

基于语义 TRIZ 的专利技术挖掘研究*

胡正银^{1,2} 方曙¹ 张娴^{1,2}

1 (中国科学院成都文献情报中心 成都 610041)

2 (中国科学院大学 北京 100049)

【摘要】在介绍语义 TRIZ 概念、分析语义 TRIZ 与专利技术挖掘关系及不足基础上, 本文提出了一个针对特定领域专利, 构建个性化语义 TRIZ 的框架。该框架从概念空间、索引空间、应用空间三个维度构建语义 TRIZ; 通过微观层 SAO 基础语义单元、中观层 P&S 本体、宏观层技术范畴三个层面描述语义 TRIZ 索引结构。基于该语义 TRIZ, 本文分析了典型专利技术挖掘应用场景; 并选择大口径光学元件专利, 进行了技术主题聚类实证研究。结果显示: 对中等规模专利数据, 本方法能有效构建语义 TRIZ, 实现专利技术挖掘应用。

【关键词】语义 TRIZ; 专利分析; 技术挖掘; 主题模型

Patent Tech Mining based on Semantic TRIZ

Hu Zhengyin^{1,2} Fang Shu¹ Zhang Xian^{1,2}

1(Chengdu Document and Information Centre, Chinese Academy of Sciences, Chengdu 610041)

2(University of Chinese Academy of Sciences, Beijing 100049)

【Abstract】: This paper proposes a personalized semantic TRIZ framework oriented to specific domain patents based on analysing the disadvantage of existed semantic TRIZ. This framework construct semantic TRIZ from three dimensions of concept space, index space and application space. And it describes the structure of semantic TRIZ based on micro-meso-macro levels. Based on this semantic TRIZ, this paper analyses some typical scenarios of patent tech mining. Finally, we select the field of large aperture optical elements for empirical analysis, and cluster the problems and solutions technology topics. The result shows that this new method can effectively construct semantic TRIZ and better support patent tech mining.

【Keywords】: Semantic TRIZ, Patent Analysis, Tech Mining, Topic Model

基金项目: 中国科学院知识产权信息服务专项(KFJ-EW-STS-032)及中国科学院西部之光项目“基于本体的专利文献技术挖掘系统研究与实践”。

作者简介: 胡正银 (1979-), 男, 四川成都人, 中科院成都文献情报中心副研究员, 在读博士研究生, 主要从事专利分析、文本挖掘研究。

方曙 (1957-), 男, 四川成都人, 中科院成都文献情报中心研究员, 情报学博士生导师, 主要从事知识产权情报研究、科技战略研究。

张娴 (1973-), 女, 四川成都人, 中科院成都文献情报中心副研究员, 在读博士研究生, 主要从事知识产权情报研究。

电子邮箱: huzy@clas.ac.cn

导师姓名: 方曙

1 概念

技术挖掘(Tech Mining)是本世纪初,美国学者 Alan Porter 等提出基于历史科技文献分析当前和未来技术发展现状与趋势的理论与方法, 它的研究范畴包括: 技术演化、技术评估、技术预测、科技指标等^[1]。专利文献具有内容新颖、系统详尽、格式规范、分类科学等特点, 是技术挖掘使用最多的信息源^[2]。专利技术挖掘内涵丰富, 从宏观管理层面上看, 可应用于面向未来技术预测等; 从中观执行层面上谈, 可帮助研发机构获取技术竞争情报信息; 也可微观层面具体技术研发提供信息服务^[1,3]。

TRIZ 是发明问题解决理论(Theory of Inventive Problem Solving)的俄文缩写, 是苏联发明家 Altshuller 等分析 200 多万份专利, 从中归纳出具有共性的原则与方法^[4]。TRIZ 理论内容十分丰富, 矛盾(Contradictions)与原则(Principles)是其重要内容之一。TRIZ 中的“矛盾”指: 发明中解决的各种技术难题(Problems), Altshuller 等将其抽象归纳成为 1201 个标准工程问题, 即: Contradictions; TRIZ 发现大量的技术问题都可采用矛盾冲突矩阵(Contradictions Matrix)来进行描述。解决这些矛盾冲突的解决方案(Solutions)也经常在不同技术领域重复出现, 它们被抽象归纳成 40 个标准解决方案, 即: Principles^[4-5]。基于矛盾与原则的专利技术挖掘, 可帮助用户快速发现采用了相似的发明原理或解决了相似技术难题的专利, 促进专利信息有效利用^[6]。但随着专利数据越来越庞大, 覆盖技术领域越来越广泛, 直接利用 TRIZ 归纳标准矛盾与原则进行技术挖掘存在较大的缺陷, 主要表现为: 首先, 它们主要依靠分析机械、工程类专利得出的结果, 无法很好反映信息技术、生命科学、材料科学等领域专利的特性; 其次, 它们是从专利中总结出来的一般性原则与规律, 难以直接映射到某一具体专利; 再次, 矛盾与原则之间的语义关系, 没有被很好地揭示; 最后, 传统 TRIZ 中, 矛盾与原则长时间结构保持稳定, 难以描述目前新增海量专利的技术特征^[4-5]。

