情报工程 ISSN 2095-915X
第2卷 第1期 59-65 2016年2月
TECHNOLOGY INTELLIGENCE ENGINEERING
ISSN 2095-915X Vol.2 No.1 59-65 Feb 2016

doi:10.3772/j.issn.2095-915x.2016.01.008

基于概念树的论文评审专家推荐

李琳娜

（中国科学技术信息研究所 北京 100038）

摘要：本文基于概念树计算论文与专家之间的相似度，然后采用基于启发式的最大相似度匹配方法将论文分配给相应的评审专家。基于概念树的相似度计算，可以充分满足主题覆盖度约束；基于启发式的最大相似度匹配算法不仅可以满足利益冲突约束，又可以满足专家工作量约束。最后实验验证了所提算法的有效性。

关键词：概念树，评审专家，评审专家推荐，利益冲突，主题覆盖度

分类号：TP391

Paper-to-Reviewer Assignment Based on Concept Tree

LI Linna

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: This paper computes the similarity between papers and reviewers based on concept tree, which is used to assign and assigns papers to appropriate reviewers based on heuristic maximum matching degree. The similarity based on concept tree can meet the constraint of topic coverage. The matching algorithm can meet the constraint of interest conflict and reviewer load balance at the same time. Finally, the experiments show verify the validity of the proposed method.

Keywords: Concept tree, reviewer, paper-to-reviewer assignment, interest conflict, topic coverage

基金项目：中国科学技术信息研究所科研项目预研基金：“基于隐语义模型和文档结构的文档关键词自动抽取”（项目编号：YY2015-14）。

作者简介：李琳娜（1981-），女，助理研究员，研究方向：个性化推荐，知识组织。通讯地址：北京市海淀区复兴路15号情报理论与方法研究中心 tel：58882035-812, Email：liln@isitc.ac.cn

1 引言

评审专家推荐是指为投稿的期刊或者会议论文推荐合适的评审专家，并从论文中选出高水平的研究论文进行出版。能否将论文分配给比较合适的审稿人是一个会议能否成功、一个期刊是否优秀的重要因素。

在实践中为投稿论文推荐合适的评审专家要满足多种约束条件：（1）利益冲突，比如：论文作者A和推荐专家B虽然没有直接合作写过论文，但是他们有可能是同事或者相关学术领域的间接好友，若将A分配给B进行评审就会造成一定程度的不公平；（2）主题覆盖度约束，一篇论文可能会涉及相关学科的多个领域，需要相关领域的多个专家共同审阅；（3）专家工作量约束，每个评审专家能评审的论文数量有限制，待评审论文需要的评审专家数量也有限制。

自动评审专家推荐问题最早由Dumais和Nielsen提出[1]。他们采用潜在语义索引技术对评审专家发表的论文及待评审论文进行特征抽取，随后计算需要评审的论文与评审专家两两之间的相似度。将每篇论文推荐给与其相似度比较高的几位专家进行评审。随后一些研究者基于其他的方法计算专家和论文之间的匹配程度，但仍采用与Dumais和Nielsen一样的专家分配算法。显然，这样的分配算法存在的问题有：（1）没有考虑评审专家的工作量约束：会导致一些候选评审专家被分配给大量的待评审论文，而一些评审专家只有少量的甚至没有论文可以评审；（2）该方法也没有考虑利益冲突约束，有可能造成评审结果的不公平。

同时，一些研究者提出了基于知识规则的评审专家分配算法，但是该方法在一定程度上需要候选评审专家的参与。这样一方面会导致若干评审专家提供的先验信息不够准确，影响后面的评审分配结果；另外一方面，候选评审专家的参与是一项费时、费力的工作，不能够满足越来越多的投稿论文数量和候选评审专家数量的需求。一些机器学习领域的研究者尝试采用基于约束优化的技术解决自动评审专家分配问题，他们将该问题转化为满足一定约束条件的目标函数求极值问题。但是由于约束优化问题是Np问题，这些方法的算法复杂度比较高并且求得的解通常是局部最优解而不是全局最优解。

本文基于概念树计算论文与专家之间的相似度，然后采用基于启发式的最大相似度匹配方法将论文分配给相应的评审专家。基于概念树的相似度计算，可以充分满足主题覆盖度约束；基于启发式的最大相似度匹配算法不仅可以满足利益冲突约束，又可以满足专家工作量约束。最后实验验证了所提算法的有效性。

