

COVID-19 中国诊疗方案的机器学习智能分析方法



蒲晓蓉^{1,2}, 陈柯成^{1,2}, 刘军池^{1,2}, 文进³, 郑尚维³, 李鸿浩³

1. 电子科技大学 计算机科学与工程学院(成都 611731)
2. 电子科技大学 大数据中心 健康医疗大数据研究所(成都 611731)
3. 四川大学华西医院 医院管理研究所(成都 610041)

【摘要】 2019年底暴发的新型冠状病毒肺炎(COVID-19)疫情是人类史上一次重大突发公共卫生事件。中国医学工作者在短时间内,经历了对该未知病毒的逐步认识、证据积累和临床实践。截至目前,中国国家卫生健康委员会在数十天内密集发布了七个版本的《新型冠状病毒感染的肺炎诊疗方案》(简称《诊疗方案》)。然而,快速准确地比较各版本的异同和掌握新版本的重点对临床医护人员和非专业人员来说存在一定困难。本文提出一种基于机器学习的计算机辅助智能分析方法,对文本主题进行无监督学习,自动分析不同版本《诊疗方案》的异同,主动给医护人员推送新版本的关注重点,降低《诊疗方案》解读的专业难度,提高非专业人员对诊疗方案的认识水平。实验证明,与人工解读方式相比较,本文方法能自动计算文本主题,实现主题的精准匹配,准确率达100%,并可自动生成关键词和语句级别的解读报告,实现《诊疗方案》的计算机自动智能解读。

【关键词】 COVID-19; 新型冠状病毒; 人工智能; 机器学习; 诊疗方案; 生物医学工程

Machine learning-based method for interpreting the guidelines of the diagnosis and treatment of COVID-19

PU Xiaorong^{1,2}, CHEN Kecheng^{1,2}, LIU Junchi^{1,2}, WEN Jin³, ZHNENG Shangwei³, LI Honghao³

1. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, P.R.China
2. Health Big Data Institute of Big Data Center, University of Electronic Science and Technology of China, Chengdu 611731, P.R.China
3. Institute of Hospital Management, West China Hospital, Sichuan University, Chengdu 610041, P.R.China

Corresponding author: LI Honghao, Email: 243424375@qq.com

【Abstract】 The outbreak of pneumonia caused by novel coronavirus (COVID-19) at the end of 2019 was a major public health emergency in human history. In a short period of time, Chinese medical workers have experienced the gradual understanding, evidence accumulation and clinical practice of the unknown virus. So far, National Health Commission of the People's Republic of China has issued seven trial versions of the "Guidelines for the Diagnosis and Treatment of COVID-19". However, it is difficult for clinicians and laymen to quickly and accurately distinguish the similarities and differences among the different versions and locate the key points of the new version. This paper reports a computer-aided intelligent analysis method based on machine learning, which can automatically analyze the similarities and differences of different treatment plans, present the focus of the new version to doctors, reduce the difficulty in interpreting the "diagnosis and treatment plan" for the professional, and help the general public better understand the professional knowledge of medicine. Experimental results show that this method can achieve the topic prediction and matching of the new version of the program text through unsupervised learning of the previous versions of the program topic with an accuracy of 100%. It enables the computer interpretation of "diagnosis and treatment plan" automatically and intelligently.

【Key words】 COVID-19; novel coronavirus; artificial intelligence; machine learning; diagnosis and treatment plan; biomedical engineering

DOI: 10.7507/1001-5515.202003045

基金项目: 四川大学华西医院新型冠状病毒肺炎疫情科技攻关项目(HX-2019-nCoV-023); 四川省科技厅科技计划项目基金(2020YFS0119, 2019JDKP0067); 四川省科技厅软科学研究项目基金(2019JDR0144)

通信作者: 李鸿浩, Email: 243424375@qq.com

引言

自 2019 年底以来全球相继暴发新型冠状病毒肺炎 (COVID-19)，其感染速度之快、范围之广属历史罕见。中国医学工作者在短时间内，经历了对该未知病毒的逐步认识、证据积累和临床实践。截至目前，中国国家卫生健康委员会组织相关专家研究、制定并发布了七个《新型冠状病毒感染的肺炎诊疗方案》(以下简称《诊疗方案》) 试行版本^[1-7]。该系列版本内容差异较大，主要内容涵盖冠状病毒病原学特点、临床特点、病例定义、鉴别诊断、病例的发现与报告、治疗、解除隔离和出院标准、转运原则和医院感染控制等项目。《诊疗方案》是全国各医疗机构医务人员对新型冠状病毒肺炎患者进行诊断和救治的指导性文件。该系列诊疗方案的高度专业性和不同版本之间存在的诸多差异，导致临床医护人员和非专业人员在短时间内快速、准确解读并掌握相关信息存在一定困难。因此，有必要研制计算机辅助的智能文本理解方法，自动区分不同版本的差异，并标识新版本的关注重点。

本课题针对国家卫生健康委员会正式发布的《诊疗方案》七个版本，基于认知计算理论^[8]，提出一种基于机器学习的隐狄利克雷分配 (latent Dirichlet allocation, LDA) 模型^[9]的无监督聚类方法，自动挖掘不同版本《诊疗方案》的主题，进行版本间异同分析比较。