基于上述不足, 语义 TRIZ(Semantic TRIZ)应运而生。语义 TRIZ 是在专利数量与日俱增大背景下, 基于传统 TRIZ 中矛盾与原则思想, 采用现代语义技术对海量专利中隐含技术信息进行自动或半自动语义索引^[7], 以深入揭示其技术内涵。语义 TRIZ 中, 标准化矛盾与原则还原成具有动态结构的技术难题与解决方案(Problems & Solutions, P&S), Problems 与 Solutions 之间语义关系得到充分描述与揭示。技术领域不同, 专利数量不同, P&S 的结构也会随之变化。

总之, 语义 TRIZ 核心是对专利语义索引; 索引重点是 P&S 信息。语义 TRIZ 通过对专利中隐含的 P&S 信息进行语义索引, 拓展了传统 TRIZ 应用范围、深化专利技术挖掘的程度、更新专利技术挖掘的角度。

2 文献综述

2.1 语义 TRIZ

虽然传统 TRIZ 研究历史悠久, 并取得众多研究成果, 但语义 TRIZ 研究尚处于起步阶段。2004 年, 发明机器公司(Invent Machine, Inc.)正式提出了语义 TRIZ

概念，并注册了 Semantic TRIZ™ 商标^[7]，但尚没有统一关于 Semantic TRIZ 的标准。发明机器公司定义了 Semantic TRIZ 基本模式：基于 TRIZ 元素对专利语义索引、基于 TRIZ 语义索引的知识平台、自然语言服务界面^[7]，并在其旗舰产品 GoldFire 中实现了上述部分功能^[8]。CREAX 公司通过深入分析 8000 万条专利，定义了一套 CREAX 创新方法学，提供面向 Problems-Solving 的知识挖掘服务^[9]。IBM 印度研究院利用语义技术对生物医药类专利语义索引，开发了 BioPatent Miner 系统，实现了基于生物医药术语与关系的语义检索^[10]。韩国学者 Yongho 等人研究了利用句法分析与模式匹配技术，从专利中自动抽取语义 TRIZ 核心要素 Problems 与 Solutions 的方法，并进行了小规模实证^[11]。Kim 等人则进一步探讨了如何利用主-谓-宾式三元组 Subject-Action-Object (SAO) 来描述 Problems 与 Solutions^[12]。

在语义 TRIZ 方面，商业公司虽然已推出了相关产品与服务，但外界无从知晓其具体技术细节。而且，其产品一般是对整体技术领域专利进行面向 TRIZ 的语义索引，索引结构简单，粒度较粗，不能满足个性化深度专利技术挖掘需求。非盈利科研机构学者的研究则比较零散，不成体系；且实证研究数据规模很小。

2.2 专利技术挖掘

根据挖掘对象不同，专利技术挖掘可分为：专利元数据字段挖掘与专利文本挖掘两种。前者无论是方法论，还是技术手段上都比较成熟；后者是研究重点。专利文本挖掘的典型应用场景包括：技术主题聚类、专利自动分类与技术演化分析、技术预测等^[13]。

专利技术挖掘中，很少对专利本身进行聚类，多是对包含的技术主题进行聚类。Yuen-Hsien 等对技术主题聚类进行了系统研究，归纳出流程如下：关键词选取、权重计算、相似度计算、聚类算法选择、多步骤聚类、生成聚类簇标签、对聚类结果进一步分组等^[14]。Young 等在关键词聚类的基础上，计算关键词在不同聚类簇之间的数量分布，结合专利申请时间，分析出代表新兴技术的关键词^[15]。

