

基于词频分析的初中英语教学资源个性化推送

刘世豪*
LIU Shi-hao

摘要

随着我国互联网技术的快速发展,教学模式与教学方法快速变革,个性化学习越来越受到关注。但是,许多学生在学习无法有效地选择所需要匹配的学习资源,或者在搜索查找的过程中花费更多的时间,同时统一的标准的资源也不符合学生的个性化学习需求。基于词频分析的方法对初中英语题目进行分析统计,依据学生个性特点优化相关内容的推送。

关键词

词频分析; 初中英语; 个性化学习; 个性化推送

doi: 10.3969/j.issn.1672-9528.2019.07.060

1 研究背景

我国的基础教育架构发展逐渐趋于完善,在很大程度上为学生的综合素质发展提供了制度与环境支持。在教学活动中,教师一直在追求实现孔子的“因材施教”教育思想,因材施教即根据每位学生的不同身心发展状况、认知状态、个性特点以及知识掌握情况等对学生采取不同的教学方法,这就是个性化教学。许多中小学教师正在摆脱传统的教学模式束缚,深入分析学生的心理特点,利用课堂教学的“真实性”、“情境性”和“时效性”等,探索与采用不同的教学方法,以期达到个性化教学。但存在共性问题教师只会按照自己的思路讲课,完成教学任务^[1]。这是由于教师无法真正地获得、分析学生的全部学习行为,这使得教师对于学生的知识需要缺乏相应的了解,学生的学习行为反馈的缺失导致教师很难有的放矢地开展个性化教学。

但是,依托于互联网技术发展的个性化推送系统可以较好地解决这一问题。通过对初中英语题目的单词词频进行计算,发现题目之间相关联的规则,分析出每两个或多个题目之间的相似度,分析比较出用户所需的题目,由系统自动地进行个性化推送。本文拟将个性化推送系统层架构分为五层:输入层,数据层,处理层,推送层,应用层等。学习者输入结构化或半结构化信息,经数据层分析确认维度,到处理层进行切分匹配,最后将相似资源推送给学习者,以达到精准教学、精准推送的目的。

2 初中英语题目词频

2.1 词频内涵

词频(Term-frequency)即某一个特定的词在文本之中出现的次数。词是一个完整语句结构中最基本的单位互联网的发展以及大数据爆炸对人们的日常生活产生了巨大的影响,面对着海量的信息资源,人们逐渐习惯碎片化阅读以及“单元格”式的信息获取方式,但由于存在客观技术缺陷以及主观信息搜索能力较弱等问题,往往会导致目标数据或者目标内容的缺失。而关于词的频率的分析则可以通过借助统计分析学习者行为活动信息和兴趣集合,由面到点似地筛选相关题目信息、要素,更加准确、快速地匹配到目标题目。

2.2 词频分析方法

2.2.1 TF-IDF

TF-IDF(term frequency-inverse document frequency)技术常用于数据等的挖掘,词频即在限定的集合中出现的频率次数,通常某个词出现的频数越小,则区别率更高。运用TF-IDF进行分析时,因频次不同的原因,会相对偏离正确结果。TF-IDF表达式分为两部分,即 $tf_{a,b} = \frac{n_{a,b}}{\sum_k n_{k,b}}$ 、 $idf_a = \ln \frac{|D|}{|\{b:t_a \in d_j\}|}$,其中, $n_{a,b}$ 是词于文本 d_b 中出现的次数, $\sum_k n_{k,b}$ 是文本中所有词出现的次数总和,TF-IDF值为 $tfidf_{a,b} = tf_{a,b} \times idf_a$ 。

2.2.2 语义相似度

相似性是在心理学领域之中被提出的,相似性是人们通过对对象的刺激的感知产生一种心理上的反应。关于相似度,Dekang Lin从信息理论的角度给出了一组定义,他认为对于任意两个对象,它们之间的相似度取决于它们的共性和差异性,当两个对象之间的共性越多时,这两个对象的相似度越大^[2]。Dekang Lin关于相似度计算的公式为:

$$Sim(A, B) = \frac{\log(common(A, B))}{\log(description(A, B))}, \log(common(A, B)) \text{ 为 } A, B \text{ 共性}$$

* 渤海大学教育科学学院 辽宁锦州 121007

描述所需的所有信息量， $\log(\text{description}(A,B))$ 为完整地描述 A,B 的信息量。在之后的各种相似度计算公式的发展过程中，大多数是以 Dekang Lin 的语义相似度计算公式为基础的。关键词词频的分析需要借助词与词之间的相似性来进一步的分割或者聚合。