1 方法

1.1 总体架构

本文所提出的基于机器学习的计算机辅助智能分析方法的整体架构如图 1 所示。其中，对文本主题分布的学习使用若干已有诊疗方案，作为训练集，用于训练 LDA 模型。将待分析诊疗方案作为测试样本，基于训练好的 LDA 模型，预测和匹配文本主题。最后，进行文本差异性分析，标注文本之间存在的差异。

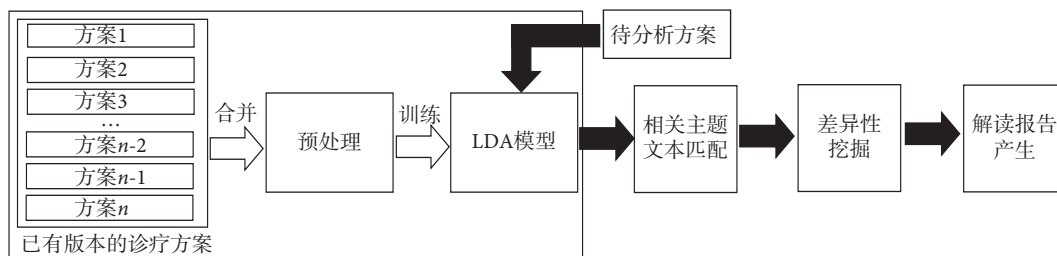


图 1 文本解析的架构

Fig.1 The framework of the text interpretation

1.2 诊疗方案主题生成的 LDA 模型

首先，假设每一版《诊疗方案》由 N 个单词组成，每一个词可以表示成 one-hot 的向量形式，则一个版本的《诊疗方案》可以表示为：

$$W = (w_1, w_2, \dots, w_N), \quad (1)$$

其中 w_n 是《诊疗方案》中的第 n 个词的向量表示。假设已有 M 个版本的《诊疗方案》，这些文本可以看作是进行主题分析的语料库 D ，可表示为：

$$D = \{W_1, W_2, \dots, W_M\}, \quad (2)$$

《诊疗方案》的文本可表示为潜在的 K 个主题上的随机混合，每个主题又被单词的分布所表征。LDA 假设语料库 D 中的每个文本 W 都有以下生成过程：

(1) 对于每一版《诊疗方案》 $d \in D$ ，得到方案 d 的主题分布参数 θ_d 。

(2) 对于每个主题 $k \in K$ ，得到主题 k 上词的多项式分布 ϕ_k 。

(3) 对于语料库 D 中的第 i 个单词 $w_{d,i}$ ，根据多项式分布 $k_{d,i} \sim \text{Multinomial}(\theta_d)$ 得到主题 $k_{d,i}$ ；根据多项式分布 $w_{d,i} \sim \text{Multinomial}(\phi_{k_{d,i}})$ ，得到词 $w_{d,i}$ 。

实际应用中，一般使用 Gibbs 抽样算法对 LDA 模型进行参数估计^[10]。其中，潜在的主题个数是一个先验数据，需要手动设置。主题数的设定一定程度上会影响主题生成的准确性。由于《诊疗方案》文本具有明显的主题分段结构，每个段落对应一个相对独立的主题，所以，《诊疗方案》的主题生成具备先验基础。基于此，本研究将《诊疗方案》按一级标题划分为 H 个主题。

1.3 诊疗方案文本主题匹配

对于一个待分析的《诊疗方案》，需要基于训练好的 LDA 模型预测每一段落所属主题分布的概率，即通过待分析的《诊疗方案》 d 及其组成单词 w ，推测潜在的主题 k 。由于 d 和 w 已知，所以可计算出 $P(w_j|d_i)$ 。基于 LDA 模型和《诊疗方案》已

知的文本-单词概率 $P(w_j|d_i)$, 可以训练《诊疗方案》的文本-主题概率 $P(k_h|d_i)$ 和主题-单词概率 $P(w_j|k_h)$:

$$P(w_j|d_i) = \sum_{h=1}^H P(w_j|k_h)P(k_h|d_i), \quad (3)$$

计算文本中每个词的生成概率:

$$P(d_i, w_j) = P(d_i)P(w_j|d_i) = P(d_i) \sum_{h=1}^H P(w_j|k_h)P(k_h|d_i) \quad (4)$$

再利用最大期望算法 (expectation-maximization, EM) [11] 估计主题概率 $P_j^d = (P_1^D, P_2^D, \dots, P_H^D)$, 即待分析《诊疗方案》 d 的第 j 段属于语料库 D 中 H 个预设主题内的哪个主题的概率。本文取预测概率最高 (即相关度最高的预设主题) 的主题 D_i 与待分析的《诊疗方案》 d 的第 j 段 d_j 相匹配。

