DOI：10.16353/j.cnki.1000-7490.2017.04.019

● 王效岳，赵冬晓，白如江（山东理工大学图书馆，山东　淄博　255049）

基于专利文本数据挖掘的技术预测方法与实证研究*
——以纳米技术在能源领域应用为例

摘　要：［目的/意义］利用文本数据挖掘技术深入挖掘专利文献主题内容，与更具有前瞻性的领域规划文本进行对比分析，发现哪些技术主题是当前饱和技术主题，哪些是未来潜在发展技术主题，以期提高技术预测的准确性，并降低预测结果的风险性。［方法/过程］在对国内外的技术预测方法进行梳理总结的基础上，利用文本数据挖掘技术挖掘专利文献技术主题，在此基础上加入时间维度，形成专利技术路线图，然后与规划文本中的规划信息进行对比分析，从而进行技术预测分析。［结果/结论］通过纳米技术在能源领域的应用进行实证研究，有效地实现了技术预测，并降低了技术预测结果的风险性。［局限］仅从专利的研究水平指标判别研究主题潜力，尚未考虑其他影响因素，指标体系还不完善；在数据源的选择上技术领域专利文本和规划文本存在不对称的情况。

关键词：专利；文本挖掘；数据挖掘；技术预测；技术路线图

Abstract: [Purpose/significance] This paper mines the main contents of patent texts by using text data mining technologies and compares the results with the planning documents in the prospective fields. By doing this, it can discover the potential technical themes to enhance the accuracy of technology prediction and lower the risks of prediction results. [Method/process] The paper first reviews the technology prediction methods both in home and abroad and mines the technical themes of patent documents by using text data mining technology. Based on that, taking time dimension into consideration, the paper draws the patent technology roadmap and compares it with information in the planning documents to conduct technology forecasting. [Result/conclusion] The empirical research of nanotechnology in energy field shows that the method proposed in the paper effectively predicts the technology and cuts the risks of prediction results. [Limitations] First, the paper predicts the potential of research theme only from research status of patents without considering other influential factors, so the indicator system is not complete. Second, patent texts and planning documents in technical field could not perfectly match each other.

Keywords: patent; text mining; data mining; technology forecasting; technology roadmap

1　专利与技术预测

技术预测[1]是指在具体的框架范围内分析技术发展的条件和潜力，是对技术的现状和发展进行持续的观察研究，从而初步确定技术的未来应用领域和前景，评估其潜能。技术预测最早在20世纪30年代开始在美国出现，在第二次世界大战中应用逐渐广泛，战后在军事和航天领域得到重视[2]，20世纪70年代，随着科技发展重心由军事用途向民用部门转移，传统的预测方法难以适应瞬息万变的市场环境，技术预测陷入低谷，直到20世纪末，在全球创新的需求下，以及新的预测方法和预测形式的推动下，技术预测又蓬勃发展起来[3]。

技术预测的分析对象是含有技术内容的各类文献，目前技术内容的主要刊载文献包括专利、学术论文、报告等，而专利是世界上最大的技术信息源，据实证统计，专利包含了世界全部科技信息的90%～95%，因此专利文献无疑是技术预测最为理想的数据源[4]。美国专利局早在20世纪70年代初就开始技术预测的工作，美国专利商标局在1971年成立了技术评价和预测办公室，建立了OTAF数据库，专门用来进行这方面的工作[5]。

目前技术预测的方法主要有定性分析和定量分析以及通过文本数据挖掘的方法，定性分析的主要代表方法是德尔菲法，该方法是利用专家群体的智慧来预见未来的技术发展，因此技术预测结果很容易带有主观色彩，而且该方法耗时耗力，准确性和可靠性都比较低，文献[6]专门针对德尔菲法存在的缺陷进行了深入的解读和分析，有些学者[7]尝试将德尔菲法与技术路线图方法相结合，由小范

* 本文为国家社会科学基金项目“未来新兴科学研究前沿识别研究”的成果，项目编号：16BTQ083。

围的专家学者绘制技术路线图，再由德尔菲法得出的结果作为补充，对技术路线图中涉及的技术、产品进行说明，虽然在一定程度上能提高可靠性，但是需要大量的专家资源进行人工分析，代价过大，因此定量分析的方法越来越多地得到使用。

