

基于条件随机域的禩国维名老中医医案挖掘分析

江启煜^{1*}, 郑美思¹, 李红毅², 梁家芬²

(1. 广州中医药大学, 广州 510405; 2. 广东省中医院, 广州 510120)

[摘要] **目的:**基于条件随机域模型对禩国维名老中医临床医案进行挖掘分析。**方法:**基于条件随机域设计一种新的临床医案文本挖掘信息模型,利用该模型对国医大师禩国维医案进行临床术语信息的自动识别提取,并使用关联规则算法进行数据挖掘分析。**结果:**基于条件随机域构建的文本信息挖掘模型对不同临床术语类型,不同的病种识别效果良好,其中在不同术语类型的识别中对“中医诊断术语”的 *F*-测度值最高,达到 89.87%,在不同病种的识别中对“湿疹”的 *F*-测度值最高,达到 83.77%。对其中禩国维名老中医治疗荨麻疹的经验进行挖掘,发现在治法上多采取益气固表、祛风止痒、除湿解毒;常用以治疗的二联药对有黄芪-白术、黄芪-防风、白术-防风等,其中白术-防风的置信度(1.00)和支持度(0.83)最高;常用的三联药物有黄芪-紫苏叶-白术等,常见的药物症状关联有地黄-风团、防风-瘙痒等,其中地黄-风团的置信度(1.00)和支持度(0.97)最高。**结论:**该研究基于条件随机域构建了一种新的文本挖掘信息模型,该模型有利于高效整理和挖掘名老中医临床医案中的经验与学术思想,对名老中医的学术思想和经验传承具有重要意义。

[关键词] 文本挖掘; 条件随机域; 关联分析; 名老中医

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2017)09-0208-06

[doi] 10.13422/j.cnki.syfjx.2017090208

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20170214.1605.046.html>

[网络出版时间] 2017-02-14 16:05

Text Mining and Analysis of Medical Records from Famous Chinese Medicine Expert XUAN Guo-wei Based on Conditional Random Fields

JIANG Qi-yu^{1*}, ZHENG Mei-si¹, LI Hong-yi², LIANG Jia-fen²

(1. Guangzhou University of Chinese Medicine, Guangzhou 510405, China;

2. Guangdong Province Hospital of Traditional Chinese Medical, Guangzhou 510120, China)

[Abstract] **Objective:** To explore and analyze the clinical medical records from famous Chinese medicine expert XUAN Guo-wei based on the conditional random field. **Method:** A new text mining model for the medical records were designed based on conditional random field. Then this model was used to automatically recognize and extract the clinical terms information in the medical records from famous Chinese medicine expert XUAN Guo-wei, and association rules algorithm was used for data mining analysis. **Result:** The text information mining model which was constructed based on conditional random fields, can effectively identify different clinical terms and different disease types; in the results of recognizing different types of terms, ‘traditional Chinese medicine (TCM) diagnosis terms’ had the highest *F*-measure value, reaching 89.87%; in the results of recognizing different disease types, ‘eczema’ had the highest *F*-measure value, reaching 83.77%. Based on XUAN Guo-wei’s clinical experience in the treatment of urticaria: professor Xuan often used the methods of tonifying Qi and strengthening exterior, dispelling wind and arresting itching, as well as removing dampness and detoxifying for urticaria patients; his commonly used herbal pairs included Astragali Radix-Atractylodis Macrocephalae Rhizoma, Astragali Radix-Saposhnikoviae Radix, Atractylodis Macrocephalae Rhizoma-Saposhnikoviae Radix, etc, and the confidence degree

[收稿日期] 20161206(015)

[基金项目] 广东省科技计划项目(2014A020221034)