2 相关工作

2.1 基于信息检索的评审专家推荐

基于信息检索的评审专家推荐方法最早由Dumais和Nielsen[1]提出。他们基于潜在语义索引（Latent Semantic Index,LSI）计算需要评审的论文与候选专家两两之间的相似度。将每篇论文推荐给与其相似度比较高的几位专家进行评审。随后一些研究者基于其他的方法计算专家与论文之间的相似度，但仍采用相似的评审分配方法。如Yarowsky和Florian[2]基于朴素贝叶斯分类方法计算专家与论文之间的相似度，Hettich和Pazzani[3][4]基于TF-IDF方法计算评审专家跟项目申请书之间的相似度，Basu等[4][5]用爬虫的方法搜索评审专家发表的文献摘要，然后基于TF-IDF的方法计算专家和论文之间的相似度。Biswas和Hasan[3][5]采用机器学习算法和基于本体驱动的主题推理的混合

方法来解决评审专家自动分配问题。Cameron 等 [6] 提出了采用语义 web 技术搜集数据，专家与论文作者之间的合著者关系，从而采用基于推理的技术为投稿论文分配评审专家。

2.2 基于约束优化的评审专家推荐

微软开发的会议管理工具包 CMT[13][7] 采用贪婪算法将一篇论文分配给具有较高偏好的专家评审。开源的在线投稿 WEB 系统—CyberChair[14][8]，采用图理论为论文分配评审专家。Philippe Rigaux[15][9] 将评审专家分配问题视为更一般的为用户推荐感兴趣的项目问题，提出了基于协同过滤的评审专家分配问题。Goldsmith 等 [16][10] 从算法复杂度的角度讨论了会议论文评审专家分配问题。Rodriguez 等 [17][11] 提出了基于合著者网络数据结构的粒子云算法。Charlin 等 [18][12] 针对会议论文评审提出了论文和专家之间匹配的优化框架。Conry 等 [19][13] 将会议论文的评审专家分配分解为两个子问题，一个是对审稿人的偏好建模问题，另一个是满足会议评审全局标准的优化问题，相应的优化方法是线性规划。Tang[20, 21][14, 15] 等采用网络流技术，将评审专家分配涉及的各种约束分别转化为图中的流量上下界约束，以及费用约束，从而将评审专家推荐转化为费用最小的网络流问题。Karimzadehgan 和 Zhai[22][16] 讨论了如何将评审专家分配问题转化为整数线性规划问题。Benferhat 等 [23][17] 详细探讨了评审专家分配问题，但是他们采用的目标函数是负效用函数而不是通常所用的效用函数。Juan 等 [24][18] 讨论了如何将贪婪搜索、遗传算法融合在一起解决会议论文的评审专家分配问题。Li 等 [25][19] 将基于偏好的方法与基于主题的方法融合在一起计算评审专家与投稿论文之间的匹配程度，随后将最大匹配程度作为目标函数求最终的评审分配结果。Long 等 [26][20] 首次提出了在评审专家分配时应该最大化分配的评审专家对投稿论文的主题覆盖度。

2.3 其他方法

Nicola 等 [7][21] 描述了会议管理系统 CMS(Conference Management System) 的一个专家模块 GRAPE—会议论文评审专家推荐系统。该系统是基于规则的方法。孙等 [8][22] 针对中国国家自然科学基金申请书的评审专家分配问题提出了基于决策模型和知识规则的混合方法。Tian 等 [9][23] 认为以前对 R&D 项目的评审主要关注于数学决策模型及其应用，忽略了决策支持过程的组织方面，进而提出了面向组织的决策支持过程，并设计开发了 NSFC 的项目申请书提交、申请书评审专家分配，项目结题验收等整个 NSFC 管理系统。吴江宁等 [11][24] 设计了一个基于本体的项目和领域专家匹配原型系统。陈媛等 [12][25] 提出了基于研究领域匹配度的科研项目评审指派方法。

3 基于概念树的文档相似度计算

3.1 概念树

概念树最早由 Praveen L 等 [26] 提出用于文献推荐。为了生成文档的概念树，首先根据一个分类树将文档分类，获得其概念向量。然后将概念向量转化为对应的具有权重的概念树，树中的每一个节点是分类树中的一个类名或者相应的标号，每个节点都有对应的权重，表示文档属于该类别的概率。概念树的生成过程具体如下：