在专利自动分类方面：He 与 Loh 系统研究了基于 TRIZ 40 条发明原理进行专利自动分类，包括：选取其中 6 条发明原理，利用多种分类算法进行的小规模专利自动分类^[16]；分析 40 条发明原理之间的相似性，对其进行重新分组^[17]；通过分析发明原理的专利句法信息，归纳出相应的句法与语法模式，基于关联规则进行面向发明原理的分类^[18]。Zhen Li 等人则基于 TRIZ 发明级别进行了专利评估与分类研究^[19]。

技术演化分析技术主题产生、发展、突破创新、转移和变化乃至湮灭的过程，是专利技术挖掘的重要内容。Janghyeok 等人研究了利用 SAO 模型分析技术演化趋势^[20-21]。Hyunseok 等将 SAO 模型与 TRIZ 演化模型结合起来，以此为基础分析技术演化历程^[22-23]。在技术预测方面：Yi 与 Alan 等人基于 GoldFire 提供的语义 TRIZ 数据，结合技术路线图，研究了如何发现定位与发现潜在新兴技术^[24-25]。

根据技术表示粒度不同，可将专利技术挖掘分为基于关键词与 SAO 两种。基于关键词的技术挖掘主要集中在中观或宏观层面，难与 TRIZ 结合起来；而 SAO 可作为语义 TRIZ 的基础语义单元，对专利深度语义索引，是目前研究的热点^[13]。

3 面向 TRIZ 的专利语义索引

专利语义索引的粒度与结构决定了各场景技术挖掘的深度、广度与维度，而现有语义 TRIZ 产品索引粒度与结构不足以满足个性化深度专利技术挖掘需求。鉴于现有研究不足，参照发明机器公司的 Semantic TRIZ 基本模式^[7]，本文从概念空间、索引空间、应用空间三个维度，提出了一种针对特定领域个性化技术挖掘需求、面向 TRIZ 的深度专利语义索引框架。该框架从 SAO、技术难题与解决方案、技术范畴三个层面对索引结构进行描述，这三个层面概念说明见表 1。

表 1 面向 TRIZ 语义索引结构概念说明

概念名称	概念层级	信息类型	概念说明
SAO	微观层面	显性信息	SAO 以主-谓-宾的形式直观显示了专利所包含的具体技术信息
P&S	中观层面	隐性信息	意义相近,关系密切一组 SAO 归纳成一对通用 P&S
Tech	宏观层面	隐性信息	多个专利中频繁共现的 P&S, 抽象成更具一般意义的技术范畴

该框架说明如下：首先，基于领域核心专利集构建语义 TRIZ 概念空间，得到领域核心 P&S 本体；其次，针对待挖掘领域专利数据集，基于领域核心 P&S 本体，构建语义 TRIZ 索引空间；最后，针对具体技术挖掘应用，基于面向 TRIZ 专利语义索引，构建语义 TRIZ 具体应用空间。具体如图一所示。

