & Head and Neck Surgery*, Vol. 17, No. 3, pp.200–203.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) ‘From data mining to knowledge discovery in databases’, *AI Magazine*, Vol. 17, No. 3, pp.37–54.

Geng, Z. and Du, J. (2014) ‘Research on web page classification-based core characteristics and web structure’, *International Journal of Wireless & Mobile Computing*, Vol. 7, No. 3, pp.253–257.

Liu, Y., Liu, Q., Zheng, D. and Deng, Q. (2012) ‘Application and improvement discussion about Apriori algorithm of association rules mining in cases mining of influenza treated by contemporary famous old Chinese medicine’, *2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, 4–7 October, IEEE, Philadelphia, PA, USA, pp.316–322.

Luo, D., Xiao, C., Sun, S. and Zheng, G. (2013) ‘Searching association rules of traditional Chinese medicine on Ligusticum Wallichii by text mining’, *2013 IEEE International Conference on Bioinformatics and Biomedicine*, 18–21 December, IEEE, Shanghai, China, pp.162–167.

Purohit, N., Purohit, S. and Purohit, R.K. (2012) ‘Data mining, applications and knowledge discovery’, *International Journal of Advanced Computer Research*, Vol. 2, No. 6, p.458.

Ray, S. and Biswas, G.P. (2014) ‘Design of an efficient mobile health system for achieving hipaa privacy-security regulations’, *International Journal of Wireless & Mobile Computing*, Vol. 7, No. 4, pp.378–387.

Setyono, A., Alam, M.J. and Eswaran, C. (2014) ‘Mobile telemedicine system application for telediagnosis using multimedia messaging service technology’, *International Journal of Wireless & Mobile Computing*, Vol. 7, No. 4, pp.348–361.

Xu, H. (2012) ‘Association rules data mining methods’, *Science Technology and Engineering*, Vol. 20, No. 1, pp.60–63.

Zhou, X.H., Li, S.L., Tian, F., Cai, B.J., Xie, Y.M., Pei, Y., Kang, S., Fan, M. and Li, J.P. (2012) ‘Building a disease risk model of osteoporosis based on traditional Chinese medicine symptoms and western medicine risk factors’, *Statistics in Medicine*, Vol. 31, No. 7, pp.643–652.