2.2.3 语义树

语义树是由图示形式构成的语义代数。在同义词词典中，所有的词组成了一颗或几颗树状的层次结构^[3]。一般同一词典都是将所有的词组织在一颗或几颗树状的层次结构中^[4]。CNKI 运用语义树构建了概念之间与概念树形之间的关系，计算模型可以计算出两个词或多个词之间的距离。在语义树之中，义原组成了每一个词的语义且每个义原的权重是不同的，语义树中的每一个节点都是代表了一个义原，关于词语义的确定为节点之间的距离计算提供了支撑。

2.3 初中英语题目词频分析

大多数软件如 CiteSpace 等自动聚类的依据主要是来源于谱聚类的算法，网络之间的连接性为聚类分析提供了前所未有的便捷性。用户浏览初中英语题目以及题目练习出错时，系统可自动地跟踪、收集此类学习行为痕迹，对具有一定长度的英语题目中出现的高频单词进行分析比对，词频统计模块对于词频的统计基于特定的目标用户的已进行完成的行为。

现对某省的多份试题进行词频分析与排列，在单词频率结果的构建中，词频分析模块会自动地清除频次较低的单词（如频次为 0 的单词），词频分布（部分）与数据透视表如表 1 和图 1 所示。

表 1 词频分布

Word	Frequency
the	1768
to	1197
and	865
in	809
you	488
he	273
will	144
more	135
time	124
Chinese	70
world	70
way	64
may	56
could	40
family	38
become	35
sports	30
water	28

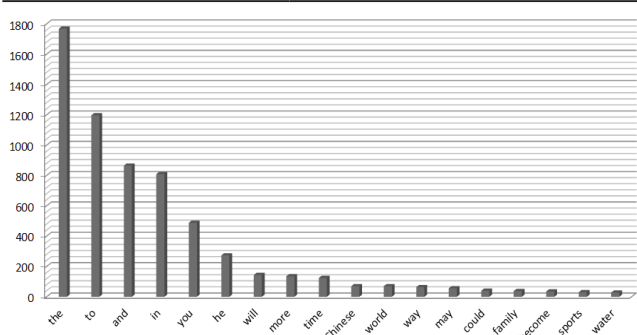


图 1

系统分析一道初中英语题目时，仅按语句结构和单词

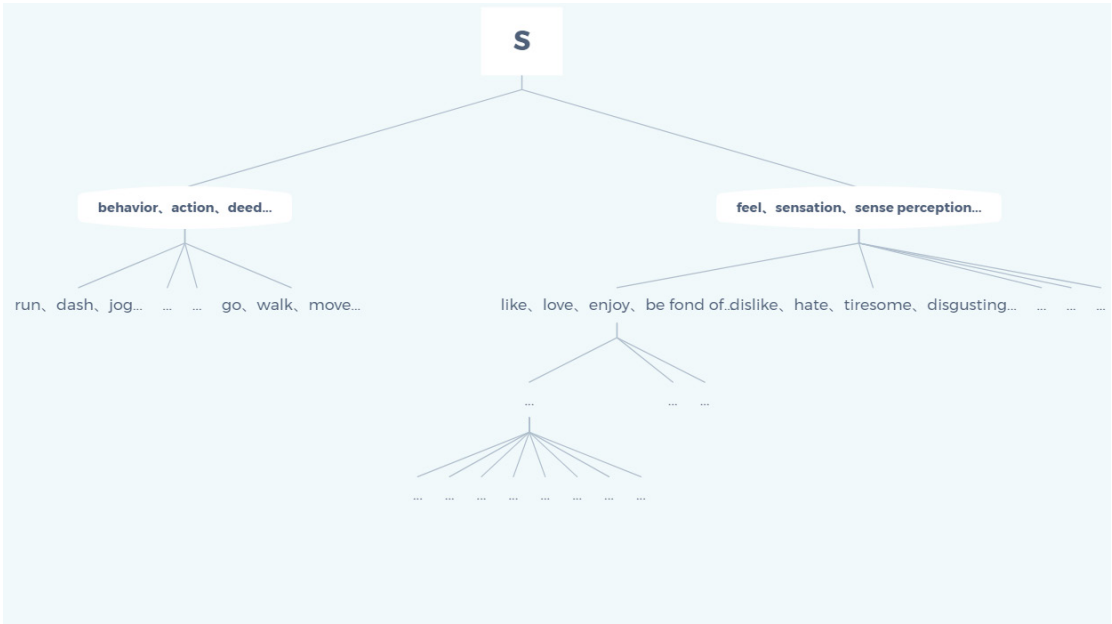


图 2 英语单词语义树

切分原题目，并根据语义树匹配中、下、同位词，语义树如图2所示，原则上默认同位词具有优先级，匹配结束后与空间内题库资源连结并推送试题。当统计最小数值至少为两到题目的词或短语时，将单词或短语作为词频模块分析对象，把多类单词或短语的复杂关系以数值定量或图表形式表现出来。词频模块进行词频分析方法的具体操作步骤：（1）确定目标语句；（2）切分目标单词；（3）统计高频词汇；（4）以高频单词的频率为基础构建矩阵；（5）分析矩阵；（6）分析结果。