[通讯作者] * 江启煜, 讲师, 从事中医药数据挖掘研究, Tel:13660765137, E-mail:jiangqiyu@gzucm.edu.cn

and support degree of *Atractylodis Macrocephalae Rhizoma-Saposhnikoviae Radix* were highest, respectively reaching 1.00 and 0.83; his commonly used triple herbal combinations included *Astragali Radix-Perillae Folium-Atractylodis Macrocephalae Rhizoma*, etc. Common TCM-symptom association pairs included *Rehmanniae Radix-wheals*, *Saposhnikoviae Radix-itching*, etc, and the confidence degree and support degree of *Rehmanniae Radix-wheals* were highest, respectively reaching 1.00 and 0.97. **Conclusion:** A new text information mining model was constructed in this study based on conditional random fields, which was conducive to efficiently excavate and succeed the clinical medical experience and academic thoughts of famous old Chinese Medicine experts.

[**Key words**] text mining; conditional random field; association rules analysis; famous old Chinese medicine experts

临床医案是名老中医表达临床经验与学术思想的主要载体,如何高效整理和挖掘临床医案中的经验与学术思想是一个值得研究的主题。但在目前数据挖掘方法中,无论使用哪一种数据挖掘算法与模型,挖掘的数据必须是已经进行人工整理(结构化)的表格数据,而并不能直接对自由书写(非结构化)的医案或电子病历进行挖掘。如何利用计算机科学技术,直接从自由书写(非结构化)的医案中智能识别并自动提取这些中医药术语数据^[1-3],并进行挖掘分析,是人们期待突破的难题。

本研究基于条件随机域模型与算法^[4-5],设计一种新的文本信息挖掘模型,把国医大师禔国维临床医案中的中医药各种名词术语信息的类别识别问题,转化为语言标记序列识别问题。使用建立的条件随机域模型,通过医案训练数据集,智能学习国医大师禔国维临床医案中的术语序列标记分布规律,并使用最终确定的模型与参数,自动识别提取测试数据集中的症状体征术语、病名术语(中医诊断)、证型名称术语(中医诊断)、治则治法术语(中医)、方剂名称术语、中药名称术语等各类的医案关键词,使用关联规则算法进行数据挖掘分析。

1 资料与方法

1.1 医案数据集的建立 获取国医大师禔国维医案 400 余份,并产生其相应的电子文档医案集。随机抽取其中的 50% 医案作为条件随机域模型的建模数据(训练数据集),用于模型的机器统计学习过程;另一半作为待识别医案(测试数据集),模型训练好后可智能识别这些医案中的临床术语。对建模数据集进行分词与字标记预处理。定义标记的符号,例如 ZM 代表证名, BM 代表病名。起始标记法采用“BIO”标记^[6],即 B 代表某个类别词开始的首字符, I 代表非首字符, O 代表不在某个类别中的字符。例如“阴虚阳亢”,第 1 个字标记为 B-ZM,其余 3 个字标记为 I-ZM。按照上述的方法,标记所有的

建模数据集。

1.2 基于条件随机域识别临床经验术语 条件随机域(CRF)是一种研究观察序列与状态序列无向图的模型^[7],它可以通过样本训练学习已知标记的序列分布,并对新样本的未知序列进行识别标记。假设病案中的若干个字组成的词串为观察序列 x ,词串所蕴含的标记为状态序列 y ,根据 CRF 模型,观察序列 x 对应状态序列 y 的总体可能性应该由两部分共同决定,一是观察序列 x 当前位置 X_i 与其对应的状态序列 y 的当前位置 Y_i 的概率;二是观察序列 x 当前位置 X_i 与其对应的状态序列 y 的当前位置 Y_i 以及前一个位置 Y_{i-1} 的概率,因此,对于给定的观察序列 x ,关于状态序列 y 的势函数积为:

$$E(X, Y) = \sum_j \lambda_j f_j(y_{i-1}, y_i, x, i) + \sum_k \mu_k S_k(y_i, x, i) \quad (1)$$