定量分析方法是根据客观的专利数据进行定量的技术预测。文献 [8] 抽取可行性数据并对技术预测的准确度影响因素进行分析，发现时间和方法对技术预测的准确度有着显著影响，并认为定量方法比定性方法准确度更高。文献 [9] 通过专利申请量与授权量比率以及平均滞后期构建模型，并在磁阻式随机存取记忆体和有机发光二极管领域进行了技术预测，证实该方法对技术预测有着明显的优越性。文献 [10] 利用矩阵映射和 KM-SVC 算法对技术空白领域进行了定量客观的预测，比较了美国、欧洲和中国的技术发展趋势和技术空白领域。文献 [11] 利用专利数据绘制专利地图，关注专利地图中密度低但面积大的技术空白区，作为识别新技术机会的依据。文献 [12] 探讨了专利与技术预测之间的联系，总结出技术预测的三个主要方面——技术活跃度、技术生命周期和技术进化方向，进一步深度分析专利分析和专利地图、技术分析和技术预测的关系，构建了基于专利地图的技术预测体系，以质子交换膜燃料电池技术为例，实证了其科学性和实用性。文献 [13] 以生命周期理论和 Logistic 模型，对 RFID 技术进行了技术预测，提出适应我国技术发展的专利策略。文献 [14] 从专利申请量、申请类别、申请人、发明人以及主要专利领域分析了吉林汽车电子技术的发展现状和趋势，并利用分析结果进行了技术预测，为技术发展模式和发展路线提供依据。文献 [15] 构建了专利情报分析方法体系，并从管理和技术两个层面对面向技术预见的情报分析方法进行探索。文献 [16] 针对技术三性：技术发展方向、技术发展水平和技术发展潜力，建立技术预测模型，选择运用相应的专利分析指标和方法，形成技术预测的专利分析方法框架。文献 [17] 将 TRIZ 理论应用到技术预测中，结合 TRIZ 理论的三大技术进化模式，识别了我国物联网关键技术的专利发展现状，绘制技术进化路径并进行技术预测分析。文献 [4] 将数据挖掘的方法应用到技术预测分析中，以中关村科技园区专利情报实证分析证实了该方法的可行性和有效性。文献 [18] 提出基于情景分析法的技术预测研究，该方法分为描述型和定量型两种，前者侧重主观想象来描绘未来的可能性，后者主要借助概率论对未来的可能情景进行定量描绘[19]，因此将定量和定性分析有效地结合。也有许多学者将文本挖掘的方法应用到专利分析中，文献 [20] 利用文本挖掘的方法对专利进行了结构化信息的抽取，并通过聚类分析展示了 CNT 领域的技术发展现状，进行了技术机会的探测，但是该方法并没有进行专利技术发展的时间动态展示，因此对于技术预测来说缺乏时效性。

定量分析弥补了定性分析方法缺乏数据支撑的问题，但仍存在一定的局限性：目前定量分析的方法都只是停留在专利的外部指标的分析，如技术生命周期、国别、地区分析、专利权人分析、IPC 分析以及引文分析和技术功效分析等，并没有深入到专利文本的内部，挖掘专利内容中的重要技术信息，实现客观准确的技术发展态势描绘。而且现阶段的技术预测都只是通过专利文本分析技术发展态势从而进行技术预测，预测结果的准确性有待考量，有一定的风险性。

为了弥补传统的技术预测分析方法的不足，本文将利用文本数据挖掘技术，通过挖掘专利文本内容中能够代表专利研究内容、方法和效果信息或信息块，构建专利技术发展路线图，描述当前的技术发展现状；通过将技术发展路线图与相应的规划文本进行对比分析，从一定程度上减小技术预测的风险性。

2 理论与方法

2.1 研究思路

本文从相关专利文本的标题和摘要信息中通过标注和抽取技术将能够代表专利的研究主题的词或短语抽取出来，每一篇专利文本都由代表研究重点、主要技术和研究结果的代表性的词或短语的组合结构进行表示，通过整理和分析构建技术发展路线图，该路线图旨在描绘当前某个主题的研究发展现状，通过与相应的规划文本之间的对比分析，从而确定各主题的发展饱和情况，分析出存在发展空间的潜在研究主题，具体技术路线见图 1。