当两个或两个单词组合（即固定搭配、常用搭配等考点）在不同题目中同时出现，采用共词分析法构建共词矩阵，矩阵构建完成后，选用聚类分析方法对题目关键词信息进行可视化分析。在词频分析的基础上，词频分析模块需要切分语句的语法结构，使匹配结果更加精确，由于切分时以英语本身语法构成为基础，词频分析的同时要进行单词的词性分析。现将英语语句结构进行分割归类，词频统计模块中关于初中英语的语句构成分类如表2所示。

表2 初中英语语句构成分类

语句名称	结构
主动句（AS）	N_do_ ...
一般现在时（SPRT）	N_be/do_ N_be/do_not_ ...
一般过去时（SPAT）	N_were/was/verb past tense_ N_were/was/verb past tense_not_ ...
一般将来时（GFUT）	N_am/is/are_going to_do_ N_will/shall/be going to_do_ ...
定语从句（FS）	RP/RD_n_verb_ RP_verb_ ...
状语从句（AC）	...

语句结构切分与分析完成后即可与初中英语资源题库或者WEB进行匹配，利用混合式推送技术向学习者以及包括教师等用户进行个性化推送，推送的默认原则为：推送试题资源的数与原题目数相一致，用户可以依据自身的学习情况与认知主动地选择其他数目选项题目的推送。

3 个性化推送

3.1 个性化推送的概念

个性化推送的概念与运用来源于电子商务。随着技术的发展，个性化推送逐渐运用到教育领域方面。网络学习平台中的个性化推送是根据学习者的个性化特征，如学习爱好、知识掌握程度或学习特点等，主动地向学习者推送适合其学

习需要或可能感兴趣的学习资源的一种推荐机制^[5]。

3.2 个性化推送优势

当前国内对于个性化推送的研究成果越来越多。从个性化推送的可应用性方面来看，面向特定的目标推送的内容边界较为清晰，可以通过个性化推送技术推动以及加快社会、学校、教师以及学生等行为主体的教育观念的改变，实现教育思想、教学模式以及教学策略和方法等的改革。同时，个性化推送相对于传统的教学方法具有特定的优势。其一，在内容过多的资源中，个性化推送会基于各种行为数据的统计分析，匹配目标群体的兴趣点，使资源的横向和纵向迁移更加的合乎规则；其二，个性化推送的技术本身是为了解决信息过量和信息超载而产生的，随着技术的创新以及观念的革新，个性化推送技术应用范畴较之以前有了多维度地分散发展，可应用于ebay、淘宝类购物平台和微博、微信、今日头条等通讯、娱乐APP以及教育教学领域；其三，个性化推送的技术可使特定的主体获取自身独特的信息资源，满足其个性化的信息需求等。

3.3 个性化推送技术

个性化推送技术主要包括基于内容的个性化推送技术、基于规则的个性化推送技术、协同过滤的个性化推送技术以及混合式个性化推送技术等。20世纪初，个性化推送技术开始被运用于教育领域，Recker等人开发的Alered Vista System是最早基于E-learning环境下的个性化推送系统，AVS系统采用了协同过滤技术，与之相似的还包括Manouselis提出的The Context E-learning with Broadband Technologies system^[6]，Anderson开发的Rule-Appling Collaborative Filtering System^[7]等。不同的个性化推送技术具有不同的特点，通过推送，能够较迅速、精确地响应目标主体的请求。基于内容的个性化推送技术更多地是基于If-Then判断语句，通过属性的相似度进行匹配；基于规则的个性化推送技术是基于规则进行推送；协同过滤的个性化推送技术是基于相似度来预测目标主体的选择；混合式个性化推送技术则结合了多种个性化推送技术，对于主体而言，此技术具有更大的选择偏向性。

4 初中英语教学资源的个性化推送

4.1 适合基于词频分析的初中英语教学资源的个性化推送技术

混合式个性化推送技术打破了单一的个性化推送技术界限，通过对学习者出错的题目以及浏览过的题目进行整理、归类，然后根据合适的英语词语词频分析的方法对试题聚类，挖掘初中英语教学资源题库中的题目资源，实现“导向性”