其中 $\lambda_j f_j(y_{i-1}, y_i, x, i)$ 表示观察序列 x 对状态序列 y 在位置 i 以及 $i-1$ 的特征函数, $\mu_k S_k(y_i, x, i)$ 表示观察序列 x 对状态序列 y 在位置 i 的状态特征函数。这两个函数均取值 0 或 1,定义特征函数 $f_j(y_{i-1}, y_i, x, i)$ 统一表示这 2 个函数,在给定观察序列 x 时,状态序列 y 对它的概率为:

$$P(y | x, \lambda) = \frac{\exp[\sum_j \lambda_j F_j(y, x)]}{\sum_y \exp[\sum_j \lambda_j F_j(y, x)]} \quad (2)$$

因此预测标记序列就是在给定观察序列 x 时求出使 $P(y | x, \lambda)$ 最大的状态序列 y ,即:

$$\max[P(y | x, \lambda)] = \operatorname{argmax}(y) \sum_j \lambda_j F_j(y, x) \quad (3)$$

对于某一确定的 λ 值,如果逐项计算所有的可能标记序列的概率会比较复杂耗时,可以使用 Viterbi 优化算法求解式(3)^[8]。而对于最佳 λ 值的确定,可以通过测试集数据样本,使用最大似然法(Maximum Likelihood)求解^[9]。

对待识别的医案,使用上述条件随机域模型进

行术语识别标记,并使用下述 3 个参数:召回率(R),准确率(P)和 F -测度(F)进行术语识别效果的评价。设能被模型识别的中医药名词术语中,识别准确的数量为 A ,识别不准确的数量为 B 。设模型需要识别的中医药名词术语总数为 D ,识别出的数量为 C 。则这 3 个参数的计算式为:

$$R = \frac{C}{D + C}, P = \frac{A}{A + B}, F = \frac{2PR}{P + R} \quad (4)$$

1.3 使用关联规则算法对识别的临床经验进行数据挖掘 对本研究识别的术语数据中的荨麻疹医案进行挖掘分析^[10-12],使用建立的文本信息挖掘模型自动识别提取荨麻疹医案中的症状体征术语,证型名称术语(中医诊断),治则治法术语(中医),中药名称术语,并自动输入到 Office Excel 2013 表格中,以医案序号为行,以医案中的各类术语为列,以数字“1”表示病案中出现相应的术语,否则为“0”,以此建立禔国维教授治疗荨麻疹的医案数据集。运用 Excel 软件中的 FREQUENCY 函数,计算各症状、证型、治法的频数,统计频率。将数据集中各 Excel 文件导入 Weka 软件,使用其中的关联规则挖掘功能进行数据挖掘^[13-14],选择 Apriori 算法^[15-17],度量类型采用置信度。参数设置如下: class Index 设置为 0, out put Item Sets 设置为 true, lower Bound Minsupport, min Metric, num Rules 等参数在挖掘结果表格后说明,其余参数默认。

2 结果

2.1 基于条件随机域的临床经验术语识别效果

2.1.1 不同术语类型的识别效果 不同术语类型的识别效果如图 1 所示。可知在各类型临床术语中,准确率(P)从 82.41% 到 92.01%,说明该模型可准确地识别病案中的各类术语类型; F -测度(F)波动于 79%~86%,说明该模型对各术语类型识别效果良好;召回率(R)相对低,但维持 76% 以上,说明该模型可识别的各术语类型个数较多但有所不足。在各术语类型中,中医诊断术语的准确率(P),召回率(R)和 F -测度(F)分别为 92.01%,87.83%,89.87%,均高于其他术语类型,说明相比于其他术语类型,该模型可最准确地识别中医诊断术语,对中医诊断术语的可识别个数最多,对中医诊断术语的识别效果最好;症状和体征术语的召回率(R),准确率(P)和 F -测度(F)均低于其他术语类型,说明该模型对症状和体征术语的识别个数、识别准确度、识别效果均不如其他术语类型。虽然该模型用于识别中医诊断术语效果最好,但从总体来说,该模型识别

