

## 社交网络智慧搜索研究进展与发展趋势

贾焰, 甘亮, 李爱平, 徐菁

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

**摘要:** 在以社交网络为代表的 Web 2.0 时代, 搜索需求已经远超出 Web 1.0 搜索引擎的能力范畴, 数据具有模式多态、快速产生、动态交互、碎片化、变化演绎等特点, 这些新特点给搜索引擎技术带来了革命性和颠覆性的挑战。归纳了在线社交网络智慧搜索研究进展和技术要点, 分析了研究发展过程将面临 3 个主要问题: 在线社交网络中智慧与知识的发掘与推演、用户真实搜索意图的理解与表示以及满足用户真实意图的智慧解答在线响应, 并以解决此 3 个问题为目标讨论了相关技术发展。

**关键词:** 社交网络; 社交网络大搜索; 意图理解; 匹配求解; 智慧解答

**中图分类号:** TP393

**文献标识码:** A

## Research progress and development trend of online social network smart search

JIA Yan, GAN Liang, LI Ai-ping, XU Jing

(School of Computer Science, National University of Defense Technology, Changsha 410073, China)

**Abstract:** In the era of Web 2.0 as the representative of online social network, the requirement of Web search has been far beyond the ability of Web 1.0 search engines, for that data has a pattern of polymorphism, rapid generation, dynamic interaction, fragmentation, change and other characteristics, these new features to search engine technology has brought revolutionary and subversive challenges. The research progressed technical essentials of online social network search are induced. Three main contents of OSN smart search were studied, including the understanding and reasoning of the wisdom and knowledge, the understanding and the expression of user's real intention, and online response of user's real intention. And then, the key technologies and the development trend of online social network search are discussed.

**Key words:** social network; big search of social network; comprehension of search intension; solution matching; intelligent solution

### 1 引言

在线社交网络是一种在信息网络上由社会个体集合及个体之间的连接关系构成的社会性结构。在线社交网络可分为 4 类: 1) 即时消息类应用, 是一种提供在线实时通信的平台, 如 QQ、微信等; 2) 在线社交类应用, 是一种提供在线社交关系的平台, 如 Facebook、人人网等; 3) 微博类应用, 是一种提供双向发布短信息的平台, 如 Twitter、新

浪微博等; 4) 共享空间等其他类应用, 是其他可以相互沟通但结合不紧密的 Web2.0 应用, 如论坛、博客等。当前, 在线社交网络应用正处在蓬勃发展期, Facebook 已拥有超过 14 亿的用户, 成为第一大“人口国”, 新浪微博用户数已达到 5.36 亿, 腾讯微博用户数已达到 5.7 亿。在线社交网络应用正深刻地影响着人们生活的各个方面。

在线社交网络数据具有丰富价值, 并蕴含着大量智慧。主要体现在: 1) 蕴含了大量用户情感、

收稿日期: 2015-10-12; 修回日期: 2015-12-17

基金项目: 国家重点基础研究发展计划(“973”计划)基金资助项目(2013CB329601, 2013CB329604, 2013CB329606); 国家自然科学基金资助项目(61472433, 61502517)

**Foundation Items:** The National Basic Research and Development Program of China (973 Program)(2013CB329601, 2013CB329604, 2013CB329606); The National Natural Science Foundation of China (61472433, 61502517)

立场和观点,进而可发掘人类的思想和行为;2)包含了各类具有时空特性的话题、事件信息,进而可对它们的起源、传播和发展规律进行揭示和挖掘;3)记录了用户和话题间丰富的关系数据,进而可发现朋友关系、社交圈子、用户与话题、话题与话题等之间关系;4)充满了针对专业问题的丰富讨论,进而可汇聚群体智慧,服务于人们的工作和生活。

传统的搜索引擎技术,主要是面向 Web1.0 静态网页,是基于关键词的“存在性搜索”,不能支持面向 Web2.0/3.0 应用,具有 5V 特性的大数据,及其满足用户需求智慧解答的搜索。因此不能发掘丰富的在线社交网络智慧,且服务于用户。本文研究在线社交网络智慧搜索技术,定义如下。