1.4 差异性挖掘

计算获得《诊疗方案》 d_j 和与之主题分布最相似的文本 D_i 以后, 就可以对某《诊疗方案》版本 d'_j 与 d_j 进行对应主题的差异性挖掘。

文本差异性挖掘的整体架构如图 2 所示。基于文本关键词级别的差异性挖掘, 可以进一步挖掘语句级别的差异, 从而最终获得关键词和语句一致性的《诊疗方案》解读结论。特别指出:

- (1) 《诊疗方案》比一般通用文本具有更强的专业性和特殊用途。所以, 需要将一些自适应的停用词添加到停用词表。
- (2) 如果某语句存在删除和新增的关键词, 该语句会被重复匹配, 所以需要进行去重操作。
- (3) 对于新增的主题文本, LDA 模型的主题预测结果一般会呈现主题概率的多峰性特点 (见后文

实验), 可结合此属性进行差异性挖掘。

2 实验

2.1 数据集和实验设计

数据集: 本课题使用《诊疗方案》第一至第六版本作为训练集, 第七版本作为测试集。实验主要解读了《诊疗方案》的第七版和第六版。特别地, 本文方法适合其他任何版本间的对比解读。

实验设计: 训练集的所有《诊疗方案》被合并成一个文本。首先需要对训练集进行预实验, 以消除噪声干扰 (第一版本除外) 和超参数设置。LDA 模型中, 一级标题主题数设置为 9, 训练迭代数为 800, passes 设置为 40, 训练集的特征数目 (单词类型数量) 为 965, 文本数目 (文本段落) 为 43。LDA 模型的主题偏差指数收敛后停止训练。

2.2 预实验

预实验对训练集的所有《诊疗方案》进行初步分析, 包括《诊疗方案》的总体分析和各方案的特征分析, 预实验为 LDA 模型的训练提供基础。

2.2.1 文本总体相似度分析 为了对不同版本的《诊疗方案》进行初步分析, 本文将余弦相似度作为评价指标来度量《诊疗方案》之间的差异度, 从而初步评估各个方案的变化程度。余弦相似度是基于语义的关键字的字向量之间的相似度评价指标, 可定义为:

$$sim(\mathbf{v}(w_i), \mathbf{v}(w_j)) = \frac{\mathbf{v}(w_i) \cdot \mathbf{v}(w_j)}{|\mathbf{v}(w_i)| |\mathbf{v}(w_j)|} \quad (5)$$

其中, $\mathbf{v}(w_i)$ 、 $\mathbf{v}(w_j)$ 表示词 w_i 和 w_j 的词向量表征, 本文词向量采用 one-hot 编码形式。由图 3 可见:

- (1) 《诊疗方案》第一版相比之后各版本, 存在显著差异。

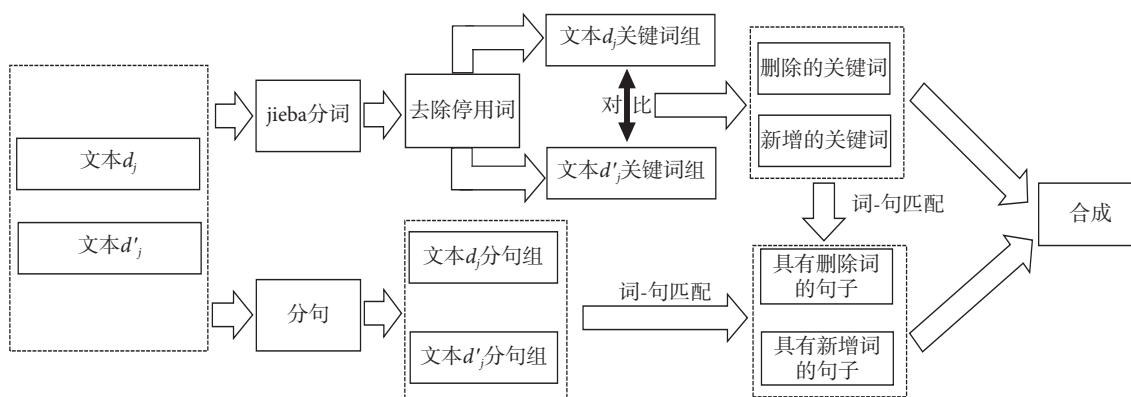


图 2 文本差异性挖掘架构

Fig.2 The framework of the text difference mining

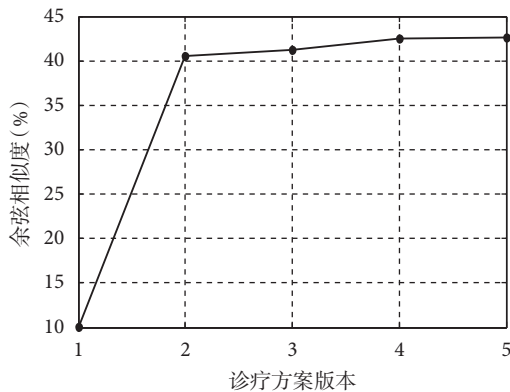


图3 不同版本《诊疗方案》的总体相似度变化(第6版分别与第1至第5版比较)

Fig.3 The variation of the overall similarity between the 6th and 1st to 5th version