针对概念向量中的每一个类别，递归地向其父节点传递权重，直至到达根节点。根据公式（1）计算父节点的权重：

$$WParent = \alpha * WChild \quad (1)$$

这里，WParent 是父节点的权重、WChild 是孩子节点的权重。α 是权重的传播因子，用来表

示孩子节点的权重传递给父节点权重的程度。若 α =0，那么孩子节点的权重将不传递给相应的父节点。概念向量中的所有概念均为相同分类树中的节点。生成的概念树将保持分类树中相应的层级结构。图 1 是我们用来展示概念树生成过程所使用的分类树示例。

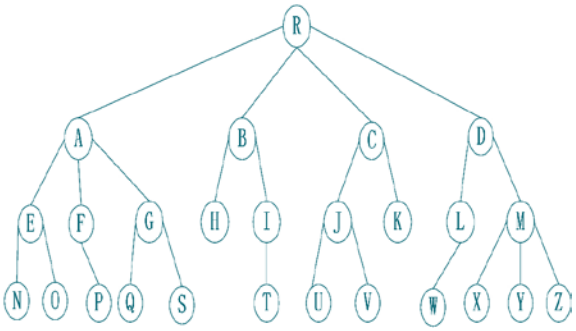

图 1 示例分类树

比如有两个文档 D1 与 D2，其对应的概念向量分别为：

D1={(F, 40), (C, 30), (E, 20), (A, 10)}

D2={(U, 60), (V, 40), (A, 20), (K, 10)}

图 2 展示了如何将文档 D2 的概念向量转化为对应的概念树，这里 α =0.5.

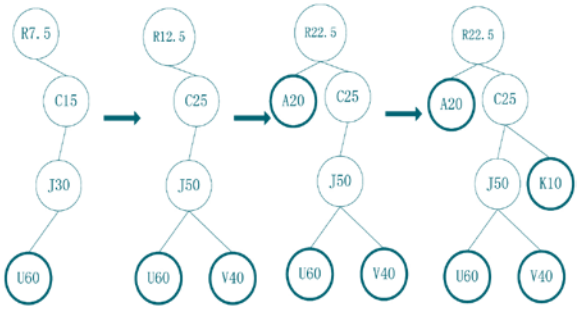

图 2 将文档 D2 转化为概念树 T2 的过程

在图示中，加粗的节点对应原始概念向量中的节点，在整个概念树的构造过程中，概念的个数保持不变。

3.2 基于概念树的相似度计算

本文基于概念树之间的编辑距离计算它们之间的相似度。概念树 A 与概念树 B 之间的相似度为在预定义的操作下从 A 变换到 B 所需要的代价。两个概念树之间的变换代价越小，那么其之间的相似度越高。本文主要采用三个预定义操作：插入、删除和替换。插入操作和删除操作的代价为插入节点或者删除节点的权重；替换操作的代价为要替换的节点与被替换节点的权重之间的差值。由于所有的概念树都是根据同一个分类树构建的，认为没有在概念树中出现的节点其相应的权重为 0。根据分类树构建的文档概念树中的每一个概念相当于文档的一个主题，相应的概念树之间的相似度反应了文档之间的主题相似度。

3.3 专家偏好构建

针对每一个评审专家，将其发表的每一篇论文都转化为对应的概念树，最终的评审专家的概念树是其发表的每篇论文的概念树的和。

4 分配算法

用在评审专家分配中的算法主要有贪婪算法、网络流等。本文借鉴文献 [27] 中的基于最大匹配度分配算法，但是又保证了专家之间的利益冲突。该算法需要在满足多个约束条件的情况下，将所有待评审论文与评审专家之间的匹配程度之和最大化。假设候选评审专家集合为 R，投稿论文集合为 P，那么可以将评审分配问题形式化为：

$$\max \sum_{i \in R} \sum_{j \in P} M(i, j) \times a_{ij}$$

这里，如果为论文 j 分配的专家为 i，则 aij=1，否则 aij=0。

上述目标函数需要满足的约束条件为：

（1）分配给每个评审专家的论文数量要小于等于 n，即 $\sum_{j \in P} a_{ij} \leq n,$

（2）每篇论文正好被 m 个评审专家评审，即，$\sum_{i \in R} a_{ij} = m$

（3）如果评审专家 i 和论文 j 的作者之间有利益冲突，则 aij=0。

本文采用了基于启发式的分配方法，该算法的思想为：如果一个值与平均值之间的距离较大，那么剩下的数据与平均值之间的距离很有可能会更小。以表 1 中的例子进行说明：

表 1 分配算法示例

	1	2	3	average	deviation
a	30	50	100	60(=180/3)	40(=100-60)
b	70	40	100	70(=210/3)	30(=100-70)