在线社交网络智慧搜索是在正确理解用户意图的基础上,基于社交网络数据进行加工、推演处理发掘知识,进而给出智慧解答。在线社交网络大搜索具有“4S”特点:1)意图感知(sensing the context),结合用户请求的上下文、时空特性、场景感知等方式,支持在语义级别上对用户搜索意图进行准确理解;2)多源综合(synthesis from multiple channels),综合、关联多通道、多来源(不同社交网络)的社交网络数据和信息,进行统一的知识发掘和推演;3)安全可信(security privacy and trust),在线社交网络搜索结果的安全可信的,并且支持隐私保护;4)智慧解答(intelligent solution),搜索的结果是基于在线社交网络数据和信息,经过发掘、推理和计算而得到的一组有序智慧解答。

## 2 相关工作

在线社交网络智慧搜索涉及的相关理论和技术包括:搜索引擎技术、在线社交网络分析、复杂对象关系建模、意图理解与匹配及知识构建与推演等。

当前的搜索引擎技术主要包括互联网搜索引擎和在线社交网络搜索。主要的互联网搜索引擎包括全文搜索、元搜索引擎和垂直搜索引擎等。在互联网搜索引擎中,为优化搜索结果,通常采用倒排索引技术对网页信息进行索引,并采用排序算法对搜索结果进行等级排名,典型的算法包括 PageRank<sup>[1]</sup>和 HITS<sup>[2]</sup>等;为提高搜索结果的关联性,Google、百度、搜狗等引入知识图谱技术;为实现搜索信息的高效存储管理,各互联网厂商纷纷

提出了解决方案,如 Google 的 Bigtable<sup>[3]</sup>, Amazon 的 Dynamo<sup>[4]</sup>、Yahoo 的 PNUTS<sup>[5]</sup>等。在在线社交网络搜索方面,360 推出的“我的搜索”,引入微博、微信等社交因素的影响,并在搜索结果中进行展示。Facebook 推出的社会搜索引擎 Graph Search,用户可在社交网络中对好友、照片、地点等进行搜索。微软推出的人立方关系搜索,自动地计算每一个人名与关键词的距离,并可展示人的社会化关系。

在线社交网络分析为社交网络中知识的获取和推演提供了相应的方法。主要的社交网络分析包括话题发现与演化、虚拟社区发现与演化、信息传播以及影响力分析等。话题发现与演化能有效支撑网络时代的信息决策。代表性工作包括: Blei<sup>[6]</sup>提出的隐含狄利克雷分布的 LDA 模型、Lin 等<sup>[7]</sup>提出的潜在扩散路径方法、美国马里兰大学研究的词项间共现频率反映语义关联原理的方法<sup>[8]</sup>。虚拟社区发现与演化有助于发现社交网络中的拓扑结构信息。代表性工作包括: Newman 等<sup>[9]</sup>提出的模块化方法、Chakrabarti 等<sup>[10]</sup>提出的社区演化模型、Mucha 等<sup>[11]</sup>提出的多层网络社区发现、Tang 等<sup>[12]</sup>研究的多模态网络社区发现等。社交信息传播机制有助于对社会网络的认识。代表性工作包括: Gruhl 等<sup>[13]</sup>基于 SIRS 传染病模型、Han 等提出的高斯条件随机场模型、Antulov-Fantulin 等<sup>[14]</sup>提出的统计推理框架溯源方法。影响力分析能发现社交网络中高影响力用户和影响强度。代表性工作包括: Ellison 等<sup>[15]</sup>研究了在线社交关系对现实社交关系的群体互动影响; Woolley 等<sup>[16]</sup>分析了心理因素、认知空间对群体聚集的影响; Wen 等<sup>[17]</sup>根据关注网络 and 用户兴趣相似性计算个体在每个话题上的影响力; Romero 等<sup>[18]</sup>综合考虑了影响力与冷漠性,提出了类 HITS 的算法。

对象关系模型是构建在线社交网络搜索知识仓库的基础。当前复杂对象关系的建模通常用图结构来表示,常用图模型包括 Property Graph<sup>[19]</sup>、RDF<sup>[20]</sup>、MultiGraph 模型<sup>[21]</sup>等。Property Graph 在节点和边上可以存在任意数量的键值对表示属性或标签,因而其表达能力很强。RDF 用三元组 SPO (subject, property, object) 来描述实体之间的关系,是当前表示实体及其关系的一种常见模式。MultiGraph 模型可在 2 个实体之间保留多条边以表示多种关系。近年来,许多研究将时空信息融入

到复杂对象关系的建模中。微软亚洲研究院分别从用户、地理位置和事件 3 个层面对基于位置的地理社交网络进行了研究,发现单纯社会网络中个体之间无法表现的关系<sup>[22]</sup>。Shekhar<sup>[23]</sup>将时空因素考虑到在线社交网络数据分析中,提出一种时间聚集的图模型。