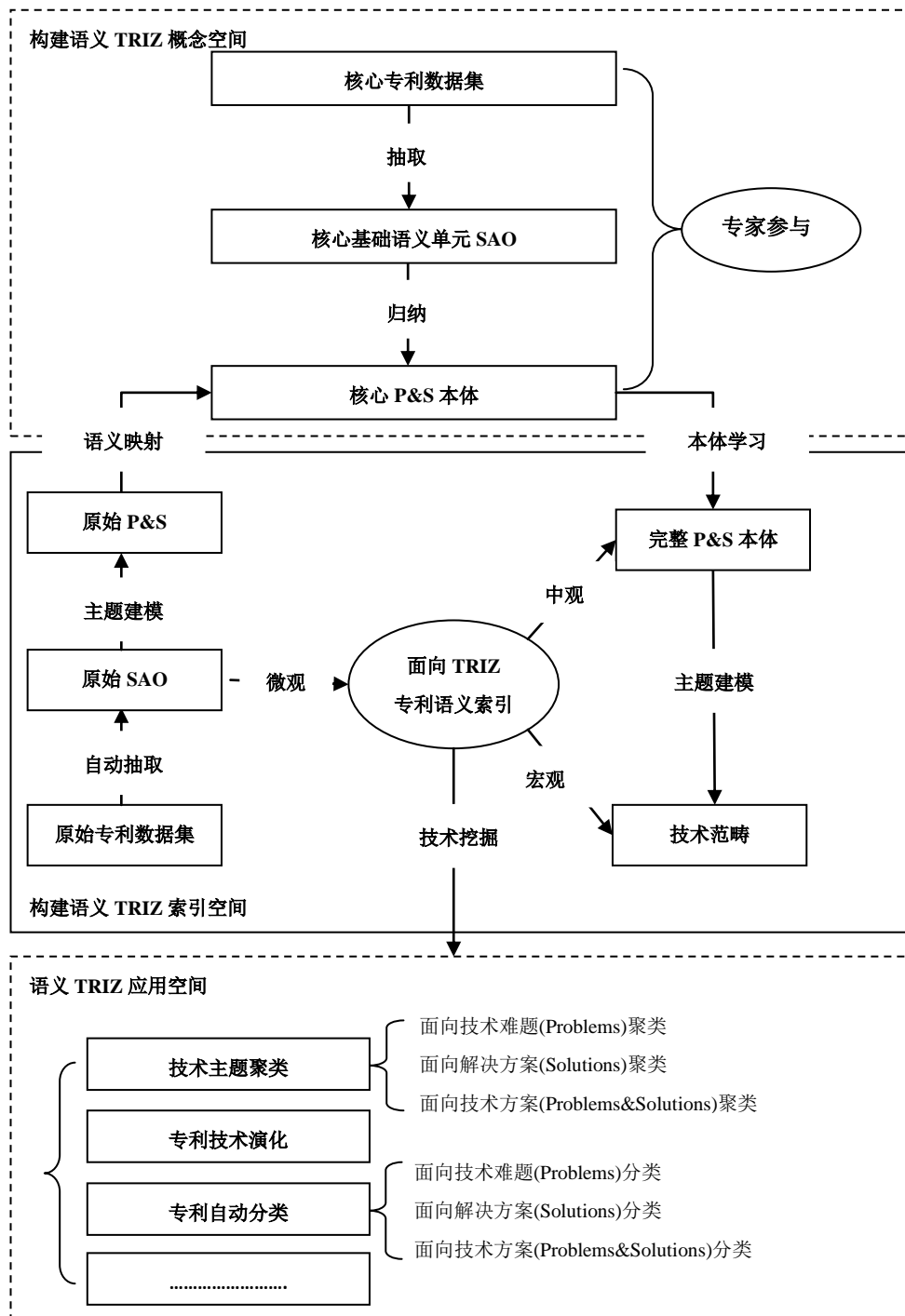


图 1. 面向 TRIZ 语义索引框架图

3.1 构建语义 TRIZ 概念空间

概念空间是领域核心概念组成的集合，它是语义 TRIZ 的核心概念集。概念空间的概念来自于领域核心专利数据集，这些概念包含两个层次：微观层的核心基础语义单元 SAO 与中观层的领域核心 P&S，其中 P&S 之间的关系采用本体进行描述。

首先，借鉴核心专利评估指标体系^[26]选取领域核心专利。然后，采用关系

抽取工具如 Reverb 等^[27]从核心专利集中抽取 SAO 作为基础语义单元。接着，利用 LDA 主题模型^[28]并结合句法分析、模式匹配等方法将其归纳为 Problems 和 Solutions。最后，邀请领域专家对数据进行校验、审核；确定 P&S 之间语义关系，得到领域核心 P&S 本体。

根据专利挖掘实际情况，归纳出P&S之间基本语义关系见表2与表3。

表2 Problems 之间基本语义关系

关系	关系定义
Super	Problem A 是 Problem B 的上位 Problem
Sub	Problem A 是 Problem B 的下位 Problem
Partof	Problem A 是 Problem B 的子集
Co-Relation	Problem A 与 Problem B 共享同一 Solution

同理，Solutions 之间也存在类似语义关系。

表3 Problems 与 Solutions 之间基本语义关系

关系	关系定义
Total Solution	Solution A 是 Problem B 的全面解决方案
Sub Solution	Solution A 是 Problem B 的下位 Problem 的全面解决方案
Part Solution	Solution A 是 Problem B 的子集 Problem(Partof) 的全面解决方案

在建立较大规模语义 TRIZ 索引时，概念空间中领域核心 P&S 本体既是自动构建完整 P&S 本体学习的“训练集”，又是其语义映射的“参照集”。核心 P&S 本体的规模要恰当，既要能满足实际技术挖掘需求，又要能保证后续较大规模专利语义自动索引的准确性。复杂的领域核心 P&S 本体，会导致后期自动索引难度加大，需要大量人工干预；而过于简单的 P&S 本体则无法满足真实技术挖掘需要。