各术语类型准确度高,可识别的各术语类型及其个数较多,识别效果良好。

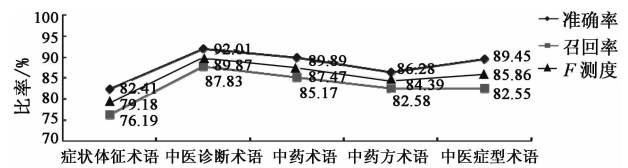


图 1 条件随机域模型对不同类别医案术语的识别效果

Fig. 1 Recognition effects of CRF models on different types of medical terms

2.1.2 不同疾病类型的识别效果 不同疾病类型的识别效果如图 2 所示。可知各疾病类型中, F -测度(F)范围从 83.31% 到 85.53%,可见识别效果从高到低的疾病类型为湿疹,过敏性皮炎,痤疮,全身性红斑狼疮,银屑病;准确率(P)从 86.62% 到 89.08%,各种类型疾病的平均准确率达到 88.012%;召回率(R)则在 83.31% 以上。在不同疾病类型中,湿疹(eczema)的准确率(P), F -测度(F)最高,分别达 89.08%,83.77%,说明该模型对湿疹的识别准确性最强,识别效果最好;银屑病的召回率(R),准确率(P)和 F -测度(F)最低,分别为 80.24%,86.62%,83.31%,说明该模型可识别的银屑病数目、识别准确度及识别效果均差于其他疾病。综合不同疾病类型的各参数指标可知,该模型可识别的不同疾病类型数量多,准确度高,识别效果良好,具有实际应用意义。

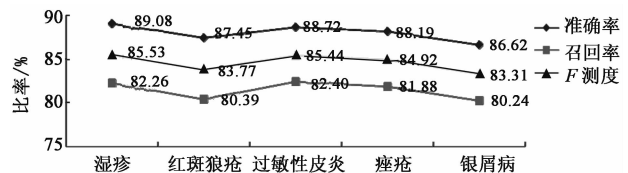


图 2 条件随机域模型对不同疾病医案术语的识别效果

Fig. 2 Recognition effects of CRF models on different disease type terms

2.2 关联数据挖掘结果

2.2.1 症状 在所有荨麻疹医案中,共有 44 个症状,各症状频率:风团、瘙痒 > 皮肤划痕征 > 脉细、脉弦 > 舌红 > 苔黄 > 抓痕 > 苔白 > 眠差,剩余症状粉刺、脉沉、脉缓、咽喉不适、大便不畅、自觉疲劳、头油腻、脱发均只出现 1 例。见表 1。

2.2.2 证型 荨麻疹各证型出现频率从大到小依次为卫表不固 > 湿热内蕴 > 风邪外袭 > 气血亏虚 > 气滞血瘀 > 风热相搏、肺肾不足 > 邪郁少阳、脾虚。见表 2。

表 1 禩国维治疗荨麻疹医案证状频数

Table 1 Symptom frequencies in Urticaria medical cases treated by XUAN Guo-wei

症状	频数/次	频率/%	症状	频数/次	频率/%
风团	58	100	胃脘不适	6	10
瘙痒	58	100	纳差	6	10
皮肤划痕征	57	98	斑丘疹	5	9
脉细	46	79	舌暗	5	9
脉弦	43	74	口苦	5	9
苔黄	36	62	脉滑	5	9
舌红	36	62	舌边齿痕	4	7
抓痕	19	33	大便干硬	4	7
苔白	19	33	盗汗	4	7
眠差	17	29	咳嗽	4	7
口干	15	26	黄痰	3	5
自汗	11	19	腰酸背痛	3	5
畏风	10	17	服药欲呕	3	5
丘疹	9	16	血痂	3	5
舌淡	8	14	脉弱	2	3
红斑	7	12	皮肤干燥	2	3
苔腻	7	12	动则气喘	2	3
便溏	7	12			

