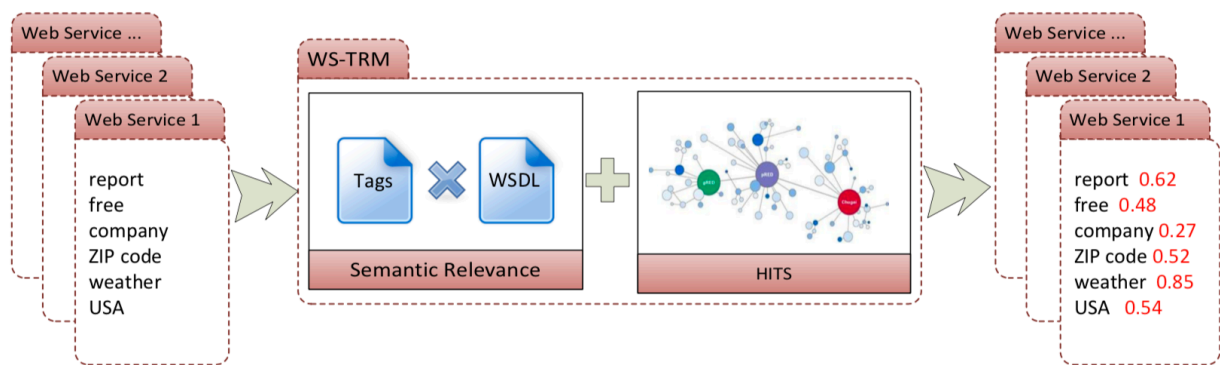


SEAI论文报告



Modeling and Exploiting Tag Relevance for Web Service Mining

雍琪 18214839

2018 年秋季

摘要

在大量的网络服务中，有许多用户定义的标签，用来对网络服务的功能、特性进行注释，这对网络服务的分类、检索等业务功能有很大的影响，所以标签的准确性非常重要。当下存在的问题是，标签的质量良莠不齐，且排序大多按照用户的添加顺序，有些其实与网络服务无关，甚至有些标签是被用户恶意添加的，如果依赖这种顺序的标签进行聚类、检索、推荐，将会非常低效，故而需要一套标准对标签的质量进行衡量，选取真正有效的标签用来注释网络服务。

本篇论文针对现状提出了标签的必要性、以及标签的有效度量算法，算法的基础是语义相关度计算和HITS模型，分别计算每个标签与网络服务的相关度和标签各自的重要程度，将两个计算结果加权求和，以此结果进行排序得到标签列表。

本文主要分为以下几个部分：1、论文主要内容，包括核心思想和主要创新点；2、本人对负责部分（实验和实际应）的论述和相关工作；3、论文改进方案的思考方向。

目录

1 主要内容	1
1.1 核心思想	1
1.1.1 相关度计算	1
1.1.2 重要性计算	1
1.1.3 加权求和	2
1.2 主要创新点	2
2 本人工作	3
2.1 实验	3
2.1.1 NDCG算法	3
2.1.2 实验结果	4
2.2 实际应用	4
3 改进方案的思考	7
参考文献	8

1 主要内容

1.1 核心思想

本篇论文的核心思想是基于标签的语义相关性和重要性，建立WS_TRM模型，获取真正有效、优质的标签序列。

1.1.1 相关度计算

网络服务由WSDL描述，WSDL（Web Service Description Language）包含Web Service的消息类型、消息、端口、操作等元素，是一个XML文档。标签语义相关度的计算前提是从WSDL中提取关键词等价代替Web Service。相关度计算的步骤有以下几个部分：

- 1) 粗糙分词。对WSDL基于空格或大小写进行初步的分词，例如WeatherReport分为Weather和Report；
- 2) 去除后缀。对类似的但后缀不同的词汇进行去重，即connected转换为connect；
- 3) 剪枝。去掉无意义词汇以及与主要内容关联较小的词汇，关联较小词汇基于出现频率得到；
- 4) 提炼。通过前三步得到的关键词有可能过于“大众化”，无法很好地区分Web Service，故而需要剔除其中意义过于普通的词汇，例如Report、building等等，至于如何界定这一类词汇，论文作者采用了计算频度设置阈值的方式，主要工作沿用文献[1][2]。

通过以上步骤将得到最终的content vector，以[w1, w2, w3, ...]方式存储，并对四种重要类型的word赋予权重，即名、值、文本、注释，权重分别f1, f2, f3, f4，接下来根据公式（1-1）、公式（1-2）、公式（1-3）计算每一个tag与content的相关度。

$$SR(t, content) = \frac{\sum_{i=1}^n Sim(t, W_i) \sum_{j=1}^4 f_j \times Occur_{ij}}{\sum_{i=1}^n \sum_{j=1}^4 f_j \times Occur_{ij}} \quad (1-1)$$

$$sim(t, w_i) = 1 - NGD(t, w_i) \quad (1-2)$$

$$NGD(t, w_i) = \frac{\max(\log^{f(t)}, \log^{f(w_i)}) - \log^{f(t, w_i)}}{\log^N - \min(\log^{f(t)}, \log^{f(w_i)})} \quad (1-3)$$

1.1.2 重要性计算

该部分主要基于HITS算法。HITS是用来计算网页重要性的算法，Hub页面（枢纽页面）和Authority页面（权威页面）是HITS算法最基本的两个定义。“Authority”页面是指与

