

doi:10.3772/j.issn.2095-915x.2015.05.003

基于专利演化模型的机构专利分析

Yunji Jang¹, Jangwon Gim¹, Jinpyo Lee¹, Do-Heon Jung¹, Hanmin Jung(原著)¹,
张津(翻译)², 李颖(编审)²

(1. 韩国科学技术情报研究院; 2. 中国科学技术信息研究所 北京 100038)

摘要: 近年来, 科学技术发展日新月异, 每天都会出现各种新型研究领域, 众多研究成果被发表与公布。为此, 来自各个国家的组织机构做了巨大的努力要为自己的机构带来外部与内部的变化。事实上, 对于很多机构来说, 研究的意义在于通过分析研究成果得到有意义的结论。因此, 本文通过对具体机制下的专利标题(发明名称)进行分析从而提出一种演化模型。首先, 我们根据关键词的特性对标题的关键词进行分类, 然后定义专利的相互关系。我们建议演化模型的相互关系建立在时间轴基础上, 并且可以适用于实际数据。通过应用于实际数据的模型可预测出未来专利的关键词。

关键词: 专利数据, 专利分析, 专利关键词, 专利演化

Patent Analysis for Organization based on Patent Evolution Model

Yunji Jang, Jangwon Gim, Jinpyo Lee, Do-Heon Jung, Hanmin Jung¹, Zhang Jin², Li Ying²

(1. Korea Institute of Science and Technology Information, KISTI;

2. Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: With the rapid development in science and technology in recent years, new research fields are being discovered and studied each day and numerous research findings are being published and presented. Considering this, organizations from various countries have been investing considerable efforts to carry out internal and external changes in their organizations. In fact, in many organizations, studies are being carried out to derive

编者寄语: 本文为韩国科学技术情报研究院 Yunji Jang 等授权提供。

基金支持: 本文为中央级公益性科研院所基本科研业务费专项资金“数字资源标识与服务标准规范及关键技术研究”项目支持。

原文请在《情报工程》期刊网站 <http://tie.istic.ac.cn/> 下载。

meaningful results by analyzing research outcomes. We thus in the paper propose an evolution model through analysis of the patent titles from one specific institution. First, we classify the keyword of title according to properties of keyword and then define the relation case of patent. We suggest the evolution model of relation based on timeline and applied to the actual data. It can predict keyword of future patent by applying to actual data.

Keywords: Patent Data, Patent Analysis, Patent Keyword, Patent Evolution

1 前言

近年来，科学技术发展日新月异，每天都会出现各种新型研究领域，众多研究成果被发表与公布。在一些国家中，如韩国、美国和日本，各种加强保护知识产权的政策已经开始实施。此外，研究成果的专利申请数量也处于增长状态^[1]。为此，来自各个国家的组织机构做了巨大的努力要为自己的机构带来外部与内部的变化。换言之，组织机构认为如果不进行创新，他们很难在当今的竞争中存活，因此，必须在这方面付出巨大努力^[2]。事实上，对于很多机构来说，研究的意义在于通过分析研究成果来得

到有意义的结论^[3,4]。通常情况下，在这些研究成果中，专利数据是一种典型的研究成果，可以作为评测某组织机构技术和创新能力的指标。专利不仅在产品研发的版权与发行问题上有着至关重要的作用，同时也有助于未来的研究与开发计划^[5]。

专利数据包括专利标题、技术实施细节、技术范畴码、引证信息以及申请人信息。分析这些专利信息非常重要，有助于诠释技术趋势、水平和商业价值方面的变化。专利分析方法多种多样，例如频率分析、共享分析、时间序列分析、引证分析和权利分析。其中，时间序列分析和二维分析尤为普遍^[6]。