表 2 禩国维治疗荨麻疹医案证型频数

Table 2 TCM syndromes frequencies in Urticaria medical cases treated by XUAN Guo-wei

证候	频数/次	频率/%	证候	频数/次	频率/%
卫表不固	42	72	风热相搏	2	3
湿热内蕴	22	38	肺肾不足	2	3
风邪外袭	9	16	邪郁少阳	1	2
气血亏虚	4	7	脾虚	1	2
气滞血瘀	4	7			

2.2.3 治法 荨麻疹各治法出现频率由大到小依次为:祛风 > 益气固表 > 止痒 > 除湿 > 清热 > 养血 > 活血、解毒、补肺敛肾 > 健脾、和解少阳。见表 3。

2.2.4 中药对关联 荨麻疹治疗中药关联如表 4 所示。可知二联药对中黄芪->白术、黄芪->防风、白术->防风的置信度达 1, 为最高值, 其中白术->防风的支持度最高, 为 0.83; 黄芪->紫苏叶、黄芪->徐长卿、黄芪->牡丹皮、黄芪->地黄的支持度均高达 0.98; 可见白术与防风关联性最佳, 黄芪与白术、防风、紫苏叶、徐长卿、牡丹皮、地黄关联性良好。三联药对中, 黄芪->白术 防风的支持度为 0.80, 置信度为 1, 均为最高值, 提示黄芪、白术、防风 3 药的关联性最强; 此外, 由置信度来看, 黄芪,

表 3 禩国维治疗荨麻疹医案治法频数

Table 3 Therapy frequencies in Urticaria medical cases treated by XUAN Guo-wei

治法	频数/次	频率/%	治法	频数/次	频率/%
祛风	51	88	活血	2	3
益气固表	45	76	解毒	2	3
止痒	34	59	补肾敛肺	2	3
除湿	24	41	和解少阳	1	2
清热	18	31	健脾	1	2
养血	5	9			

表 4 禩国维治疗荨麻疹医案药对关联

Table 4 Associations of medicines in Urticaria medical cases treated by XUAN Guo-wei

关联规则	支持度	置信度
黄芪->白术	0.79	1
黄芪->防风	0.79	1
黄芪->紫苏叶	0.78	0.98
黄芪->徐长卿	0.78	0.98
黄芪->牡丹皮	0.74	0.98
黄芪->地黄	0.78	0.98
黄芪->甘草	0.72	0.98
白术->防风	0.83	1
黄芪->白术, 防风	0.80	1
黄芪, 紫苏叶->白术	0.76	1
黄芪, 徐长卿->白术	0.76	1
黄芪, 地黄->白术	0.78	1
黄芪->白术, 紫苏叶	0.78	0.98
黄芪->白术, 徐长卿	0.78	0.98
黄芪->白术, 地黄	0.78	0.98

注: 支持度 ≥ 0.5 , 置信度 ≥ 0.5 , nmuRules = 100; 仅选取挖掘结果中的二联药对、三联药对。

紫苏叶->白术, 黄芪, 徐长卿->白术, 黄芪, 地黄->白术的置信度均为 1, 提示 3 组三联规则存在良好的关联性。

2.2.5 症状与中药关联 症状与中药关联如表 5 所示。可知各二联规则的置信度均为 1, 其中地黄->风团、地黄->瘙痒的支持度为 0.97, 为最高值, 其余二联规则的支持度均在 0.9 以上, 可见地黄、防风、徐长卿、紫苏叶与风团、瘙痒关联性良好; 临床上荨麻疹之风团、瘙痒症状可用地黄、防风、徐长卿、紫苏叶组合加以治疗。

三联规则中, 以牡丹皮->皮肤划痕征 (+) 风团, 甘草->风团瘙痒, 牡丹皮, 地黄->风团的置信度为 1, 为最高值; 徐长卿, 地黄->皮肤划痕征 (+) 的支持度为 0.93, 为最高值, 可见牡丹皮对皮肤划痕征 (+), 风团症状, 甘草对风团瘙痒症状, 牡丹皮、地黄对风团症状关联性良好。