3.2 构建语义 TRIZ 索引空间

索引空间是针对整个领域专利数据的语义索引。它从微观层 SAO、中观层 P&S 本体、宏观层技术范畴三个层面对专利进行索引。

首先，从原始专利中抽取原始 SAO，将专利表示成 SAO 词袋子模式。然后，利用 LDA 主题模型生成原始 P&S。接着，原始 P&S 往领域核心 P&S 本体语义映射，得到中观层的完整 P&S 本体。相比领域核心 P&S，原始 P&S 是从大量专利中通过 LDA 自动生成的。这些 P&S 一般数据量庞大，而且语义发散。如果不进行必要的清洗与映射，使其收敛，那么其提供的技术信息依旧是杂乱无章，难以利用的^[29]。原始 P&S 往领域核心 P&S 语义映射是一个从庞杂、发散的原始数据集中发现、归纳出有意义 P&S 的过程，是一个本体学习、语义降维的过程。映射前，原始 P&S 需要进行必要的清洗；映射中，原始 P&S 往核心 P&S 定向语义降维。定向语义降维的本质是一种分类。映射后，将那些不能准确映射的 P&S 进行降维后与核心 P&S 合并。最后，利用 LDA 对 P&S 主题进一步挖掘，自动生成一系列更宽泛主题，由专家将这些主题抽象成宏观层技术范畴。

语义索引结构构造完成之后，利用 LDA 计算专利与技术范畴、P&S、SAO 语义单元之间的条件概率，得到本领域专利的语义 TRIZ 索引。专利语义 TRIZ 索引结构示意图图二所示。

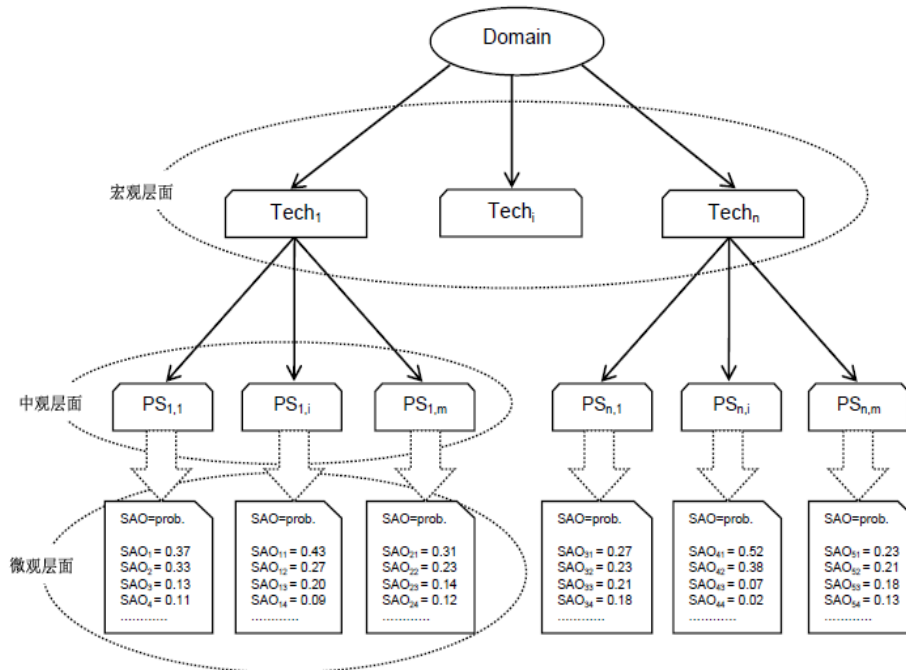


图 2 专利语义 TRIZ 索引结构示意图

构建语义 TRIZ 索引空间的关键是：原始 P&S 往领域核心 P&S 本体的语义映射。一般来说，原始 P&S 不可能做到严格的一一映射，即：既存在 1 对多的映射、也有多对 1 的映射，还会有一定量的 P&S 无法准确映射。为了保证索引覆盖专利集绝大部分技术领域，需要将那些不能准确映射的 P&S 进行降维处理，然后与核心 P&S 共同组成完整的领域 P&S 本体。

3.3 语义 TRIZ 应用空间

应用空间是基于语义索引进行具体专利技术挖掘的中间或最终结果。这些结果本身又可以作为语义 TRIZ 的应用空间，丰富语义 TRIZ 的内容。语义 TRIZ 的应用空间是一个开放性的体系，可以将各种不同类型、不同层级的技术挖掘应用结果，作为应用空间的一部分，叠加到语义 TRIZ 中。如利用语义索引可对领域专利进行面向技术难题或解决方案技术聚类。而聚类结果，可以作为语义 TRIZ 的应用空间支持其它专利技术挖掘应用。