表 1 专利分析方法的类型

类型	细节
频次分析	国家、发明人、申请人、技术分类……
共享分析	国家、发明人、申请人、技术分析、详细技术分类……
时间序列分析	申请率分析、国籍分析、发明人分析、申请人分析、技术分析、最新申请分析、新发明人分析……
关联分析	发明人关联示意图、申请人关联示意图，技术关联示意图……
引证分析	引证相关性分析、核心专利分析……
专利权分析	专利家族图谱……

现在一些研究尝试着分析各公司专利引证关系的特点，例如，试图通过引证关系来决定竞争公司技术战略研究项目的数量与日俱增^[7-10]。

这些研究主要分析公司已经申请的专利信息；但是，据我们所知，目前还没有研究对专利数据采用时间序列分析。因此，在本文中，我们提出了基于时间序列分析的专利演化模型。此外，与之前的文献不同，在本文中，我们只分析标题，因为在众多的专利数据中，专利标题可以充分起到介绍专利特征信息的作用。

2 专利分析模型

在本文中，我们采用四个阶段分析了专利标题演化过程。专利分析的过程如图1所示。首先，将无用、无意义的词语从专利标题中剔除，剩下的便是可以代表专利核心技术和专利目的的词语和关键词。然后，通过绘制关系网络图来映射关键词的关系类型，这种关系类型的演化是通过时间序列分析来确定。

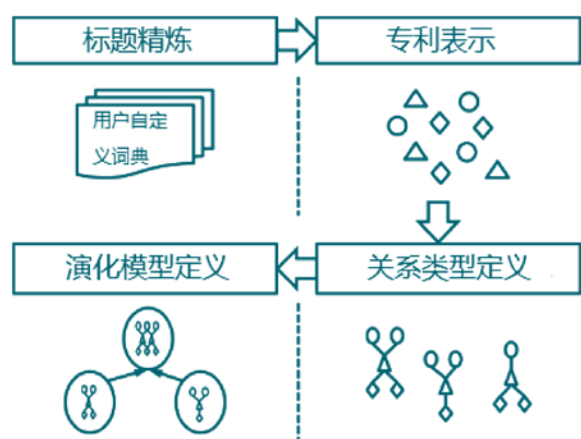


图1 专利分析模型流程

2.1 专利标题精炼

精炼专利标题这一阶段是在提取有意义的关键词之前进行，比如目的与方法。在这一阶段，多余的词语将会从专利标题中被剔除。根据空格，

切分专利标题，考虑被拆分词语的统计数量，他们将被一一剔除。

2.2 专利表示

针对某一特定组织机构观察其基于关键词概念的演化过程，精炼标题中的词语被分为方法、目标对象、目标谓词三大类。目标代表了阐述专利发明目的关键词。当某项发明的名称被命名为专利标题后，我们就可以得到反映核心技术的目标关键词。根据专利的类型，方法关键词被用于描述发展目标技术的核心技术的关键词。在本文中，我们构建了一种将关键词分为方法和目标两大类的词典。已区分的方法和目标关键词将通过预先构建的词素分析工具被标记为对象和谓词。当词素分析完成时，方法、目标对象和目标谓词之间关系类型将会以关系网络图的方式绘制出来。

2.3 关系类型定义

针对两个或更多专利标题间的专利表示关系类型，本文定义了所有类型。本文中定义的关系是指那些在三种关键词之间有一个或多个关键词重叠的实例，这三种关键词被区分成方法、目标对象、目标谓词。

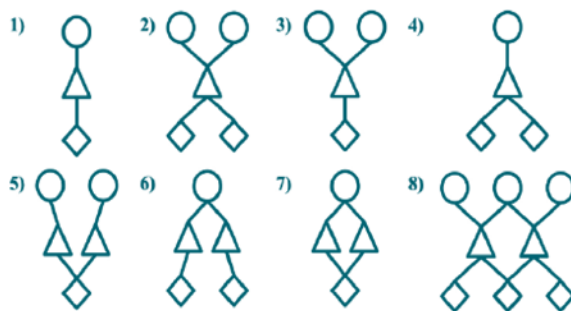


图2 专利关系类型

如图2所示，关系类型可以在方法、目标对象、目标谓词的基础上被推导出来。圆圈代表方法，三角代表目标对象，菱形代表目标谓词。图2-1)表示一种专利类型。图2-2)是X型的关