表 5 禩国维治疗荨麻疹医案症状与中药关联

Table 5 Associations between medicines and symptoms in Urticaria medical cases treated by XUAN Guo-wei

关联规则	支持度	置信度
地黄->风团	0.97	1
地黄->瘙痒	0.97	1
防风->风团	0.95	1
防风->瘙痒	0.95	1
徐长卿->风团	0.93	1
徐长卿->瘙痒	0.93	1
紫苏叶->风团	0.93	1
紫苏叶->瘙痒	0.93	1
牡丹皮->皮肤划痕征(+)风团	0.90	1
牡丹皮,地黄->风团	0.90	1
甘草->风团,瘙痒	0.90	1
徐长卿,地黄->皮肤划痕征(+)	0.93	0.98
紫苏叶,风团->皮肤划痕征(+)	0.92	0.98
紫苏叶->风团,皮肤划痕征(+)	0.92	0.98

注:支持度 ≥ 0.5 , 置信度 ≥ 0.5 , numRules = 100, 仅选取挖掘结果中的二联、三联规则。

3 讨论

3.1 条件随机域模型能够较好地识别临床经验术语 医案中的中医药各种临床术语信息的识别与提取问题,其本质上是语言标记序列识别问题。而序列标记的机器学习模型主要包括贝叶斯理论(Naive Bayes),隐马尔科夫模型(Hidden Markov Model),最大熵模型(Maximum Entropy),以及条件随机域(Conditional Random Field)等^[9]。这些序列标记模型通过观察序列与状态序列之间的有向图或无向图关联,及其关联权重矩阵与状态特征函数,使用数学算法学习训练数据集中的序列联合概率,并使用已训练好的模型,根据测试数据集中的信息观察序列,预测或识别其实际的状态序列。相对于隐马尔科夫模型和最大熵模型,条件随机域在信息序列计算时权重偏置更小,序列特征灵活性更高。对于自由的书写病案文本,每一个词都可被视为有一个内含的标记序列,则词串是观察序列,词串的标记序列是其对应的状态序列。因此可以用上述模型研究病案中词串的分布与识别规律,并使用算法加以挖掘。对于复杂变化的中医药语言而言,条件随机域模型更能充分反映中医药名词之间的联系和内涵。

3.2 禩国维荨麻疹经验挖掘分析

3.2.1 症状分析 从症状的频数分析结果可知,风团、瘙痒、皮肤划痕征阳性、抓痕、眠差、舌红、苔黄或苔白、脉弦细出现频率较高,为荨麻疹患者的普遍症状。荨麻疹是一种临床较为常见的皮肤黏膜过敏性

疾病。禩国维教授认为^[18],荨麻疹属虚者多因血虚生风、或因产后受风、或因皮疹反复加之风邪外袭;属实者多为饮食不节,脾失健运,外感风邪,风热湿邪困阻肌肤而成。风邪善行数变,为本病致病关键,风袭则起风瘙瘾疹,舌红苔黄提示风热病机;苔白可为风寒袭于肌肤腠理;脉弦细提示患者可能为情志抑郁或产后受风。

3.2.2 证型分析 从证型的频数分析结果可知,卫表不固证和湿热内蕴证是荨麻疹病案中发病的主要证候类型。一般而言,若机体卫表不固,常致外邪侵袭,而六淫邪气中风邪为首;风邪袭表可致多种病变,起风瘙瘾疹是风邪的致病表现之一。若患者饮食失节,过食辛辣及膏粱厚味,致脾胃失调,酿湿生热,湿热内蕴,化热动风,亦可致病。实际临床上,禩教授认为荨麻疹的证型可分以下 6 种:卫气不固型、胃肠失调型、气滞血瘀型、肺肾不足型、冲任失调型、气血两虚型^[18]。其中卫气不固型、气滞血瘀型、肺肾不足型、气血两虚型均可见于病案中。