表中，尽管 a 与 b 的最大值相同，由于 a 的最大值与其余数据的偏离较大，故 a 其余的数据要小于 b 其余的数据。如果将分配赋予 a3，那么在 b 的剩余数据中发现的赋值将很可能大于在 a 的剩余数据中能够搜索到的赋值。所以，该情况下的赋值应该为 a3,b1。

为了计算每个值与平均值的偏离度，需要构造匹配程度矩阵并计算每行、每列的平均值。相应的每个值的偏离度为：

$$Dij=Mij-rai \quad (2)$$

$$Qij=Mij-caj \quad (3)$$

这里，D 是行偏离度、表示专家 i 与论文 j 之间的匹配程度与 j 的平均匹配程度的差异，Q 是列偏离度，表示专家 i 与论文 j 之间的匹配程度与 i 的平均匹配程度的差异。下图是基于偏离的匹配算法伪代码。需要注意的是，为了满足利益冲突约束，若专家 i 与论文 j 之间存在利益冲突关系，则他们之间的匹配程度为 0。

```
输入：  候选专家与待评审论文匹配程度的矩阵
        每篇论文的评审专家数量 (m)
        每个专家的审稿量 (n)

输出：赋值矩阵
(1) 设置 assignnum=0;
(2) While(assignnum ≠ m*M 的行数 )
(3) {
(4)    For 每一行 ri
(5)    {
(6)       While(ri 中的赋值个数 <m) do
(7)       {
(8)          匹配程度最大的列的 ColumnID → list maxColumnIDs
(9)          If(|maxColumnIDs|>1)
(10)               j=columnID 使得在 maxColumnIDs 中 Q 最大
(11)          Else if(|maxColumnIDs| =1)
(12)               j=maxColumnIDs 中的第一个 ID
(13)          Aij=1
(14)          Mij=0
(15)          Assignnum+=1;
(16)       }
(17)    }
(18) For 每一列 cj
(19) {
(20)    assignRowIDs=cj 中被分配的文章专家 Id
(21)    While ( cj 中的赋值个数 >n )
(22)    {
(23)       i=assignRowIDs 中使得 D 最小的 rowID
(24)       Aij=0
(25)       Assignnum+=-1;
(26)    }
(27) } (28) }
```

5 实验分析

我们在32位处理器Intel Core 2.7GHz的、内存4G Windows 7上对我们所提算法进行了实验。实验数据集由ACM Recsys（推荐系统领域的国际顶级会议）2012，2013，2014三年的会议论文组成，共247篇。候选评审专家为2012年会议的所有会议委员共100位专家。将程序的分配结果跟人工的分配结果进行比较，采用准确率对算法结果进行评价。设系统分配的匹配结果为A，人工标注的匹配结果为R，则相应的准确率的计算为：

$$precision=|A \cap R|/|A| \times 100\% \quad (3)$$

最终的准确率为所有论文准确率和召回率的均值。在概念树的计算中，我们选取α的值分别为0，0.33，0.5，0.67。实验结果如表所示。

	precision
α=0	0.332
α=0.33	0.56
α=0.5	0.618
α=0.67	0.503

从实验结果可以看到，当α=0时，分配的准确率最低，这是因为当α=0时，概念树中低层概念的权重没有向上传递给父节点，这样所对应的概念树仅仅为概念向量。在本试验中当α=0.5时，准确率最高，但是这并不代表所有的实验都是α=0.5时取得的分配结果最好，因为最终的比较结果是人工标注的，不同的人工标注结果会有不同的实验结果。

6 结论

对于会议组委会而言，为投稿论文选择合适的评审专家至关重要。传统的基于关键词与人工辅助的评审论文分配方式由于缺乏语义理解，其合理性及效率都比较低，并且不能考虑到分配过程中的主题覆盖度约束、专家工作量约束等。本文基于概念树计算投稿论文与专家之间的相似度，然后采用基于启发式的最大相似度匹配方法将论文分配给相应的评审专家，所提算法可以同时满足分配过程中的主题覆盖度、专家工作量等各种约束。实验证明，本文所提算法为会议组委会将所有投稿论文分配给评审专家提供了一个新的思路。

参考文献

[1] Dumais S, Nielsen J. Automating the assignment of submitted manuscripts to reviewers [C] // Research and Development in information Retrieval. New York: ACM, 1992: 233-244.

[2] Yarowsky D, Florian R. Taking the load off the conference chairs: towards a digital paper-routing assistant [C] // Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in NLP and Very-Large Corpora,2000. 220-230.