系类型。X 型是指若干方法和目标谓词映射到一个目标对象上。图 2-3) 的关系类型是 Y 型。Y 型是指若干方法映射到单一目标对象和目标谓词上。图 2-4) 是倒转 Y 型。倒转 Y 型是若干谓词映射到单一方法与目标对象上。图 2-5) 是 V 型。在 V 型类型中, 若干目标对象和若干方法映射到一种目标谓词上。图 2-6) 是倒转 V 型。这一类型是指若干目标对象和谓词映射到单一方法上。菱形关系是若干目标对象映射到单一方法和目标谓词。最后是图 2-8) 所示的双 X 型, 是含有若干方法、目标对象、目标谓词的关系类型。上述类型中, X 型可以演化成多种类型, 但在本文中只对双 X 型进行研究。

2.4 演化模型定义

本节内容对演化模型进行定义。演化模型可以反映基于特定组织机构的时间序列数据的关系类型的特点。图 3 展示了可根据关系类型进行演化的模型。列举一个演化模型的例子, “A 型演化成 B 型” 是指 B 型满足当 A 型和 B 型合并时的情况。换句话说, 当 A 型和 B 型合并时, 即 B 型。

所有类型都可以演化为另一种类型。∅ 型和 I 型可以演化成所有相关类型。此外, 所有关系类型都可以演化为与之相关的类型。但是, Y 型和 Λ 型可以演化成 X 型。

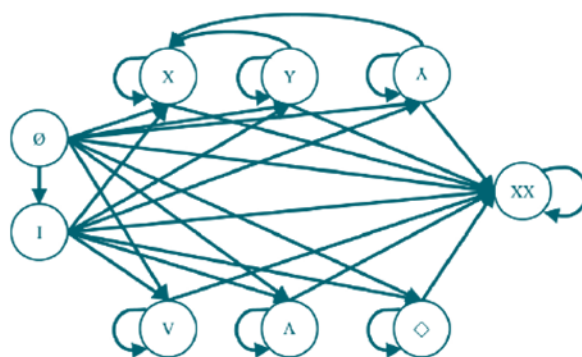


图 3 演化模型

3 实验

3.1 数据集

数据集由 82 种专利组成, 这些专利数据是由韩国科学技术情报研究院 (KISTI) 计算机智能实验室提供的 2005 年到 2013 年间的专利数据。最初我们共选出 99 组数据, 但其中 17 组标题因为重复被排除。

3.2 用户自定义词典

整个过程包括语句结构的分析和标记。在专利标题集合的基础上, 构建了数个词典。通过认知分析建模来确定相似的结果。表 2 和表 3 是用语将韩文翻译成英文的词典。表 2 展示了用于提炼无意义词语词典的一部分。

表 2 语义停用词词典

Semantic Stopword (语义停用词)	General Stopword (普通停用词)
System and Method (系统和方法)	Optimum (最佳的)
Apparatus and Method (设备和方法)	Effective (有效的)
Method and System (方法和系统)	For (为了)
System and Apparatus (系统和设备)	About (关于)
Framework (架构)	Included (包括)
System (系统)	At (在)
...	...