4 基于语义 TRIZ 的专利技术挖掘应用

本文选择大口径光学元件(Large Aperture Optical Elements, LAOE)专利进行构建语义 TRIZ 研究。基于该语义 TRIZ，进行了技术主题聚类实证研究。聚类是一种无监督学习的过程，不依赖先验知识，自动按照数据自身蕴含的规律对其进行归类划分。基于语义 TRIZ 的主题聚类可进一步对专利涵盖的 Problems、

Solutions 进行自动分析，深入揭示技术主题的分布、技术主题之间关联等信息。

数据源方面，我们选择德文特专利(Derwent Innovations Index, DII)作为专利检索数据库，邀请专家制定检索策略，选择 2000-2011 申请年作为检索时间，共得到 1364 条专利。

首先，从中挑选了 284 条核心专利来构建语义 TRIZ 概念空间。概念空间包括：382 个核心 SAO、32 个核心 P&S。核心 P&S 之间语义关系示例见表 4。

表 4 Problems 与 Solutions 之间语义关系示例

Problems	Solutions	关系
Test spherical optical surface	Offset parallel diffraction apertures	Total Solution
Measure object rough surfaces	Interferometer assembly	Sub Solution
Monitor surface quality	Optical systems focus reference beams	Part Solution

接着，基于全体专利数据，构建领域 TRIZ 索引空间。索引空间包括：2372 个 SAO、124 个有效 P&S 主题与 4 类技术范畴。语义索引结构示例见表 5。

表 5 LAOE 语义 TRIZ 索引空间

序号	技术范畴	P&S 主题	SAO
C1	Measuring surface shape	P&S ₁₀ ($p=0.557$)	check large lens convex surface;
		P&S ₁₁ ($p=0.213$)	measure surface roughness;
		P&S ₁₄ ($p=0.117$)	analyze object surface profile;
	
C2	Surface measuring method	P&S ₁ ($p=0.628$)	method measure diffraction;
		P&S ₃₉ ($p=0.124$)	method measure optical curvature;
		P&S ₁₁₄ ($p=0.017$)	method analyze interference-fringe;
	
C3	Surface measuring device	P&S ₁₅ ($p=0.415$)	device measure wave aberration;
		P&S ₇₉ ($p=0.354$)	shear interferometer for flatness;
		P&S ₁₀₂ ($p=0.203$)	device measure lens deflection;
	
C4	Online monitoring	P&S ₂₇ ($p=0.813$)	monitor surface quality;
		P&S ₄₂ ($p=0.102$)	control optical surface quality;
		P&S ₇₈ ($p=0.005$)	inspect surface shape;
	

然后，基于该语义索引，采用 k -Means 聚类算法，进行了聚类研究。发现 LAOE 专利技术难题集中在“凹面检测 (concave test)、分析元件表面轮廓(analyze object surface profile)、光学元件表面质量控制(optical surface quality control)”等

领域；解决方案集中在“衍射法 (diffraction method)、全息图片法(artificial hologram)、条纹分析法(fringe analysis)”等领域。聚类的结果组成了 LAOE 语义 TRIZ 的应用空间，可被其他技术挖掘应用。

语义 TRIZ 其它应用场景包括：面向 TRIZ 的专利智能检索、专利自动分类、技术演化分析、核心专利自动筛选、热门专利预测等^[7-8]。现简述如下：

专利智能检索：传统专利检索主要是基于关键词与专利分类号的布尔匹配检索。而面向 TRIZ 的专利智能检索则可基于专利的具体技术主题，如：采用的技术方案、解决的技术难题进行语义检索；帮助用户快速发现同类专利。

专利自动分类：传统专利分类主要依据专利分析号实现，而分类号系统由于其表示过于宽泛，没有反应专利特有的 Problems & Solutions 信息，无法满足深度技术挖掘需求。基于语义 TRIZ 构建面向 P&S 的个性化分类体系，可满足面向具体技术领域 Problems & Solutions 进行专利自动分类的需求。

专利技术演化分析：专利技术演化过程本质上是一个不断解决旧问题，产生新问题的过程。Problems 与 Solutions 的交替推进的过程，客观反应了专利技术演化的过程。以前，技术演化分析大多基于单个表示维度来进行，如：关键词等。而基于面向P&S的专利语义知识表示模型，可分别单独对Problems与Solutions进行演化建模；然后利用Problems与Solutions之间语义关系，将Problems演化模型与Solutions演化模型融合，基于P&S交替演进来构建专利技术演化模型，以全面、深入揭示专利技术演化历程及趋势。