[3] Hettich S, Pazzani M. Mining for proposal reviewers: lessons learned at the national science foundation [C] // Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006.

[4] Basu C, Hirsh H, Cohen W. Recommendation as Classification: Using Social and Content-based Information in Recommendation [C] //Proceedings of the 15th national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence. USA: AAAI, 1998:714-720.

[5] Biswas H K, Hasan M. Using publications and domain knowledge to build research profiles: an application in automatic reviewer assignment [C]//Proceedings of the 2013 workshop on Computational scientometrics: theory & applications, 2013: 19-24.

[6] Cameron D, Aleman-Meza B, ARPINAR B. Collecting expertise of researchers for finding relevant experts in a peer-review setting [C] // First International ExpertFinder Workshop, 2007.

[7] The microsoft conference management toolkit [SW/OL] [2016-04-06]. http://msrcmt.research.microsoft.com/cmt/.

[8] The cyberchair software[SW/OL] [2016-04-06]. http://www.cyberchair.org.

[9] Rigaux P. An iterative rating method: Application to web-based conference management [C] // Proceedings of the ACM International Conference on Applied Computing. New York:ACM, 2004: 1682-1687.

[10] Goldsmith J, Sloan R H. The AI conference paper assignment problem [C] // Proceedings of AAAI 2007 Workshop on Preference Handling for Artificial Intelligence. British Columbia: AAAI, 2007: 53-57.

[11] Rodriguez M A, Bollen J. An algorithm to determine peer-reviewers[C] // Proceedings of the 17th ACM conference on Information and knowledge management. New York:: ACM, 2008:319-328.

[12] Charlin L, ZemeL R, Boutilier C. A framework for optimizing paper matching[R]. Report number: UAI-P-2011-PG-95, 2012.

[13] Conry D, Koren Y, Ramakrishnan N. Recommender systems for the conference paper assignment problem[C] // Proceedings of the third ACM conference on Recommender systems. New York: ACM,2009: 357-360.

[14] Tang Wenbin, Tang Jie, Tao Lei, et al. On optimization of expertise matching with various constraints[J]. Neurocomputing, 2012,76(1): 71-83.

[15] Tang Wenbin, Tang Jie, TAN Chenhao. Expertise matching via constraint-based optimization [C] // Web Intelligence and Intelligent Agent Technology (WI-IAT). New York: ACM, 2010:31-41.

[16] Karimzadehgan M, Zhai Chengxiang, Belford G. Multi-aspect expertise matching for review assignment[C] // Proceedings of the 17th ACM Conference on Information and knowledge management. New York: ACM, 2008: 1113-1122.

[17] Benferhat S, Lang J. Conference paper assignment[J]. International Journal of Intelligent Systems, 2001,16: 1183-1192.

[18] Merelo-G S, Castillo V P. Conference paper assignment using a combined greedy/evolutionary algorithm [C] // Proceedings of 8th International Conference on Parallel Problem Solving from Nature. Berlin: Springer, 2004: 602-611.

[19] Li X L, Watanabe T. Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers[C]// Proceedings of 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, 2013, 22:633-642.

[20] Long C, Wong Rcw, Peng Y, et al. On good and fair paper-reviewer assignment[C]// Proceedings of 13th International Conference on Data Mining, 2013:1145-1150.

[21] Nicola d, Teresa M, Stefano F. Grape: an expert review assignment component for scientific conference management systems[C] // Proceedings of the 18th International Conference on Innovations in Applied Artificial Intelligence, 2005: 789-798.

[22] SUN Y H, MA J, FAN Z P, et al. A hybrid knowledge and model approach for reviewer assignment[J]. Expert Systems with Applications, 2008, 34(2): 817-824.

[23] Tian Q J, Ma J, Liang JZ, et al. An organizational decision support system for effective R&D project selection[J]. Decision Support Systems, 2005, 39(3): 403-413.

[24] 吴江宁 , 杨光飞 . 基于本体的项目和领域专家匹配原型系统 [J]. 计算机应用研究 , 2009, 26(10): 3787-3790.

[25] 陈媛 , 樊治平 . 基于研究领域匹配度的科研项目评审指派方法 [J]. 中国管理科学 , 2011, 19(2): 169-173.

[26] Kannan C, Susan G, Preveen L, et al. Concept-based Document Recommendations for CiteSeer Authors[C] // Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 2008: 83-92.

[27] Li X L, Watanabe T. Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers[J]. International Journal of Knowledge and Web Intelligence, 2014, 5(1): 1-20.