表3 辨析词词典

A-G Distinguishing word (A-G 辨析词)	A-G Distinguishing word (A-G 辨析词)
Based on (基于)	Applied (应用的)
Based (基准)	Through (通过)
Using (使用)	By (通过、被)
Centered (居中)	According to (根据)

表3列举的是术语词典，用来区分方法和目标。区分出方法和目标后，剩下的词语就被认为是无用词。事实上，韩语词的数量多于英语。因此，表3中词典词语数量少于韩语词典中的词语数量。

3.3 统计

这里显示了将非必要词词典应用于82个实例

的统计过程。目标对象和目标谓词在上述82个实例中全部被标引，但方法仅在44个中被标引。我们对关系类型和演化模型进行统计。对2005年至2013年间关系案例的累积分析如图4所示。其中X型最常见，其次是V型。

表5是演化模型的统计。其中，从 \emptyset 型向I型的演化最为常见。这种演化方案意味着KISTI发表了新专利。

表4 关系类型统计

	Base	X	Y	人	V	△	◇
2005	1	0	0	0	0	0	0
2006	0	1	0	0	0	0	0
2007	1	1	0	0	1	0	0
2008	3	1	0	0	0	0	0
2009	2	3	0	0	1	1	0
2010	2	4	1	1	4	2	0
2011	2	4	2	1	5	3	1
2012	3	5	3	1	5	3	1
2013	3	6	3	1	5	3	1
合计	3	6	3	1	5	3	1

表5 演化模型统计

	I → XX	I → X	X → X	X → XX	XX → XX	Y → XX	\emptyset → I	\emptyset → V	\emptyset → Y
2005	0	0	0	0	0	0	1	0	0
2006	0	1	0	0	0	0	0	0	0
2007	0	0	0	1	0	0	1	0	0
2008	0	0	0	0	0	0	2	0	0
2009	1	1	0	0	1	0	2	1	0
2010	0	1	0	1	3	0	1	0	1
2011	0	0	2	1	0	1	0	0	0
2012	0	1	1	0	1	0	1	0	0
2013	0	0	0	0	1	0	0	0	0
合计	1	4	3	3	6	1	8	1	1

3.4 结论

在结论部分，我们将本文中演化模型与实际结果进行了比较。

2013年度方法、目标对象与目标谓词间的最终关系网络图如图4所示。网络图使用NodeXL^[11]绘制完成。如图4所示，关系类型的演化是通过绘制2005年至2013年专利标题的关系网络图来实现的。图4和图5是KISTI专利的演化模型。



图4 KISTI专利的关系网络图

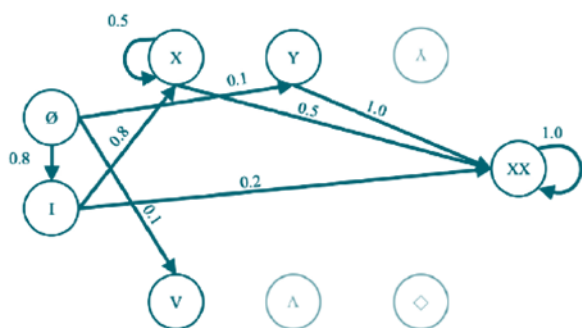


图 5 KISTI 演化模型

在 KISTI 专利的 30 种演化模型中可被发现的模型有 9 种。图 5 中的数字表示向箭头方向演化的概率。I 型、Y 型、V 型都是由 Ø 型演化而来。Ø 型演化成 I 型的概率是 0.8，演化成 Y 型和 V 型的概率都是 0.1。XX 型是由 X 型和 I 型演化而来。I 型演化成 X 型和 XX 型的概率分别是 0.8 和 0.2。X 型可演化成 XX 型，亦可演化成其本身。两者的概率皆为 0.5。XX 型是唯一能由 Y 型演化而来的类型，XX 型也可以演化成其本身。在 KISTI 专利中，倒转 Y 型、Λ 型和◇型没有被关注。基于以上研究结果，可以通过预测 KISTI 专利的未来演化模型将提取的关键词与演化类型搭配起来。例如，当专利由 I 型演化成了 X 型，我们就可以预测出 KISTI 将会对目标对象进行研究。