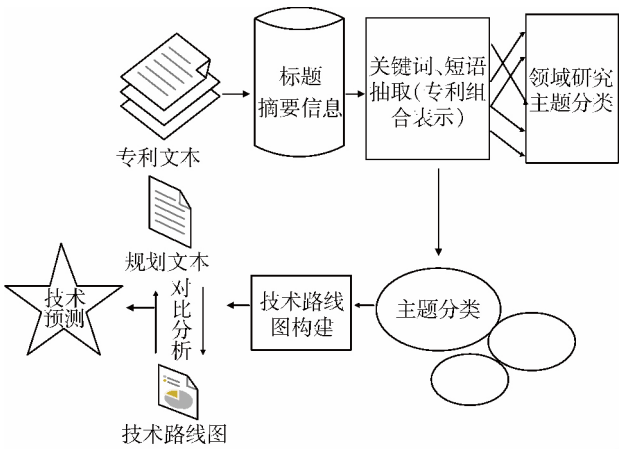

图 1 技术路线图

2.2 实施步骤

1）确定文献数据库和检索式。本文将专利数据作为

技术路线图的数据源主要出于以下考虑：技术路线图要反映当前某个主题发展的实际状况，而专利文本信息的一个要求就是可操作性，因此凡是在专利文本中体现的主题发展的情况就可以认为超越了实验室研究和理论研究阶段。由于本文的研究目的是更细致地体现某个研究主题的发展现状，因此检索式的确定也必须更具体、专业，需要对某个领域的主要研究方向有大致了解，可以通过专家咨询或通过专业资料查询完成。

2）信息抽取和数据清洗、整理。随着计算机技术和文本挖掘技术的发展，目标信息的自动标注和抽取成为可能，本文将对专利标题和摘要中的关键信息，如研究主题信息、主要技术信息和技术功效信息进行标注和抽取，标题信息能够明确表示专利的主要研究主题，摘要中的信息作为有效补充。

3）关键信息组合的专利表示。从专利的摘要和标题信息中抽取出关键词信息或短语信息，这些信息能够反映专利的主要研究内容、研究结果，有些还能反映主要的方法，通过这几个方面就能够对专利进行比较简要、完整的表示，进一步与领域研究主题的大类进行映射，从而完成专利的主题分类。

4）技术发展路线图构建。技术发展路线图主要展示的是某一个主题及其关键技术在一定的时间区间内的发展现状。本文将从整理后的数据中选择主要的研究方向进行技术路线图构建，将该研究方向研究状况在时间轴上进行展示，包括关键技术的发展程度以及技术应用状况。

5）对比分析。潜在发展空间的确定是将某主题的某个研究方向的技术发展路线图与对应的规划文本进行对比分析后的结果，分析该技术发展规划目标以及响应情况，判断其饱和状况，从而判定潜在研究主题。

3 实证分析

本研究选择 2010—2016 年间纳米技术在能源领域的应用情况作为研究对象，通过检索式为：（nanoporous or nanostructured or nano or nanotube or nanostructure）and energy 查找到的美国公开、美国授权、日本、韩国这几个专利技术申请较为领先的国家为专利来源的专利作为原始数据源，共查找到 413 篇相关专利，通过去重和去除不相关专利，最后得到 207 篇专利数据，抽取专利数据的标题和摘要部分作为数据源，标题部分可以大致显示该专利的研究内容，而通过摘要部分的补充可以精确地确定专利的研究所属能源应用中的哪一个部分。

将专利的标题和摘要部分进行关键信息的标注和抽取，则每一篇专利都可以用一个由关键信息组成的组合结构进行表示，见图 2、图 3。

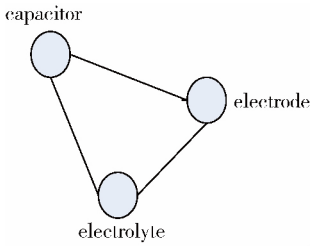

图 2　专利信息组合表示（电容器）

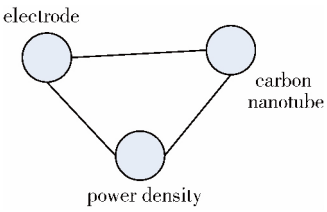

图 3　专利信息组合表示（电极）

图 2 展示的是一项电容器的专利研究，从题目中我们仅能获得“capacitor”的信息，再结合摘要中的“electrode”和“electrolyte”就很容易得知该专利重点在通过电极和电解质方面的改善。图 3 展示的是一项电极方面的创新，但是通过其他的信息我们可以了解到，其中涉及“carbon nanotube”碳纳米管材料的应用，其研究目的是提高能源密度。通过以上方法在对专利进行表示后，与能源领域的主要研究主题进行映射，确定专利在能源领域所属的研究主题，见图 4。