试题的推送。

混合式个性化推送技术适合中小学教师与学生的需求与推送的机制，有利于扩充初中英语教学资源，实现国家倡导的联通，通过混合式个性化推送技术构建本校、本地区的初中英语教学资源的个性化推送模型，提高学习者的学习效率与学习的自主性。在英语及其他科目的教育教学活动过程中，教师通过个性化推送的结果，针对性地对学生提供个性化指导；学习者能够在课堂和自主学习的过程之中产生兴趣偏好的学习。学生与教师的直接性或间接性的交互行为形成重要数据，此类行为数据反馈给系统，系统对交互中产生的开放的、异构的数据进行聚合，进一步促进学生的自主学习能力。由此可见，个性化推送教学资源技术在目前的教学环境中将拥有广泛的接受群体与受益群体。

4.2 混合式个性化推送技术与语义相似度计算的结合

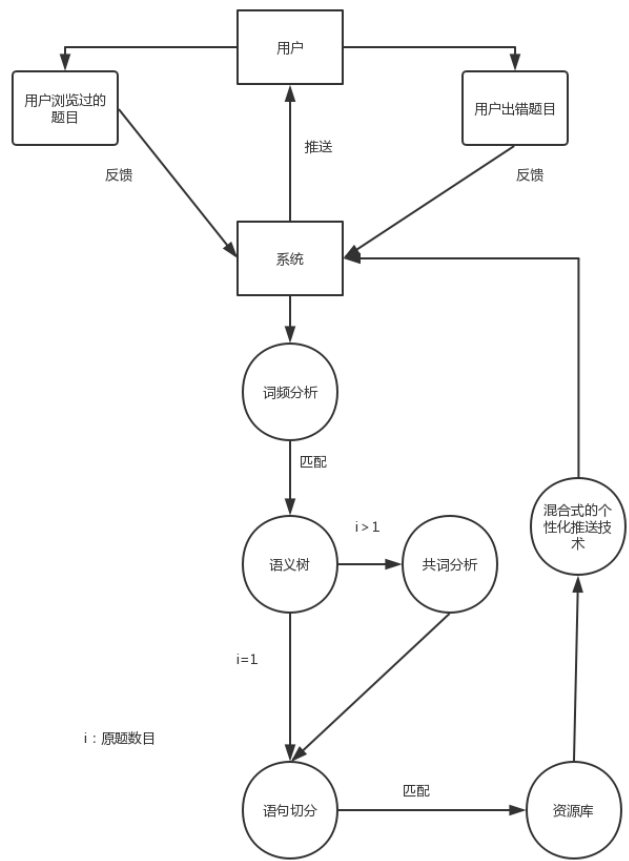


图3 混合式个性化推送技术与词频分析结合的过程图

混合式个性化推送技术是基于联通主义学习理论，联通主义学习理论是由 George Siemens 提出的。他认为学习不再仅仅是一个人的活动，学习是连接专门节点和信息源的过程^[8]。混合式个性化推送技术具有许多传统教学方式不能比拟的有点，它的灵活性、精确性、高效性大大地提高了教师教学与学生学习的效率，且理论上混合式个性化推送具有无

穷性，不存在达到设定值即终止的学习模式。

词频分析采用不同的方式与方法，对不同的、海量的初中英语教学资源进行词频高低多维度的统计与分析，分类出相似度较大的题目。混合式个性化推送技术与词频分析的过程如图3所示。

通过基于词频分析的初中英语教学资源的个性化推送技术，可以更好地满足教学需求，提高学习的效率，得到学习者的喜好信息，实现个性化教学与个性化学习的教育愿景，激发学生自主学习的动力，形成有推广价值的资源推送技术。

参考文献

[1] 姜强,赵蔚,王朋娇,王丽萍·基于大数据的个性化自适应在线学习分析模型及实现[J]·中国电化教育,2015(01):85-92·
[2] 李飞·基于本体的概念语义相似度测量研究[D]·广西师范大学,2016·
[3] 张琨昱·基于HowNet的个性化学习系统的研究与设计[D]·江西师范大学,2016·
[4] 张亮,尹存燕,陈家骏·基于语义树的中文词语相似度计算与分析[J]·中文信息学报,2010,24(06):23-30·
[5] 方琦·中小学网络学习平台个性化推送系统的研究与设计[D]·浙江工业大学,2013·
[6] Gulden Uchyigit, Matthew Y Ma· Personalization Techniques and Recommender Systems· World Scientific Publishing Company, 2008-04-04 ·
[7] M· Anderson, M· Ball, H· Boley, S· Greene, N· Howse, D· Lemire, S· Mcgrath· RACOFI: A Rule-Appling Collaborative Filtering System, <http://www.citeulike.org/user/koles/article/1803210>·
[8] 王佑镁,祝智庭·从联结主义到联通主义:学习理论的新取向[J]·中国电化教育,2006(03):5-9·

【作者简介】

刘世豪(1994-),男,山东淄博人,渤海大学硕士研究生,主要研究方向为现代教育技术。

(收稿日期:2019-06-26)