3.2.3 治法分析 从治法的频数分析结果可知,益气固表、祛风、止痒、除湿、清热为病案中出现的较频的治法。此与荨麻疹多见卫表不固、湿热内蕴证型及致病关键因素“风邪”相关。禩教授治疗荨麻疹一病时,根据临床具体情况也采用养血活血、补肾敛肺、和解少阳、健脾益气等治法。

由此可见,禩教授临床治疗荨麻疹时常祛邪扶正兼施,坚固机体正气又兼祛风除湿止痒,清热解毒。

3.2.4 中药对关联分析 从中药的关联规则分析结果可知,禩国维教授治疗荨麻疹的药对为黄芪与白术、防风、紫苏叶、牡丹皮、徐长卿、地黄相配,或白术与防风相配,或黄芪、白术与防风、紫苏叶、徐长卿、地黄之间的组合。可见荨麻疹之治疗药对或为益气药之间相配伍,或益气药合祛风药同用,或益气药与祛湿药相伍,或益气药配清热凉血药,或益气固表药、祛风除湿药兼施。

其中关联性最强的三联药对为黄芪-白术-防风。黄芪、白术、防风 3 药相伍,共成益气固表之玉屏风散。《本草新编·黄芪篇》云“黄芪得防风而功愈大”;防风与黄芪、白术相配,祛邪而不伤正;黄芪、白术与防风相伍,固表而不留邪;3 药合用,固表气,实肌腠,兼疏风邪,补中寓散,散不伤正,补不留邪,共奏益气固表实卫之功。此外,黄芪、白术合用紫苏叶,则益气兼疏风除湿;黄芪、白术合用徐长卿,则益气兼祛风止痒;黄芪、白术合用地黄,则益气兼

清热凉血。

3.2.5 症状与中药关联分析 风团、瘙痒常用治疗药物为地黄、防风、徐长卿、紫苏叶、甘草,其中地黄与风团、瘙痒关联性最强;皮肤划痕征常用牡丹皮、徐长卿、地黄等,其中以牡丹皮最佳。荨麻疹之风团、瘙痒等证,其关键致病因素为风邪,或风邪袭表,或血虚生风,或机体阴血不足而风动;或湿热或肝郁化热而热盛动风。故用以防风解表祛风。防风素有“风中之润剂,治风之通药”之美名;《本草纲目》更盛赞其为除风祛湿仙药,可见凡风邪为患之疾病,防风必为首选之药。徐长卿善祛风化湿、止痛止痒;现代药理研究表明,徐长卿中的丹皮酚具有抗炎、抗过敏、免疫调节等药理作用,可用治皮肤科之荨麻疹、湿疹等疾患^[19]。地黄入血分,可清热凉血,滋阴生津,善针对荨麻疹之阴血虚而虚热内生、引动内风的情况。紫苏叶可解表散寒,行气宽中,此处用紫苏叶,源于其较强的抗过敏性^[20]。甘草性味甘平,可清热解毒、补脾益气,具有抗过敏和抑制抗体产生的作用,同时具有抗炎和皮质激素样作用^[21]。牡丹皮可清热凉血,活血化瘀,与地黄共寓“治风先治血,血行风自灭”之意。由上可知,地黄、防风、徐长卿、紫苏叶、甘草、牡丹皮均是治疗麻疹主证的有效药物。