核心专利自动筛选：核心专利有多种选取指标。语义 TRIZ 则从具体技术的角度提供了判断标准，如：某专利解决的技术难题(Problems)或提供的解决方案(Solutions)是很多其他专利的基础，那么该专利就有可能该技术角度较为核心的专利。

热门专利预测：从技术角度看，专利热门与否的评价指标是：该专利解决的技术难题(Problems)或提供的解决方案(Solutions)在未来能否大量被其他专利所借鉴使用。将预测模型与语义 TRIZ 结合起来，可以实现从技术角度对未来专利的热门度进行预测与分析。

5 结果与讨论

本文设计一个开放的语义 TRIZ 框架，从概念空间、索引空间、应用空间三个维度构建语义 TRIZ，采用微观层 SAO 基础语义单元、中观层 P&S 本体、宏观层技术范畴三个层面描述语义 TRIZ 索引结构。基于语义 TRIZ 的专利技术挖掘能深化专利技术挖掘程度，拓展专利技术挖掘范围，更新专利技术挖掘视角，是未来的发展趋势与研究热点。目前，我们已经在 大口径光学元件(1364 条)、精准农业(3012 条)两个中等规模专利数据集上进行了实践与应用。基于该框架构建的大口径光学元件专利辅助创新知识库系统已成功在中科院上海光机所部署应用。

该框架在实际应用中发现一些问题。首先，LDA 生成的原始 P&S 向核心 P&S 本体自动语义映射，准确率较低，需要大量人工干预。其原因可能是：直接基于

LDA 模型生成的 P&S 主题聚焦度不够, 有些主题包含的 SAO 语义单元相互之间关系还较分散, 难以集中凝练成有意义的技术主题。其次, P&S 之间的语义关系需要专家人工标注, 如何自动或半自动挖掘 Problems、Solutions 之间的语义关系, 是未来研究的难点与重点。再次, 该方法在数据量较小(如几百条)时, 效果较差, 与 LDA 对数据量有一定要求有关; 在较大数据量级(上万条)场景下, 该方法有效性尚未得到验证。最后, 虽然领域专家参与, 是专利技术挖掘必不可少的部分。但如何规范领域专家参与模式、减少其主观影响, 增强该方法的稳定性与通用性, 也是未来研究重点。

未来, 我们将进一步规范与清洗该框架的流程; 制定一套可量化的标准, 帮助专家更准确、客观的参与构建语义 TRIZ。此外, 将本方法更进一步标准化、自动化, 并应用到其他技术领域, 辅助专利分析也是今后研究应用重点。

参考文献

- [1] Alan L. Porter, Cunningham S. Tech mining: exploiting new technologies for competitive advantage [M]. John Wiley & Sons, 2005:17-23.
- [2] 王朝晖. 专利文献的特点及其利用[J]. 现代情报, 2008, (9):151-152, 156.
- [3] 吕详惠, 仇宝艳, 乔鸿. 基于本体的专利知识发现体系研究[J]. 计算机与信息技术, 2008, (7):43-46.
- [4] Katie B., Ellen D. & Michael S. TRIZ-What is TRIZ. Resource Document. *Triz-journal*. (2006) [2012-12-10]. http://www.triz-journal.com/archives/what_is_triz/. Accessed 10 Dec 2012.
- [5] Terninko, J., Zusman, A., Zlotin, B. Systematic Innovation: An Introduction to TRIZ. *St. Lucie Press, Boca Raton, FL*. 1998.
- [6] He Cong, Loh Han Tong. Grouping of TRIZ Inventive Principles to facilitate automatic patent classification. *Expert Systems with Applications*, 2008, vol. 34, p.788-795.
- [7] Mikhail V. Semantic TRIZ [EB/OL]. (2004) [2012-09-10]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.1907&rep=rep1&type=pdf>.
- [8] IHS Inc. Optimize Decision-Making Across the Product Lifecycle. *White Paper*. [EB/OL]. (2012) [2012-12-10]. http://inventionmachine.com/Portals/56687/docs/OptimizingDecisionMakingAcrossTheProductLifecycle_WhitePaper_InventionMachine.pdf.
- [9] CREAX Inc. (2012). Accelerate your R&D-process with CREAX CreationSuite. *White Paper*. [EB/OL]. (2012) [2012-12-10]. <http://www.creationsuite.com/Content/CREAX%20CreationSuite%20folder.pdf>.
- [10] Mukherjea S. & Bamba B., Kankar P. Information Retrieval and Knowledge Discovery Utilizing a BioMedical Patent Semantic Web. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 2005, vol. 17, Issue 8, p. 1099-1110.
- [11] Youngho K. & Tian, Y, Jeong, Y, etc. Automatic Discovery of Technology Trends from Patent Text [C]. SAC '09 Proceedings of the 2009 ACM symposium on Applied Computing, New York, NY, USA, 2009.
- [12] Kim, H.B., Hyeok, Y.J & Kim, K.S. Semantic SAO network of patents for reusability of