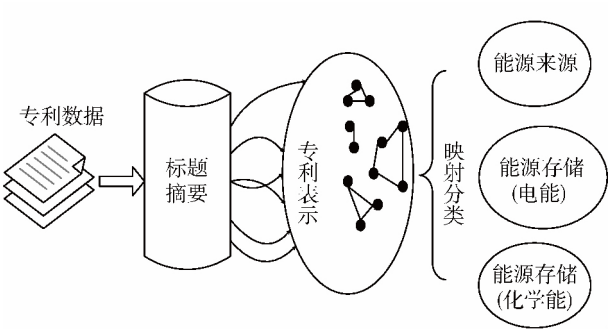

图 4　专利主题分类步骤

通过与当前能源的主要研究主题进行映射可以将数据进行粗分类，207 篇专利中有 47 篇属于有关能源来源的研究，其主要纳米技术包括染敏、量子技术和纳米薄膜技术的应用；154 篇属于能源存储方面主要是电能的研究，主要体现的纳米技术包括电极、电容器和电解质以及锂电池研究；有 6 篇专利属于能源存储中化学能方面的研究，主要是纳米的多孔技术和催化剂技术。通过以上简要的分析，可以看出纳米技术在能源存储方面研究占的比重最大，尤其是对电能的存储上研究最为广泛，从图 5 可以看出纳米技术在该方面的研究在最近几年发展都比较稳

定，因此可以看作一个研究热点，该研究在 2013 年已经达到过一次研究高峰，近两年的研究逐渐减少，说明研究热度降低，但是该主题技术发展情况如何，其饱和程度如何以及还有哪些潜在研究主题还值得进一步挖掘，本文将锁定该主题作为下一步研究对象。

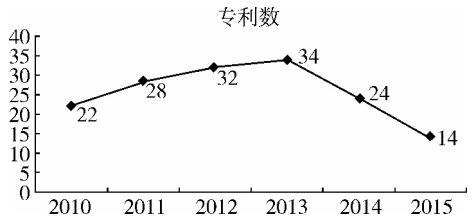

图 5　纳米技术在能源存储（电能）方面的专利数量

将有关纳米材料在电能储存中的应用研究的 154 篇专利进行主要研究主题的发现，将主要的研究热点内容进行展示，见图 6。

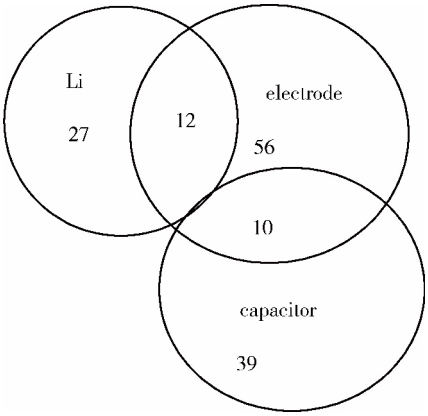

图 6　纳米技术在能源存储（电能）方面的主题分类

纳米技术在能源存储（电能）方面的研究主要分布在锂电池的优化、超级电容器以及电极，这三个方面既有独立改良的研究，同时还存在技术交叉合作研究，即主要通过电极优化传统的锂电池的研究，如图 6 中所示有 12 项研究；提高电极储能密度；以及通过纳米技术电极改善电容器的能源密度，图 6 中所示有 10 项研究。这些基础研究所得的成果在其他各领域都有着广阔的应用前景，现将通过抽取的数据构建这三个基础研究的技术发展路线图，以观测其发展现状，见图 7。