(2)《诊疗方案》第二版至第五版的内容总体趋近稳定。特别指出,实验使用《诊疗方案》各版本对应的余弦相似度都是与第六版进行比较的结果。

2.2.2 《诊疗方案》各版本的特征分析 不同版本《诊疗方案》的总体相似度一定程度上反映了不同方案的相似性,但是无法精准刻画每个方案的特征。本研究进一步基于关键词的频率统计对不同版本《诊疗方案》的总体特征进行刻画。

针对《诊疗方案》进行分词、去除停用词等预处理以后,得到各方案主要关键词的统计数据,见图4所示。分析发现:

(1)相比其他诊疗方案,《诊疗方案》第一版的主要关键词频率差异显著。《诊疗方案》第一版主要关注疫情的防护,出现的高频包括“消毒”“口罩”“防护”“防护用品”等关键词。该结论与图3的文本整体相似度的结论一致。

(2)《诊疗方案》第二版至第六版中,“呼吸道”“肺炎”(图4中红色标注)两个关键词出现的频率逐步提高。由此可以推断,临床现象和相关研究确定了新冠病毒主要造成呼吸道和肺部感染。因此,在后续多个迭代的版本中逐步增加了这些关键词的权重。

(3)《诊疗方案》第六版中的高频词与此前各版本相似,但是第二至第五版的高频词主要包括“治疗”“患者”“病例”“冠状病毒”等,而第六版出现了关键词“注射液”(见图4中绿色标注)。通过查阅《诊疗方案》发现“注射液”与中药相对应,即《诊疗方案》第六版出现了中药注射液推荐治疗方案。

结合上述两部分的预实验可以发现,《诊疗方

案》第一版的总体相似度和主要关键词都与其他版本存在显著差异。因此,本研究的LDA训练数据不采用第一版,而只使用《诊疗方案》第二至第六版。

2.3 模型特点

表1展示了LDA主题生成模型的训练结果,包括对训练集中所有诊疗方案相似的主题进行聚类,并产生主题对应的权重值前10的关键词和相应权重。从表1可见:

(1)每个主题对应的关键词都与某特定主题具有较强的相关性。观察表1中加粗的关键词可以发现这些关键词与主题的相关性,如“流行病学”主题的“传播”“途径”关键词,“治疗”主题的“注射液”“疗程”关键词,以及“鉴别诊断”主题的“鉴别”“腺病毒”和“合胞病毒”关键词等。

(2)各主题的高权重关键词与该主题的语义之间存在较强的关联性,如主题“病原学特点”与关键词“病毒”(权重为0.04)存在较强的语义关联。

基于《诊疗方案》第二至第六版训练LDA模型以后,将《诊疗方案》第七版用于验证本文所提的整体架构。众所周知,《诊疗方案》的版本越高其内容越翔实,解读难度越大。本实验选用《诊疗方案》第六版与第七版进行对比解读,例如,使用LDA模型能将第七版的主题“临床分型”和“病例发现与报告”,与第六版相应的主题进行精准匹配。

研究显示,本文方法能精准识别任意两个《诊疗方案》之间的不变主题和新增主题。图5示意了不变主题的情况,即第七版《诊疗方案》中的“临床分型”和“病例发现与报告”主题精准匹配了第六版的对应主题,两个对应主题之间的相关性接近1,即匹配准确度接近100%。实验证明,任意两个《诊疗方案》版本之间,只要主题保持不变,都能准确匹配。图6示意了版本之间新增主题情况下的实验结果,即《诊疗方案》第七版新增了主题“病理改变”,用LDA模型进行主题匹配时,发现文本整体主题相关性呈现多峰性特点,且都具有较低的相关性。利用这两个特点,可以识别出哪个是新增主题。

2.4 实验结果

利用本文所提的文本差异性挖掘方法,对比《诊疗方案》第六版和第七版,分别针对“临床分型”“病例发现与报告”“治疗”“解除隔离和出院标准”四个主题自动生成解读报告,见表2到表5。