用户意图理解与匹配是搜索中的关键技术。在用户意图理解方面, Wolframalpha 通过从公众的和获得授权的资源中发掘、构建的数据库,能够理解用户问题并直接给出答案。搜狗的“知立方”通过引入“语义理解”技术,试图理解用户的搜索意图,对搜索结果进行重新优化计算。Etzioni 等提出了基于规则模板抽取实体/概念之间的关系来描述和理解搜索意图。Madhu 等<sup>[24]</sup>利用语义网工具和技术提供分层模块的方法解决搜索引擎对语义内容的理解。在意图匹配方面,主要包括文本模型和图模型。基于文本模型的意图匹配通过将以关键词查询检索的方式来把用户的意图进行语义转换和目标文档的匹配,并获取相关度排序。基于图模型的意图匹配通过图搜索来实现搜索意图与搜索空间中目标项的查找和匹配,主要包括<sup>[25]</sup>: 关键词图搜索技术、子图匹配技术和近似图匹配技术等。

知识是实现智慧搜索的关键。当前,知识构建较多地从知识图谱构建角度加以展开,以互联网网页为来源的典型知识图谱包括 KnowItAll<sup>[26]</sup>、TextRunner<sup>[27]</sup>和 Probase<sup>[28]</sup>,以在线百科为数据来源的知识图谱包括 YAGO<sup>[29]</sup>和 DBpedia<sup>[30]</sup>等。知识推演是在给定目标的情况下,在知识库或网络空间中进行推演求解,以获得答案并产生新的知识。当前

知识推演的操作过程包括利用统计、知识推理和众包等方法。其中,主要的推理方法包括:正向推理、逆向推理、双向推理、非精确推理、基于语义的推理和基于案例的推理等。

上述技术的发展为在线社交网络智慧搜索的研究奠定了研究基础,在理论、方法和技术方面存在诸多挑战,主要包括:在线社交网络中智慧与知识的发掘与推演、用户真实搜索意图的理解与表示、满足用户真实意图的智慧解答在线响应。

### 3 研究进展及技术要点

目前,社交网络智慧搜索与当前的社交网络搜索的区别主要体现在智慧的能力,而智慧处理过程是以知识图谱为基础,主要研究内容可划分为在线社交网络知识发掘与推演、知识聚合与组织管理、用户搜索意图理解、用户意图的搜索与匹配等部分,各研究点间交互形成总体框架如图 1 所示。

社交网络知识发掘与推演。可支持对在线社交网络空间中的数据获取和推理,包括微博、博客、论坛、维基、共享网站等空间中采集文本、图片、语音、视频等各种类型的多模态数据,以及各类已存在的对象知识和关系知识。数据获取与采集过程不间断进行,采集后的数据和知识是后续推理和搜索的基础。

知识聚合与组织管理。面向在线社交网络空间的海量对象知识及关系知识进行建模;在此模型实例化的基础上通过知识聚合,构建知识仓库空间,并通过索引、关联和演算等聚合操作预先形成知识聚合体。知识仓库中的知识是不断经过二次加工的,经过用户的查询、修改、反馈和自演化的过程,

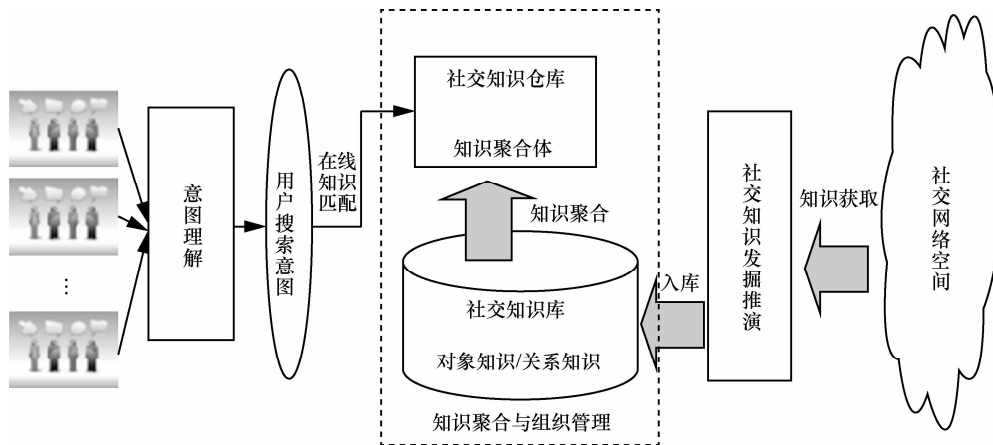


图 1 总体框架

逐步完善, 根据应用建立各类索引, 同时满足用户搜索时的准确性需求和实时性需求。

用户搜索意图理解。面向意图理解的准确性和歧义消除的基本需求。结合用户的上下文和语义知识等方法, 迅速、准确地理解用户的真实意图, 并转变成与知识仓库可匹配推演的表示方式。