本文将锂电池、电容器以及电极方面的研究现状通过技术路线图的方式进行展示，然后将与对应的规划文本进行对比分析。规划文本的选择对于本文提出的技术预测方法在降低预测风险方面有着重要的影响，本文选择美国国家航空航天局（NASA）发布的规划文本作为对比对象，该机构除了在航空航天领域方面居世界一流外，在材料技术、生物物理研究以及地球学研究处于世界领先水平，NASA 致力于各种服务于航空航天技术发展的研究，并且会在第一时间发布最新最权威的相关技术规划文本。NASA 在 2010 年发布了纳米技术的发展规划，该规划涉及了纳米技术在能源、工程材料以及传感器等方面的应用，本文选择能源部分作为下一步的对比分析对象。如图 7，锂电池电极的发展在 2012、2013 年经历了一个高峰期，之后趋于平缓，该研究目标是在 2016 年左右实现锂电池的纳米电极发展，该研究目标已经得到了充分的响应，其研究成果主要在于改变锂电池的电极纳米材料、提高锂电池的效率和储能密度以及在低温情况下的稳定性。电极方面的研究近几年都比较稳定，有下降的趋势，但是其独立性研究目前主要是通过改造电极内的构造以及纳米新材料的应用，如 CNT、传感性纤维材料以及氧化金属材料来提高电极的储能密度，电极作为电能存储设备的一部分大多数的研究都是与其他技术的合作性研究，NASA 规划文本中是在 2020 年左右实现燃料电池膜电极达到 50% 以上的储能密度，就目前来看其响应的却还比较少，在燃料电池中的应用研究还未出现。电容器方面在 2011 年和 2012 年较多，随后有所减少，近年又有回温的趋势，其研究内容主要是提高电容器的能源密度，目前研究主要是通过纳米多孔材料、碳复合物材料等进行改善，而且这个方向的研究还比较少，与

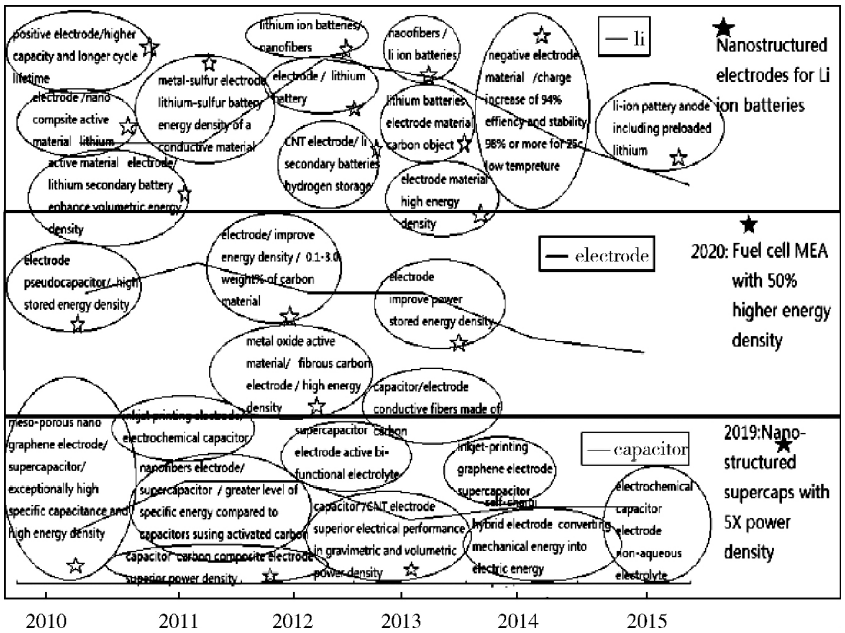

图 7　技术发展路线图

规划目标中2019年左右能源密度增加5倍相比，可以说有所响应，但是并不充分。通过以上分析，对本文研究的三个主题按照发展潜力大小进行排序应是：electrode（电极）方面的研究，尤其是在燃料电池上的应用和改善；capacitor（电容器）方面的研究，尤其是具有超大能源密度的超级电容器的研究；Li电池电极的研究。

4 结论

当前的技术预测方法存在的两个缺陷：①停留在专利外部指标的统计分析，没有深入到专利文本内容中去，技术预测结果也缺乏准确性和细致性；②单纯基于专利文献进行技术预测，预测结果存在一定的风险性。针对这两个缺陷，本文提出了通过文本数据挖掘技术进行专利文本的内容挖掘，构建技术发展路线图，并通过与对应的规划文本进行对比分析的方法，进行技术预测，并就纳米技术在能源存储（电能）应用领域进行了实证研究。该方法有效地解决了研究前沿探测中的两个问题，但是也存在一定的局限，如只是从专利的标题和摘要信息出发，进行了信息的标注和抽取，下一步还可以实现专利全文信息的挖掘，更加丰富地体现技术发展态势；仅从技术主题在专利中体现的研究水平信息（方法和功效等）来判断其研究潜力，没有考虑研究主题的发展时间因素，以及研究人群、国别等社会因素；潜在研究主题的发现需要将技术路线图与相应的规划信息进行对比分析，但是难免会出现信息不对称的情况，因此文本数据挖掘的进一步深化、结合文本数据挖掘技术与社会因素分析在内的技术预测体系构建以及数据源的改善等都是下一步研究的方向。□