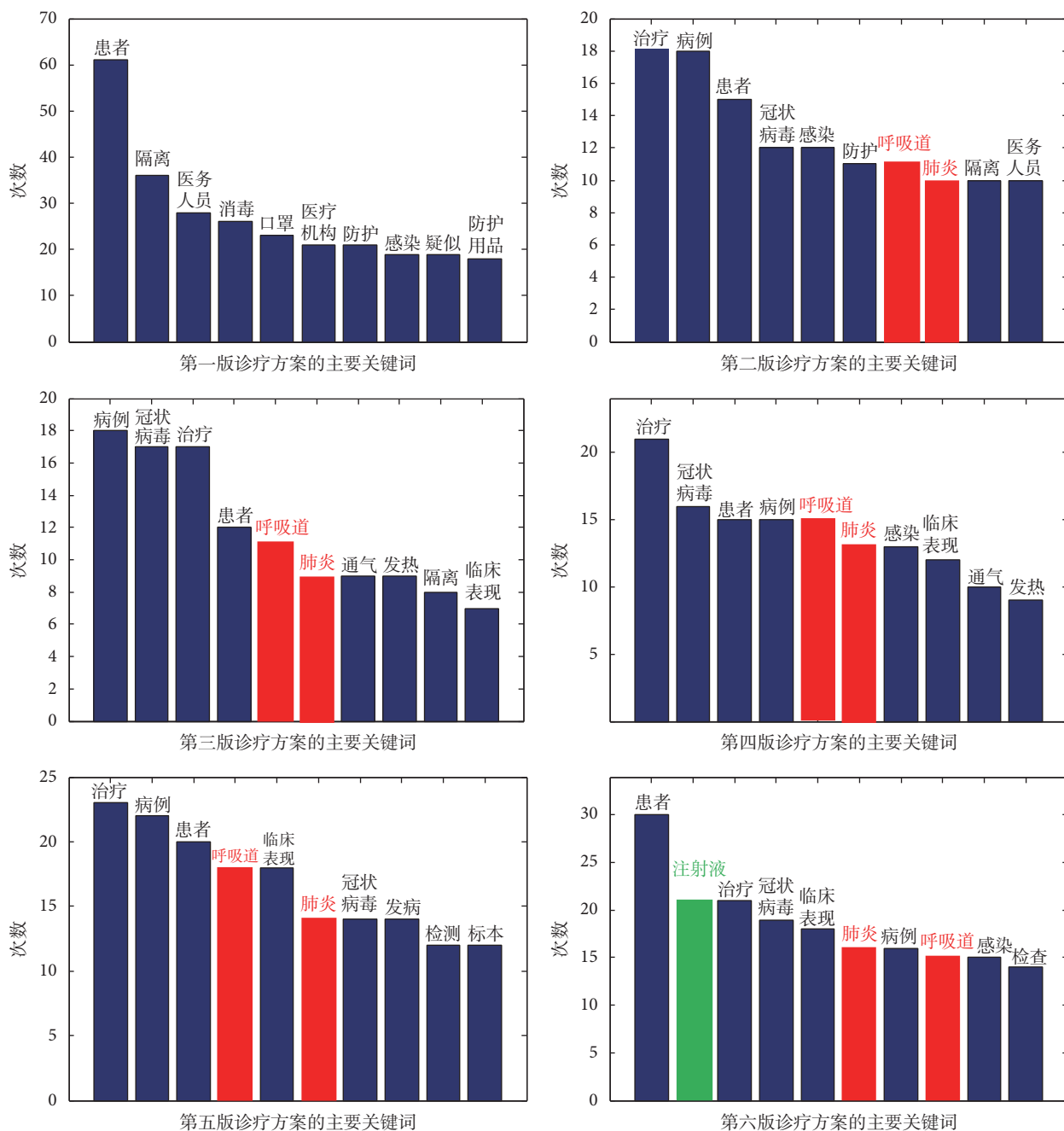


图 4 第一版至第六版诊疗方案排名前 10 的高频关键词统计图

Fig.4 Statistics of the top 10 high-frequency keywords from the 1st to 6th version

由表 2 到表 5 可见，本文方法能以符合人类认知的角度自动解读《诊疗方案》，自动区分不同版本之间的差异性语句和关键词，帮助医生和非医学专业人士快速、精准识别不同版本的差异，掌握新版本的重点。

3 讨论和结论

本文使用的 LDA 模型是一种基于贝叶斯概率的数学模型，可以广泛用于文本聚类、主题划分和预测。LDA 模型包含词、主题、文本三层结构，可

以通过词向量对文本中的词进行表征，以进行文本的相似性推测。LDA 是概率隐含语义分析 (probabilistic latent semantic analysis, PLSA) 模型^[12]的一种改进方法，改进了 PLSA 模型参数随训练文本增加线性增长、产生的语义可定义性差等不足，为本文进行有效的主题聚类提供了保障。

本文的主要贡献包括：

(1) 提出一种基于 LDA 模型的无监督学习方法，对不同版本《诊疗方案》的文本主题进行无监督学习。相比基于频率的匹配方法，如 TF-IDF^[13]，

表 1 LDA 主题生成模型训练结果。

Tab.1 The training results of the LDA generative model

主题	主题对应的关键词权重*关键词									
病原学原理	0.040*病毒	0.037*冠状病毒	0.027*研究	0.027*分离	0.027*灭活	0.024*培养	0.018*蛋白	0.015*呼吸道	0.014*小时	0.014*显示
流行病学原理	0.033*传播	0.017*感染	0.017*途径	0.017*传染源	0.009*接触	0.009*基础	0.009*病情	0.009*患者	0.009*人群	0.009*飞沫
临床表现	0.054*患者	0.020*严重	0.016*发热	0.016*表现	0.015*减少	0.014*危重	0.014*升高	0.013*症状	0.013*病例	0.013*多发
诊断标准	0.041*发病	0.033*病例	0.027*流行病学	0.026*病毒	0.024*临床表现	0.023*呼吸道	0.021*标本	0.020*符合	0.020*疑似	0.020*前14天
临床分型	0.037*符合	0.035*影像学	0.027*肺炎	0.027*情况	0.027*症状	0.024*轻型	0.024*合并	0.022*功能	0.022*衰竭	0.020*重型
鉴别诊断	0.088*肺炎	0.070*鉴别	0.042*流感病毒	0.042*病毒	0.024*已知	0.024*腺病毒	0.024*与非	0.024*肺炎	0.024*合胞	0.024*支原体
病例发现与报告	0.047*检测	0.044*疑似病例	0.029*病毒	0.026*呼吸道	0.025*医疗机构	0.022*病例	0.022*病原	0.020*会诊	0.020*核酸	0.017*隔离
治疗	0.038*注射液	0.017*推荐	0.015*服法	0.014*剂水	0.014*服用	0.012*生石膏	0.012*处方	0.012*患者	0.012*疗程	0.012*氯化钠
解除隔离和出院标准	0.034*呼吸道	0.034*出院	0.020*隔离	0.020*解除	0.017*症状	0.017*两次	0.017*时间	0.017*间隔	0.017*连续	0.017*采样