用户意图的搜索与匹配。基于意图理解表示和知识仓库, 经过匹配、推理、计算、乃至众包等技术和方法的处理, 形成若干个满足用户真正意图的智慧综合的解决方案, 并通过结果评价排序方式给出其优先级, 为用户提供智慧的解答方案。

### 3.1 在线社交网络知识表示模型

在线社交网络中的对象知识具有多样化特性, 可通过文本解析、实体抽取、关系抽取、元数据分析、指代消解等技术来获得在线社交网络中的不同侧面对象知识, 并以特征关联的形式对其进行描述, 建立针对个体对象的内容语义描述模型。

在线社交网络中的内在多层次、演化的关系型知识需要进一步提取和挖掘, 一般可采用支持语义关系的语义图模型表达; 综合对象知识和关系型知识, 可借鉴目前时态地理信息系统以及数据分析领域中的资源描述框架 (RDF)、属性图 (property graph)、多图 (multi-graph) 等模型方法, 通过模型的组合以及扩展等方法, 并通过整合现有语义库 (包括 Freebase 和 Probase 等) 来统一表示语义信息。

### 3.2 在线社交网络知识的发掘与推演

在线社交网络知识获取与推演具有多样化、关系复杂与演化等需求, 可从个体行为及立场分析、群体社区发现及极化规律、话题的缘起与发展和信息传播规律等在线社交网络的角度出发, 进行发掘和推演。研究主要针对社交实体的对象交互特性、时空特性、规模特性、多源特性等方面。

在线社交网络中的对象具有丰富的交互关系进行推理挖掘, 可采用基于时序语义图的关联算法。支持时空特性是社交网络知识的主要特性, 可基于相似性计算与多尺度空间匹配等方法, 以及面向在线社交网络的时态逻辑推理算法, 利用关系传递和协同过滤等技术, 对在线社交网络知识推理。在线社交网络中的对象属性具有个数规模大的特点, 可通过目标驱动的基于属性依赖关系的可伸缩的模态推理技术, 实现基于刻面的社交网络大规模属性推理。社交网络中的知识含有

大量多源异构交互信息, 可通过离线众包推理与反馈相结合的多源知识融合方法, 实现社交网络交互信息的众包推理与多专家信息的智慧解答的有机融合。

### 3.3 面向在线社交网络知识聚合与组织管理

发掘和推演生成的知识是粗糙、低层次的, 可通过知识聚合来生成精炼、物化和泛化的知识来满足用户搜索的需求, 并形成知识仓库。

知识预先聚合、组织并生成知识聚合体过程应具有有效性、准确性和顺序性, 在社交网络知识表示模型基础上, 可参考 Wikipedia/DBpedia/ Freebase 等多种语义概念层次, 参考联机分析处理的聚合计算机理, 建立在线社交网络中面向领域的对象知识、关系知识间的聚类方法, 以及各概念层次间的聚合函数。在此基础上, 基于特征空间的降维分解方法研究高维空间中各维度的可聚合性及相应的聚合函数。其中, 对于时空属性的聚合计算, 将采用多时间粒度聚合、基于地理位置的空间聚合等方法, 研究时空聚合计算函数和有效计算方法; 在聚合计算的基础上, 采用基于时空相似度散列的知识聚合体模型表示和存储方法, 将时、空上相似或相近的对象和关系就近存储并建立高效索引; 在概率 Skyline 和概率 Top- $k$  算法框架下, 可研究时变、不确定环境下的知识聚合体的动态排序与更新演化算法。

### 3.4 用户搜索意图理解

用户搜索意图主要体现在用户的历史行为、场景环境、语言表达等方面, 其研究也基于各个方面的综合感知和理解。

用户搜索历史行为, 可按照由个体到群体, 从点到轨迹的思路, 采用频繁模式挖掘相关技术, 挖掘用户的行为模式和搜索模式, 建立用户搜索时空场景知识库, 用以识别用户的行为、情感、意图、经验和生活模式。