参考文献

[1] TÜBKE A，DUCATEL K，GAVIGAN J P，et al. Strategic policy intelligence：current trends. the state of play and perspectives [EB/OL]. [2016-07-15]. http://ipts. jrc. ec. europa. eu/publications/pub. cfm?id=1012.

[2] 李健民. 全球技术预见大趋势 [M]. 上海：上海科学技术出版社，2002.

[3] COATES V，FAROOQUE M，KLAVANTS R. On the future of technological forecasting [J]. Technological Forecasting and Social Change，2001，67：1-17.

[4] 袁冰，朱东华，任智军. 基于数据挖掘技术的专利情报分析方法及实证研究 [J]. 情报杂志，2006（12）：99-104.

[5] 胡安朋. 我国专利活动状况的分析和专利技术的评价及预测 [J]. 情报业务研究，1991，8（2）：65-72.

[6] 张冬梅，曾忠禄. 德尔菲法技术预见的缺陷及导因分析：行为经济学分析视角 [J]. 情报理论与实践，2009，32（8）：24-27.

[7] 徐磊. 技术预见方法的探索与实践思考——基于德尔菲法和技术路线图的对接 [J]. 科学学与科学技术管理，2011，32（11）：37-41.

[8] FYE S R，CHARBONNEAU S M，HAY J W，et al. An examination of factors affecting accuracy in technology forecasts [J]. Technological Forecasting and Social Change，2013，80（6）：1222-1231.

[9] CHEN Darzen，LIN Changpin，MU Hsuan. Technology forecasting via published patent applications and patent grants [J]. Journal of Marine Science and Technology-Taiwan，2012，20（4）：345-356.

[10] JUN S H，PARK S S，JANG D S. Technology forecasting using matrix map and patent clustering [J]. Industrial Management & Data Systerms，2012，112（5）：786-807.

[11] LEE S J，YOON B G，PARK Y R. An approach to discovering new technology opportunities：Keyword-based patent map approach [J]. Technovation，2009，29（6）：481-497.

[12] 王兴旺，汤琰洁. 基于专利地图的技术预测体系构建及其实证研究 [J]. 情报理论与实践，2013，36（3）：51-55.

[13] 赵莉晓. 基于专利分析的RFID技术预测和专利战略研究——从技术生命周期角度 [J]. 科学学与科学技术管理，2012（11）：24-30.

[14] 王旭超，吴腾枫，江小蓉，罗锐戈. 面向技术预测的专利情报分析实证研究 [J]. 情报科学，2014，32（7）：139-144.

[15] 谢学军，周贺来，陈婧. 面向技术预见的专利情报分析方法研究 [J]. 情报科学，2009，27（1）：132-136.

[16] 张韵君，柳飞红. 基于专利分析的技术预测概念模型 [J]. 情报杂志，2014，33（3）：22-27.

[17] 张亚斌，王洵迪. 基于TRIZ理论的物联网关键技术专利发展态势及预测分析 [J]. 系统工程，2015，33（3）：130-136.

[18] 王知津，周鹏，韩正彪. 基于情景分析法的技术预测研究 [J]. 图书情报知识，2013（5）：115-122.

[19] SWART R J，RASKIN P，ROBINSON J. The problem of the future：sustainability science and scenario analysis [J]. Global Environmental Change，2004（14）：137-146.

[20] YOON J，KIM K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection [J]. Scientometrics，2012，90（2）：1-17.

作者简介：王效岳（ORCID：0000-0002-7100-7758），男，博士，教授。

赵冬晓（ORCID：0000-0002-9518-4281），女，硕士生。通讯作者。

白如江（ORCID：0000-0003-3822-8484），男，博士，副教授。

作者贡献声明：王效岳：提出论文的研究思路和框架。

赵冬晓：资料的整理和论文的撰写。

白如江：论文的修订。

录用日期：2016-09-26