注：加粗项目表示符合认知的主题关键词

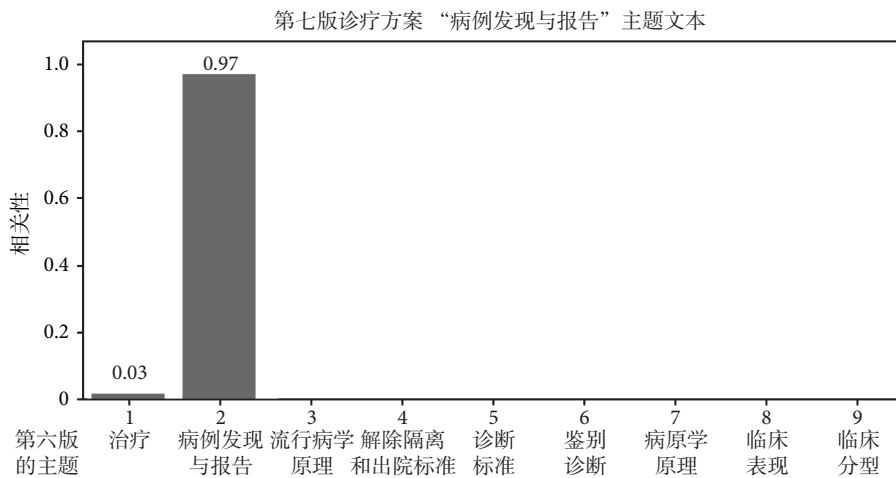
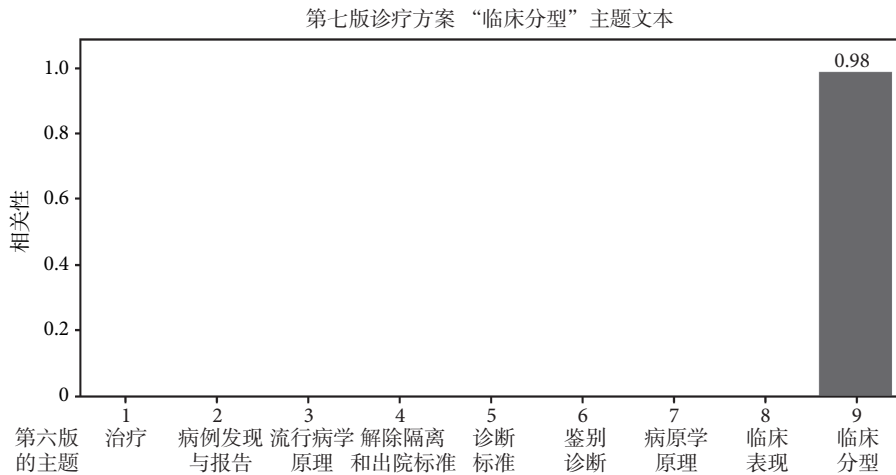


图 5 文本主题预测与匹配

Fig.5 Text topic prediction and match

第七版诊疗方案“病理改变”主题文本

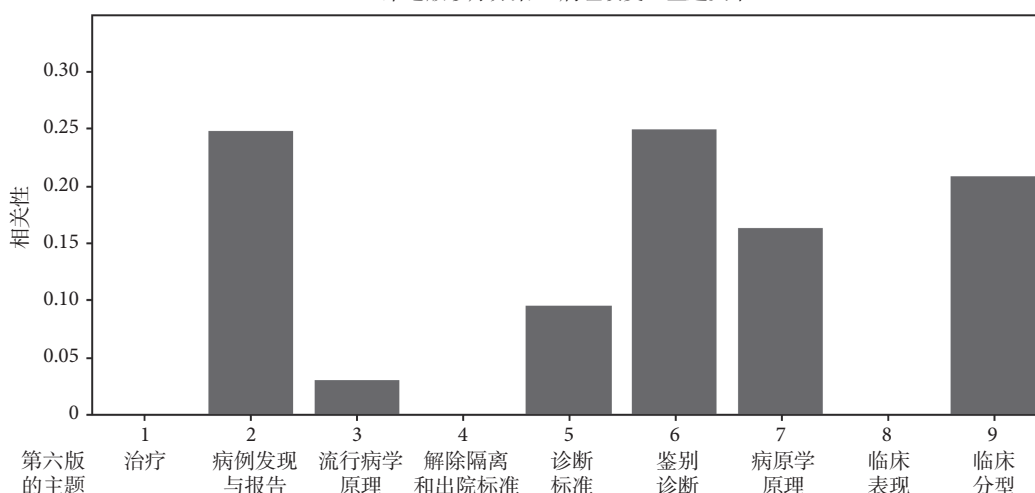


图6 主题“病理改变”的相关性匹配结果

Fig.6 Correlation match results of the topic "pathological changes"