用户场景环境, 是用户所处的时间上下文、空间上下文、历史行为上下文、社交关系上下文等环境, 一般采用基于内容以及协同过滤等推荐方法和机器学习相结合的方法, 增强用户意图理解的准确性, 并结合用户偏好和当前位置, 按照用户的满意度、兴趣度选择与用户当前需求相关的信息, 进行空间信息的语义搜索, 并建立基于语义的信息聚合模型, 将个性化需求的信息进行整合。

用户语言表达, 是用户的自然语言文字或语音

表述，在此方面有大量研究成果。考虑搜索过程的特殊性，应针对用户的搜索意图的一些不定和模糊表达等特征，在稀疏的搜索空间中，通过一些数据降维嵌入和相近分析等方法来进行有效的推理演算，更好地支持用户意图的理解和匹配，并通过交互、反馈等方式对理解有偏差的意图进行纠正。

### 3.5 在线知识匹配求解

在线知识匹配求解是知识仓库中知识的查询匹配、推理求解以及搜索答案生成过程。

知识聚合体中的文本类知识的快速匹配算法，可基于深度学习思想，研究不同关键词间的深度语义关联，并在语义空间中研究基于时空相似的快速匹配算法，实现知识聚合体能快速准确满足搜索用户需求。针对知识聚合体中的关系类知识的快速匹配算法，可基于图的分布式处理方法，研究大图和巨图并行匹配的分解算法及优化方法，支持大图和巨图的高效查询。对于混合属性查询请求，可采用地理信息等时空特性的知识快速匹配为重点，处理地理信息、时空特性的快速知识匹配算法。针对用户意图的解答排序与评估反馈，可采用半监督增强学习方法和自反馈理论，突破在学习因素和反馈特征因子数量大、维度高的情况下高效反馈学习算法，实现搜索过程的自我演化与更新需求。

## 4 研究发展趋势

社交网络智慧搜索发展研究，将聚焦于 3 个主要问题，包括社交网络中智慧与知识的挖掘与发现、用户真实搜索意图的理解与表示、快速给出满足用户需求的智慧解答。在应用方面，重点考虑满足政府决策的民意调查、舆情分析，以及企业市场

的社会化营销等各行业领域现实需求。为满足以上 3 个问题，主要关键技术发展将包括以下 5 个方面内容，如图 2 所示。

### 4.1 支持时空特性的在线社交网络知识表示模型

针对在线社交网络中的人物情感立场、事件缘起发展、群体互动与聚集等巨规模、复杂、演化的对象和关系，需要研究支持时空特性的社交网络知识表示模型，实现对社交网络知识的建模。关键技术发展将包括以下几方面。

1) 在线社交网络中的对象知识表示方法。针对在线社交网络中的对象知识多样化特性，包括人物、话题、信息等各种各样的社交网络对象，且每个对象属性多样，深层挖掘对象的各种属性的特点及其随时间演化的规律，研究统一的对象建模与知识表示模型。

2) 在线社交网络巨复杂关系型知识表示方法。针对在线社交网络中各种关系规模巨大、种类繁多，粒度不同、时间演化的特点，包括用户、社区、话题之间的各种关系，需要分析关系的不同特点及演化规律，建立适合社交网络对象关系的统一语义表示模型。

3) 在线社交网络对象与关系统一融合的表示计算模型。针对不同来源、跨通道的在线社交网络中的复杂对象与关系，需要研究能够统一融合表示的在线社交网络知识归一化表示模型，实现对巨规模、复杂、演化的在线社交网络知识建模。