inventive knowledge. *Management of Innovation and Technology (ICMIT) 2012 IEEE International Conference*, Singapore, 2012.

- [13]胡正银, 方曙. 专利文本技术挖掘研究进展综述[J]. 现代图书情报技术, 2014,30(6):62-70.
- [14]Yuen-Hsien T., Chi-Jen L., Yu-I L. Text mining techniques for patent analysis[J]. *Information Processing and Management*, 2007, 43, 1216-1247.
- [15] Young K. & Jong S., Park, S. Visualization of patent analysis for emerging technology[J]. *Expert Systems with Applications*,2008,34 :1804–1812.
- [16]Loh Han Tong , He Cong & Shen Lixiang. Automatic classification of patent documents for TRIZ users[J]. *World Patent Information*, 2006, 28,6-13.
- [17]He Cong & Loh Han Tong. Grouping of TRIZ Inventive Principles to facilitate automatic patent classification[J]. *Expert Systems with Applications*, 2008, 34, 788-795.
- [18] He Cong & Loh Han Tong. Pattern-oriented associative rule-based patent classification[J]. *Expert Systems with Applications*, 2010,37, 2395-2404.
- [19] Zhen Li, Derrick Tate, Christopher Lane & Christopher Adams. A framework for automatic TRIZ level of invention estimation of patents using natural language processing, knowledge-transfer and patent citation metrics. *Computer-Aided Design*, 2012,vol. 44, p.987-1010.
- [20] Janghyeok Yoon & Kwangsoo Kim. An automated method for identifying TRIZ evolution trends from patents[J]. *Expert Systems with Applications*, 2010,37,2882-2890.
- [21] Janghyeok Yoon & Kwangsoo Kim. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks[J]. *Scientometrics*, 2011,88,213-228.
- [22] Hyunseok Park, Jason Jihoon Ree & Kwangsoo Kim. An SAO-based approach to patent evaluation using TRIZ evolution trends[C]. *Management of Innovation and Technology (ICMIT) 2012 IEEE International Conference*, Singapore,2012.
- [23] Hyunseok Park , Jason Jihoon Ree , Kwangsoo Kim. Identification of promising patents for technology transfers using TRIZ evolution trends[J]. *Expert Systems with Applications*, 2013, 40,736-743.
- [24] Yi Zhang, Alan L. Porter & Zhengyin Hu. An Inductive Method for “Term Clumping”: A Case Study on Dye-Sensitized Solar Cells[C]. *the International Conference on Innovative Methods for Innovation Management and Policy*, Beijing, China, 2012,May 21-25.
- [25] Yi Zhang, Alan L. Porter, Ying Guo & Xiao Zhou. Discovering emerging technology trends: with TRIZ and technology road mapping[C]. *the 2nd Global TechMining Conference*, Montreal, Canada,2012, Sept. 5.
- [26] Xian Zhang, Shu Fang, Chuan Tang, Guohua Xiao, Zhengyin Hu, Lidan Gao. (2009).Study on Indicator System for Core Patent Documents Evaluation. *Proceedings of ISSI 2009 - 12th International Conference of the International Society for Scientometrics & Informetrics*, Vol. 1: 154-164. July ,14-17,, 2009, Rio De Janeiro, Brasil.
- [27] Anthony F., Stephen S.& Oren E.. Identifying Relations for Open Information Extraction [EB/OL].(2011) [2013-03-02]. <http://ai.cs.washington.edu/www/media/papers/reverb.pdf>.
- [28] David M. Machine Learning with MALLETT. [EB/OL]. [2013-03-03]. <http://mallet.cs.umass.edu/mallet-tutorial.pdf>.
- [29] Zhang Y., Alan P., Hu ZY., etc. “Term clumping” for technical intelligence: A case study on dye-sensitized solar cells [J]. *Technological Forecasting and Social Change*, 2014, 85, 26–39.