4 结论

本研究基于条件随机域模型与算法,设计的禩国维名老中医临床医案的文本信息挖掘模型,能够较好地识别医案中的临床经验术语序列标记分布规律;基于关联规则挖掘出禩国维教授治疗荨麻疹的临床经验:临床上荨麻疹多见卫表不固、湿热内蕴证型,因此治疗上禩国维教授多采取益气固表、祛风止痒、祛风除湿、清热除湿解毒等方法,根据临床实际情况尚采取养血活血、补肾敛肺、和解少阳、健脾益气等治法,祛邪扶正兼顾;常用到的中药包括防风、紫苏叶、徐长卿、黄芪、白术、蝉蜕等,其常用二联药对为黄芪、白术、黄芪、防风;常用三联药对为黄芪、白术、防风,黄芪、白术、紫苏叶,黄芪、白术、徐长卿,黄芪、白术、地黄;多为益气药配伍益气药、祛湿药、祛风药、清热凉血药,或益气固表药、祛风除湿药兼施;此外针对患者的风团、瘙痒症状,常用地黄、防风、徐长卿、紫苏叶等中药。本研究提出构建了老中医临床医案的文本挖掘一种新的信息模型,有利于高效整理和挖掘名老中医临床医案中的经验与学术思想,对名老中医的学术思想和经验传承具有重要意义。

[参考文献]

- [1] 周雪忠. 文本挖掘在中医药中的若干应用研究[D]. 杭州:浙江大学,2004.
- [2] 赵宇平,王慧,杨光,等. 基于文本挖掘技术探索青蒿的药理作用规律[J]. 中国中药杂志,2016,41(16): 3072-3077.
- [3] 刘建华,张智维,徐健,等. 自动术语识别——对科技文献进行文本挖掘的重要技术方法[J]. 现代图书情报技术,2008 168(8):12-17.
- [4] 杨永贵. 中文信息抽取关键技术研究及实现[D]. 北京:北京邮电大学,2008.
- [5] Sager N. Natural language information processing: a computer grammar of English and its applications [M]. Massachusetts: Addison Wesley Publishing Co., 1981: 158-169.
- [6] 刘豹,张桂平,蔡东风. 基于统计和规则相结合的科技术语自动抽取研究[J]. 计算机工程与应用,2008, 44(23):147-150.
- [7] 张榕. 术语定义抽取、聚类与术语识别研究[D]. 北京:北京语言大学,2006.
- [8] 施水才,王锴,韩铧铎,等. 基于条件随机场的领域术语识别研究[J]. 计算机工程与应用,2013,49(10): 147-149.
- [9] 郑强,刘奇军,王正华,等. 生物医学命名实体识别的研究与进展[J]. 计算机应用研究,2010,27(3): 811-815.
- [10] 金宗泽,冯亚丽,纪博,等. 大数据分析中的关联挖掘[J]. 计算机与数字工程,2014,42(10):1924-1928.
- [11] 李芸. 数据挖掘中关联规则挖掘方法的研究及应用[D]. 西安:西安电子科技大学,2007.
- [12] 程丹丹. 关联分析在中医数据挖掘中的应用研究[D]. 北京:北京交通大学,2007.
- [13] 司晓梅. 数据挖掘中关联规则的研究及应用[D]. 武汉:武汉理工大学,2007.
- [14] 王卉,张红君. 关联挖掘研究综述[J]. 软件导刊, 2009,8(3):7-8.
- [15] 吴少莹. 多维关联规则挖掘算法研究[D]. 天津:天津理工大学,2008.
- [16] 马猛. 关联挖掘的若干研究[D]. 合肥:安徽大学,2005.
- [17] 左艇,范欣生,田硕,等. 基于关联规则方法对不同地区乌头组反药的临床调查研究和配伍特点分析[J]. 中国中药杂志,2015,40(6):1198-1202.
- [18] 范瑞强,谢长才. 禩国维教授治疗慢性荨麻疹的经验[J]. 中医药研究,1999,15(5):27-28.
- [19] 周国茂,沈琴,高进. 徐长卿抗变态反应作用的实验研究[J]. 中国中西医结合皮肤性病学期杂志,2004, 3(2):126-128.
- [20] 黄丽,韦保耀. 紫苏叶抗过敏有效成分的研究[J]. 食品科技,2005,38(5):90-93.
- [21] 陈红. 甘草药理作用概述[J]. 海峡药学,2005,17(4):37-41

[责任编辑 邹晓翠]