### 4.2 在线社交网络知识的发掘与推演

针对在线社交网络中知识多样化，关系复杂及时空演化等特点，包括个体行为及立场分析、群体社区发现及极化规律、话题的缘起与发展和

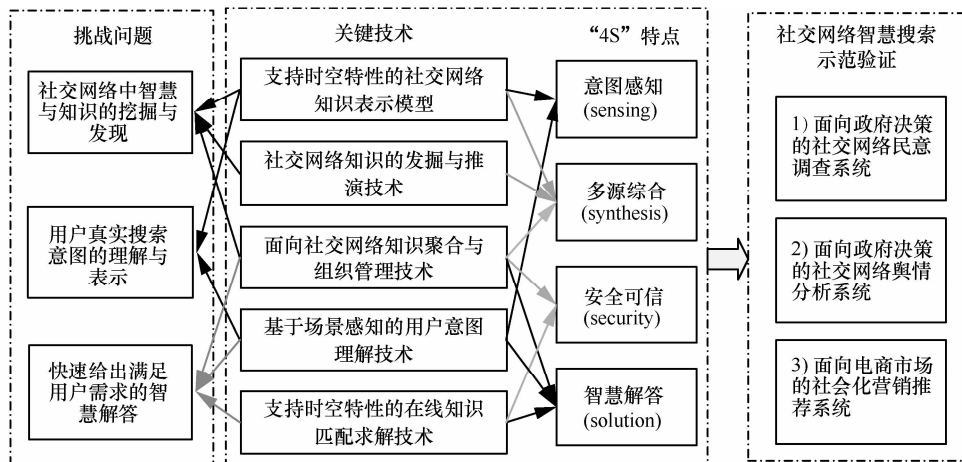


图 2 技术发展及相互关系

信息传播规律等, 研究以复杂社会计算为基础的在线社交网络知识的发掘和推演。关键技术发展如下所示。

1) 面向在线社交网络关系的新型关联推理机制。针对在线社交网络中的对象具有丰富的交互关系, 以及知识之间具有关联关系的特点, 需要研究适合于在线社交网络知识的表性、隐性、虚实结合的关联关系挖掘与发现推理方法。

2) 面向在线社交网络知识时空属性的推演机制。针对在线社交网络知识具有丰富的时空属性的特点, 需要研究时空相似性计算和时空推理技术, 及面向社交网络的时态逻辑推理技术, 实现多尺度的时空数据融合推理。

3) 基于刻面的在线社交网络大规模属性推理机制。针对在线社交网络中的对象属性个数规模众多的特点, 需要研究目标驱动的基于属性依赖关系的可伸缩的模式推理模型, 以及面向多属性刻面的冲突消解方法。

4) 面向在线社交网络交互信息的众包推理与融合。针对在线社交网络中的知识含有大量交互信息的特点, 需要研究基于离线众包推理的智能知识的挖掘框架, 以及基于标注与反馈相结合的多源知识融合机理。

#### 4.3 面向在线社交网络知识聚合与组织管理

针对在线社交网络知识具有层次性和多粒度特点, 以及用户不可预测的在线知识查询需求, 需要对知识进行预先聚合与组织, 生成在线社交网络知识聚合体, 并进行有效排序, 从而支持实时多维度的搜索请求。关键技术发展将包括以下几方面。

1) 高维特征空间的知识聚合计算。针对在线社交网络知识跨时空、多层次、多维度的特点, 基于发掘与推演技术, 需要研究在线社交网络知识聚合计算方法, 及各知识聚合计算的相关性与计算策略技术。

2) 社交知识聚合体的表示与存储模型。针对在线社交网络聚合体归一化表示问题和存储空间随维度增加呈指数增长的维灾难特点, 需要研究面向在线社交网络搜索的社交知识聚合体的模型表示和存储管理方法。

3) 社交知识聚合体的排序与更新演化模式。针对用户搜索需求不断变化和无法预测的特点, 需要研究在线社交网络知识聚合体的索引排序, 以及获取的数据不断更新过程中, 知识聚合体的自我演化

与更新模型。

#### 4.4 基于场景感知的用户意图理解

面向用户查询输入的关键词、语音、手势等内容, 结合用户手机终端、所处运动轨迹的时空场景以及历史记录和个人偏好等信息, 准确理解用户的意图, 并采用支持高效查询推演的统一模型进行表示。关键技术发展将包括以下几方面。

1) 支持时空属性的用户搜索意图建模。针对在线社交网络的各种时空轨迹数据, 体现了用户当前在真实世界中的场景的特点, 需要研究基于时空信息的用户行为模式分析技术, 实现对用户真实搜索意图的建模。

2) 支持上下文的语义级用户意图理解方法。针对用户搜索意图与上下文紧密相关的特点, 需要研究基于上下文感知的用户意图理解方法, 及基于情感分析的用户意图理解方法, 实现语义级用户意图的理解。

3) 基于交互式的用户搜索意图理解方法。针对用户搜索意图单次表述具有二义性等特点, 需要研究基于交互式的用户意图理解方法, 实现交互步骤最小优化模型, 以最少交互开销了解用户潜在搜索意图。