4 总结

在本文中，通过对特定机构的专利标题分析提出了专利演化模型这一概念。通过目前并不活跃的韩语标题的研究，我们揭示了一种新的可能性。我们剔除了特定机构专利标题中的停用词，然后对方法、目标对象和目标谓词进行区分。通过区分出的三种关键词，可在时间序列基础上描绘出关系网络图。可通过这种关系网络图定义指定年份里可能出现的专利关系类型。根据这个实验结果，该组织机构的关系实例即可明确。

目前，专利分析系统和相关的分析方法主要

依靠基础性可视化技术以及专利地图来进行，例如柱状图、饼形图、表格以及气泡图。专利间的关系类型和关系的演化模型的建议仅适用在了专利标题。

在今后的研究中，我们计划将相同的分析实验用在不同的组织机构中，然后与本文中的结论进行对比；下一步，我们计划验证本文模型的可扩展性。我们旨在进一步开发这种专利演化模型，以便它能被用于除了韩国之外的其他国家的专利标题中。此外，我们希望通过预测专利关系类型，将它用于融合技术预测中去，特定关系类型的专利关系就可以通过演化模型进行演化了。

参考文献

- [1] Sang-cheol Kil, Sung-min Kang, "The study of an Analysis on Patent Management Affecting the Company Performance," In Proceedings of the Korea Technology Innovation Society 2008, Vol. 11, No. 2, pp. 171-193, 2008.
- [2] SeongTaek Park, Youngki Kim, Taeung Kim, "A Study on Influencing Factors of Patent Activities on Magnagement Performance," In Journal of Entrue Journal of Information Technology 2013, Vol. 12, No. 3, pp. 121-129, 2013.
- [3] Hwan Lim, Hosun Rhim, Myungho Sohn, "Study for Improving the Patent Management Schema by Using Citation Index in Public-sector R&D Institute : Case Study on K Institute," In Journal of the Korea Society of IT services, Vol. 7, No. 3, pp. 115-127, 2008.
- [4] Mu-hsuan Huang, Li-yun Chiang, Dar-zen Chen, "Constructing a patent citation map using bibliographic coupling : A study of Taiwan's high-tech companies," In Proceedings of the Scientometrics , Vol. 58, No. 3, pp 489-506, 2003
- [5] Sunghae Jun, Sang-Sung Park, Young-Geun Shin, Dong-Sik Jang, HoSeok Chung, "Forecasting Vacant Technology of Patent Analysis System using Self Organizing Map and

Matrix Analysis," In Journal of the Korea Contents Association , Vol. 10, No. 2, pp. 462-480, 2010.

[6] Jin-Hyo Joseph Yun, Oh-Jin Kwon, Jinseo Park, Eui-seob Jeong, "A Study on the Development and Adaption of Open Innovation Analysis Model," In Proceedings of the Korea Technology Innovation Society 2010, Vol. 13, No. 1, pp. 99-123, 2010.

[7] Aidong, Peng. "The Application of Patent Citation Analysis in Enterprise Competitive Intelligence," In Proceedings of the Information Studies: Theory & Application, Vol. 27, No. 3, pp no. 16, 2004.

[8] Lingyun Sun, Yu Song, "Research on Clustered Patent Mapping Visualization and Interaction," In Proceedings of the Computer-Aided Industrial Design and Conceptual Design,

pp. 1130-1133, 2008.

[9] Jinho Choi, Junguk Kim, "Analysis of Assortativity in the Keyword-based Patent Network Evolution," In Journal of Internet Computing and Services, Vol. 14, No. 6, pp. 107-115, 2013.

[10] Xian Wu, Li Zhang, " SIMPLE: A Strategic Information Mining Platform for Licensing and Execution," In Data Mining Workshops, pp. 270-275, 2009.

[11] Marc Smith, Ben Shneiderman, Natasa Milic-Frayling, Eduarda Mendes Rodrigues, Vladimir Barash, Cody Dunne, Tony Capone, Adam Perer, Eric Gleave, "Analyzing (Social Media) Networks with NodeXL," In Proceedings of the fourth international conference on Communities and technologies, pp. 255-264, 2009.