表2 产生的第七版与第六版对比解读建议（临床分型）

Tab.2 Recommendations for comparative interpretation between the 7th and 6th version (clinical typing)

“临床分型”主题产生的解读建议
1. “儿童符合下列任何一条：1.出现气促(<2月, R ≥ 60次/分; 2~12月, RR ≥ 50次/分; 1~5岁, RR ≥ 40次/分; >5岁, RR ≥ 30次/分除外发热和哭闹的影响; 2.静息状态下, 指氧饱和度 ≤ 92%; 3.辅助呼吸(呻吟、鼻翼扇动、三凹征), 发绀, 间歇性呼吸暂停; 4.出现嗜睡、惊厥; 5.拒食或喂养困难, 有脱水征”为 新增句 。 新增的关键词包括 : 儿童符合一条气促月龄次月龄发热哭闹静息状态指氧饱和度辅助呼吸呻吟鼻翼扇动间歇性呼吸暂停嗜睡惊厥拒食喂养困难脱水

表3 产生的第七版与第六版对比解读建议（病例发现与报告）

Tab.3 Recommendations for comparative interpretation between the 7th and 6th version (case discovery and reporting)

“病例发现与报告”主题产生的解读建议
1. “疑似病例连续两次新型冠状病毒核酸检测阴性(采样时间至少间隔24小时)且发病7天后新型冠状病毒特异性抗体IgM和IgG仍为阴性可排除疑似病例诊断”为 新增句 。 新增的关键词包括 : 疑似病例连续两次新型冠状病毒核酸检测阴性采样时间至少间隔发病新型冠状病毒特异性抗体阴性排除诊断

表4 产生的第七版与第六版对比解读建议（解除隔离与出院标准）

Tab.4 Recommendations for comparative interpretation between the 7th and 6th version (de-isolation and discharge criteria)

“解除隔离与出院标准”主题产生的解读建议
1. “痰、鼻咽拭子等呼吸道标本核酸检测阴性连续两次阴性, 采样时间至少间隔24小时”为 新增句 。 新增的关键词包括 : 小时

LDA模型方法具有语义分析功能和更小计算复杂度的优势^[14], 在实践中往往具有更好的鲁棒性。

(2) 提出一种轻量级的相同主题文本内容差异化的挖掘方法。用于确定主题的文本, 能自动搜索不同文本之间新增或删减的关键词, 以及与之对应的语句。

本文提出的基于机器学习的计算机辅助智能分析方法, 对文本主题进行无监督学习, 自动计算文本主题, 实现主题的精准匹配, 准确率达100%; 自动分析不同版本《诊疗方案》的异同, 主动给医护人员推送新版本的关注重点; 自动生成关键词

和语句级别的解读报告, 降低《诊疗方案》解读的专业难度, 提高非专业人员对诊疗方案的认识水平。本文方法能替代人工解读, 实现《诊疗方案》的计算机自动智能解读。

实验证明, 相比于人工解读, 本文方法通过对现有《诊疗方案》多个版本的主题进行无监督学习, 能够实现目标版本的主题匹配, 准确率达100%。此外, 本方法能够自动生成语句级和关键词级的《诊疗方案》解读报告, 有助于非专业人士理解医学专业知识。下一步, 我们会将《诊疗方案》与全国的COVID-19确诊和治愈出院等数据相结合, 循证研究二者之间的相关性和《诊疗方案》的

表 5 产生的第七版与第六版对比解读建议 (治疗)

Tab.5 Recommendations for comparative interpretation between the 7th and 6th version (treatment)