#### 4.5 支持时空特性的在线知识匹配求解

知识匹配求解是获取匹配用户需求搜索结果的关键技术。针对用户的在线社交网络搜索请求复杂多样和要求在线响应的特点, 需研究在社交网络知识聚合体中的快速匹配求解推演算法, 以及相应的实验验证系统。关键技术发展将包括以下几方面。

1) 基于文本的快速在线匹配与解答排序。针对用户搜索需求多以关键词等文本进行表示的特点, 需研究基于文本进行的在线社交网络搜索快速在线匹配与排序方法, 实现社交网络知识聚合体中快速准确满足搜索用户需求的方法。

2) 基于子图的面向用户意图的快速匹配求解。针对在线社交网络知识多以图结构表示的特点, 需要研究面向大图结构的特性分析技术, 基于大图的高效查询及其优化技术, 以及基于大图的用户意图高效推演等技术。

3) 支持时空特性的社交网络知识快速匹配。针对在线社交网络知识的时空特性, 研究基于语义图模型和协同过滤等方法, 以及支持时空特性的社交网络知识匹配算法, 高效完成对用户搜索意图的快

速准确匹配。

4) 面向用户意图的解答排序与评估反馈。针对在线社交网络搜索自我演化与更新的需求, 研究异构信息聚合搜索评价技术, 评价返回的各种类型的信息之间的相互作用与评估体系, 实现不同搜索解答的评估与反馈优化。

## 5 结束语

在 Web2.0 蓬勃发展的环境, 在线社交网络智慧搜索具有迫切的需求, 本文采用理论研究和实证研究相结合的方法对当前研究进展及未来发展趋势进行了归纳和预测。主要探索用户搜索真实意图的理解与表示机理; 解决社交网络中复杂海量的知识以及知识之间关系的模型构建与知识发现以及推演机制描述方法; 研究用户意图在搜索空间的快速匹配模型和方法, 以构建面向社交网络大搜索的运行支撑平台及环境。

长期发展将以面向政府决策的民意调查、舆情分析和社会化营销等重大需求为目标, 构建在线社交网络大搜索的示范验证系统, 从而提升我国在社交网络大搜索的原始创新能力和国际影响力, 推动信息技术发展, 抢占 IT 技术的战略高地。

## 参考文献:

- [1] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: bringing order to the web[A]. Stanford Info Lab[C]. 1999.1-14.
- [2] KLEINBERG J. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM, 1999, 46(5): 604-632.
- [3] CHANG F, DEAN J, GHEMAWAT S. Bigtable: a distributed storage system for structured data[J]. ACM Transactions on Computer Systems. 2008, 26(2):205-218.
- [4] DECANDIA G, HASTORUN D, JAMPANI M. Dynamo: amazon's highly available key-value store[A]. SOSP'07[C]. 2007.205-220.
- [5] COOPER B F, RAMAKRISHNAN R, SRIVASTAVA U, et al. PNUTS: Yahoo!'s hosted data serving platform[J]. Proceedings of the VLDB Endowment, 2008, 1(2): 1277-1288.
- [6] BLEI D, NG A, JORDAN M. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003(3): 993-1022.
- [7] LIN C X, MEI Q, HAN J, et al. The joint inference of topic diffusion and evolution in social communities[A]. IEEE 11th International Conference on Data Mining (ICDM)[C]. IEEE, 2011.378-387.
- [8] SAYYADI H, RASCHID L. A graph analytical approach for topic detection[J]. ACM Transactions on Internet Technology (TOIT), 2013, 13(2):992-999.
- [9] MARK N, ELIZABETH L. Mixture models and exploratory analysis in networks[J]. Proc Natl Acad Sci 2007, 104(23): 9564-9569.
- [10] CHAKRABARTI D, KUMAR R, TOMKINS A. Evolutionary clustering[A]. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining[C]. 2006.554-560.
- [11] MUCHA P, RICHARDSON T, MACON K. Community structure in time-dependent, multiscale, and multiplex networks[J]. Science, 2010, 328 (5980): 876-878.
- [12] TANG L, LIU H, ZHANG J. Identifying evolving groups in dynamic multimode networks[J]. IEEE Transation on Knowledge and Data Engineering, 2012, 24 (1) : 72-85.
- [13] GRUHL D, MEREDITH D N, PIEPER J H, et al. The Web beyond popularity: a really simple system for web scale rss[A]. Proceedings of the 15th International Conference on World Wide Web[C]. 2006. 183-192.
- [14] ANTULOV-FANTULIN N, LANCIC A, STEFANCIC H, et al. Statistical inference framework for source detection of contagion processes on arbitrary network structures[A]. 2014 IEEE Eighth International Conference on Self-Adaptive and Self-Organizing Systems Workshops (SASOW)[C]. 2014.78-83.
- [15] STEINFELD C, ELLISON N B, LAMPE C. Social capital, self-esteem, and use of online social network sites: a longitudinal Analysis[J]. Journal of Applied Developmental Psychology, 2008, 29(6): 434-445.
- [16] WOOLLEY A W, CHABRIS C F, PENTLAND A, et al. Evidence for a collective intelligence factor in the performance of human groups[J]. Science, 2010, 330(6004): 686-688.
- [17] WENG J, LIM E P, JIANG J, et al. Twitterrank: finding topic-sensitive influential twitterers[A]. The 3rd ACM International Conference on Web Search and Data Mining (WSDM'10)[C]. New York, USA, 2010. 261-270.
- [18] ROMERO D M, GALUBA W, ASUR S, et al. Influence and passivity in social media[A]. Proc European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)[C]. 2011. 18-33
- [19] ADAM N, ATLURI V, JANEJA V, et al. Semantic graph based knowledge discovery from heterogeneous information sources[A]. Conference on Public/Private R&D Partnerships in Homeland Security[C]. 2005.
- [20] ZOU L, MO J, CHEN L, et al. GStore: answering SPARQL queries via subgraph matching[J]. Proc VLDB Endow. 2011, 4(8): 482-493.
- [21] BERLINGERIO M, COSCIA M, GIANNOTTI F, et al. Multidimensional Networks: Foundations of Structural Analysis [J]. World Wide Web, 2013,16 (5-6): 567-593.
- [22] BAO J, ZHENG Y, MOKBEL M F. Location-based and preference-aware recommendation using sparse geo-social networking data[A]. ACM SIGSPATIAL GIS 2012[C]. 2012.199-208.
- [23] GEORGE B, SHEKHAR S. Time-aggregated graphs for modeling spatio-temporal networks[A]. Conceptual Modeling - Theory and Practice[C]. Springer Berlin Heidelberg, 2006. 85-99.
- [24] MADHU G, GOVARDHAN A, RAJINIKANTH T. Intelligent semantic web search engines: a brief survey[J]. International Journal of Web & Semantic Technology, 2011, 2(1): 34-42.
- [25] FAN W, LI J, MA S, et al. Graph homomorphism revisited for graph matching[J]. Proceedings of the VLDB Endowment, 2010, 3(1-2): 1161-1172.
- [26] ETZIONI O, CAFARELLA M, DOWNEY D, et al. Web-scale information extraction in knowitall:(preliminary results)[A]. Proceedings of the 13th International Conference on World Wide Web[C]. 2004. 100-110.
- [27] YATES A, CAFARELLA M, BANKO M, et al. Textrunner: open

information extraction on the Web[A]. Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, Association for Computational Linguistics[C]. 2007. 25-26.

[28] WU W, LI H, WANG H, et al. Probbase: a probabilistic taxonomy for text understanding[A]. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data[C]. 2012. 481-492.

[29] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[A]. Proceedings of the 16th International Conference on World Wide Web[C]. 2007. 697-706.

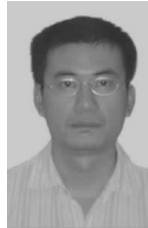
[30] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: a nucleus for a Web of Open Data[M]. Springer Berlin Heidelberg, 2007.

[31] 方滨兴, 刘克, 吴曼青, 等. 大搜索技术白皮书[M]. 北京: 电子工业出版社, 2015.

FANG B X, LIU K, WU M Q, et al. White Paper of Big Search in CyberSpace[M]. Beijing: Electronic Industry Press, 2015.



**甘亮** (1977-), 男, 江西樟树人, 国防科技大学博士后, 主要研究方向为知识图谱、网络信息安全和数据挖掘。



**李爱平** (1974-), 男, 山东诸城人, 博士, 国防科学技术大学研究员, 主要研究方向为语义网络、数据挖掘和网络信息安全。

**作者简介:**



**贾焰** (1960-), 女, 四川成都人, 博士, 国防科学技术大学教授, 主要研究方向为大数据、网络信息安全和社交网络。



**徐菁** (1989-), 女, 山东济南人, 国防科学技术大学博士生, 主要研究方向为大数据、文本信息抽取和文本挖掘。