“治疗”主题产生的解读建议
1. “有条件可采用氢氧混合吸氧 (H ₂ /O ₂ : 66.6%/33.3%) 治疗” 为新增句。新增的关键词包括: 采用 氢氧混合
2. “对孕产妇患者的治疗应考虑妊娠周数, 尽可能选择对胎儿影响较小的药物, 以及是否终止妊娠后再进行治疗等问题, 并知情告知” 为新增句。新增的关键词包括: 妊娠 胎儿 终止 知情告知
3. “根据气道分泌物情况, 选择密闭式吸痰, 必要时行支气管镜检查采取相应治疗” 为新增句。新增的关键词包括: 气道 吸痰
4. “相关指征: ① 在 FiO ₂ > 90% 时, 氧合指数小于 80 mm Hg, 持续 3-4 小时以上; ② 气道平台压 ≥ 35 cm H ₂ O” 为新增句。新增的关键词包括: 气道
5. “循环支持: 在充分液体复苏的基础上, 改善微循环, 使用血管活性药物, 密切监测患者血压、心率和尿量的变化, 以及动脉血气分析中乳酸和碱剩余, 必要时进行无创或有创血流动力学监测, 如超声多普勒法、超声心动图、有创血压或持续心排量 (PiCCO) 监测” 为新增句。新增的关键词包括: 无创 动力学
6. “如果发现患者心率突发增加大于基础值的 20% 或血压下降大约基础值 20% 以上时, 若伴有皮肤灌注不良和尿量减少等表现时, 应密切观察患者是否存在脓毒症休克、消化道出血或心功能衰竭等情况” 为新增句。新增的关键词包括: 尿量 脓毒症 衰竭
7. “患有重型或危重型新型冠状病毒肺炎的孕妇应积极终止妊娠, 剖腹产为首选” 为新增句。新增的关键词包括: 重型 孕妇 妊娠
8. “7. 免疫治疗: 对于双肺广泛病变者及重型患者, 且实验室检测 IL-6 水平升高者, 可试用托珠单抗治疗” 为新增句。新增的关键词包括: 双肺 治疗
9. “出现机械通气伴腹胀便秘或大便不畅者, 可用生大黄 5~10 g” 为新增句。新增的关键词包括: 机械 大便 生大黄
10. “出现人机不同步情况, 在镇静和肌松剂使用的情况下, 可用生大黄 5~10 g 和芒硝 5~10 g。” 新增的关键词包括: 人机 镇静
11. 推荐中成药: 血必净注射液、热毒宁注射液、痰热清注射液、醒脑静注射液、参附注射液、生脉注射液、参麦注射液” 为新增句。新增的关键词包括: 注射液
12. “血液净化治疗: 血液净化系统包括血浆置换、吸附、灌流、血液/血浆滤过等, 能清除炎症因子, 阻断‘细胞因子风暴’, 从而减轻炎症反应对机体的损伤, 可用于重型、危重型患者细胞因子风暴早中期的救治” 为新增句。新增的关键词包括: 血液 血浆 细胞 机体 救治
13. “要注意洛匹那韦/利托那韦相关腹泻、恶心、呕吐、肝功能损害等不良反应, 同时要注意和其他药物的相互作用” 为删除句。
14. “要注意上述药物的不良反应、禁忌症 (如患有心脏病者禁用氯喹) 以及与其他药物的相互作用等问题” 为新增句。新增的关键词包括: 不良 心脏 禁忌

效果, 为医护技术评价、医院管理和卫生决策提供参考。

利益冲突声明: 本文全体作者均声明不存在利益冲突。

参考文献

- 1 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第一版)[EB/OL]. (2020-01-16) [2020-03-24]. <http://www.nhc.gov.cn/yzygj/s7659/202001/b91fdab7c304431eb082d67847d27e14.shtml>.
- 2 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第二版)[EB/OL]. (2020-01-18) [2020-03-24]. http://www.360doc.com/content/20/0202/00/397911_889144879.shtml.
- 3 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第三版)[EB/OL]. (2020-01-23) [2020-03-24]. <http://www.nhc.gov.cn/xcs/yqfkdt/202001/f492c9153ea9437bb587ce2ffcbee1fa.shtml>.
- 4 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第四版)[EB/OL]. (2020-01-27) [2020-03-24]. <http://www.nhc.gov.cn/yzygj/s7653p/202001/4294563ed35b43209b31739bd0785e67.shtml>.
- 5 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第五版)[EB/OL]. (2020-02-04) [2020-03-24]. <http://www.nhc.gov.cn/xcs/zhengcwj/202002/3b09b894ac9b4204a79db5b8912d4440.shtml>.
- 6 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第六版)[EB/OL]. (2020-02-19) [2020-03-24]. <http://www.nhc.gov.cn/yzygj/s7653p/202002/8334a8326dd94d329df351d7da8aefc2.shtml>.
- 7 国家卫生健康委员会, 国家中医药管理局. 新型冠状病毒肺炎诊疗方案(试行第七版)[EB/OL]. (2020-03-04) [2020-03-24]. <http://www.nhc.gov.cn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>.

- 8 Anderson J A. Cognitive and psychological computation with neural models. *IEEE Trans Syst Man Cybern*, 1983, SMC-13(5): 799-815.
- 9 Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation. *J Mach Learn Res*, 2003, 2003(5): 993-1022.
- 10 Casella G, George E I. Explaining the Gibbs sampler. *The American Statistician*, 1992, 46(3): 167-174.
- 11 Gaussier E, Goutte C. Relation between PLSA and NMF and implications// *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2005: 601-602.
- 12 Lu Y, Mei Q, Zhai C. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. *Information Retrieval*, 2011, 14(2): 178-203.
- 13 Ramos J. Using TF-IDF to determine word relevance in document queries// *Proceedings of the First Instructional Conference on Machine Learning*. Piscataway: 2003: 33-142.
- 14 Pu X, Long K, Chen K, et al. A semantic-based short-text fast clustering method on hotline records in Chengdu// *2019 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. Fukuoka: IEEE, 2019: 516-521.

收稿日期: 2020-03-16 修回日期: 2020-03-25

本文编辑: 李蓓兰