某个领域或者某个话题相关的高质量网页，比如搜索引擎领域的Google和百度首页；“Hub”页面指包含了很多指向高质量“Authority”页面链接的网页，比如hao123首页。依据这一核心思想，建立STNet（service-tag网络），该网络的连线包括service-service、service-tag、tag-tag，然后计算出每个节点的authority值，返回所有tag的authority。计算公式如下，其中所有节点的auth值和hub值用0初始化。

$$Auth(p) = \sum_{i=1}^n Hub(p_i) \times w(p_i, p) \quad (1-4)$$

$$Hub(p) = \sum_{i=1}^n Auth(p_i) \times w(p, p_i) \quad (1-5)$$

1.1.3 加权求和

最后将1.1.1和1.1.2两个部分求到的SR与Auth按照公式（1-6）进行加权求和。

$$Score(t, s) = (1 - \Lambda)SR(t, s) + \Lambda Auth(t) \quad (1-6)$$

当 Λ 为1时，表示只考虑标签在整个STNet中的重要性，为0时，表示只考虑标签与WSDL的相关程度。文章后文详细阐述了 Λ 的选择。

1.2 主要创新点

本文的创新主要在两个方面，一是对网络服务各项功能提出了一个重要衡量指标，即标签的相关性，前期大部分研究都是基于标签的相关度均为1这一前提，这个显然不成立的前提使得很多应用研究不具有可用性；二是基于HITS算法衍生出STNet，提出计算标签重要性的算法，建立效率高、实用性强的WS-TRM模型。

2 本人工作

我针对这篇论文主要研究了后半部分，算法的验证实验和实际应用，在这部分，作者主要阐述了WS-TRM模型对比原始算法的优势，以及 Λ 的不同取值对算法的影响。本人主要针对实验指标排序的准确性进行了调研分析，主要研究了NDCG算法，另外在实际应用中针对文献[3]提到的推荐标签模型进行了研究。

2.1 实验

实验数据来源是六大类总共240个Web Service，针对每个Service所拥有的标签进行人工评分，评分等级从1到5，相关度依次递增，测试指标是排序的准确性，计算基于NDCG算法。

2.1.1 NDCG算法

本篇论文的排序评估是一个Top-k问题，我们只关注排序中前若干名，因为最终被用作聚类、检索的只会是评分较高的标签，基于这个问题，采用了NDCG算法，一个常用的排序模型评价指标。排序中的每一项得到的评分被称为增益（gain），表示为 rel_i ，如果评分高但是排序靠后，说明这个排序的精确度有一定折损，这个折损通常是该项位置的对数，我们将每一项的增益除以各自的折损值，通过公式（2-1）得到该排序的DCG；另外由于这里的k值不定，所以需要DCG进行规范化进行统一的比较，根据公式（2-2）得到一个0到1的数，可以认为是预测排序的精确度和理想排序的精确度之间的差距，其中理想排序的精确度表示为IDCG，也就是序列中所有项的评分都是单调递减时的DCG值。

$$DCK@k = rel_i = \sum_{i=2}^k \frac{rel_i}{\log_2 i} \quad (2-1)$$

$$NDCK@k = \frac{DCG@k}{IDCG@k} \quad (2-2)$$

NDCG算法是目前相当流行的排序评估模型，它具有两个优势^[4]，1、拥有细化的评分等级，即上文提到的1-5分分别代表不同的相关程度，而其他评估算法大多只有“相关”或“不相关”两种评价；2、折损函数的引入，这一点很重要，大部分算法对每一个位置都平均分配了权值，这不能有效代表排序的精确性。这其中的折损函数也是各项研究讨论比较多的部分，为什么要采用以2为底的对数来计算？关于这一点没有十分详细的解释，主要还是因为该对数能让增益得到一个相对平滑的减少。

2.1.2 实验结果

实验结果分为两个方面，首先分别用三个方法对标签进行排序，一是基于与WSDL的相关度提取出标签，二是基于HITS算法根据重要性提取标签，三是两者的结合。图2.1中以原始的标签排序为参考，可以看出三种算法在六大类Web Service都表现优良，当k值增大时，精确度普遍有所下降。其中WS-TRM算法中需要对 Λ 进行评估选择，图2.2是不同 Λ 取值情况下对应的排序精确度，得到最优取值为0.1。

TABLE 1
NDCG@K Performance of Web Service Tag Relevance Measuring Approaches

NDCG@K	Method	Tourism	Weather	Commu	Finance	Stock	Email	Average
K=3	Baseline	0.723	0.831	0.741	0.637	0.781	0.839	0.759
	Semantic	0.793	0.878	0.912	0.869	0.921	0.934	0.884
	HITS	0.823	0.931	0.852	0.791	0.913	0.893	0.867
	WS-TRM	0.863	0.956	0.962	0.893	0.941	0.952	0.928
K=5	Baseline	0.705	0.863	0.725	0.747	0.841	0.852	0.789
	Semantic	0.805	0.912	0.905	0.913	0.947	0.936	0.903
	HITS	0.794	0.908	0.913	0.828	0.913	0.894	0.875
	WS-TRM	0.841	0.965	0.958	0.931	0.967	0.959	0.937

图2.1 不同算法的排序精确度比较

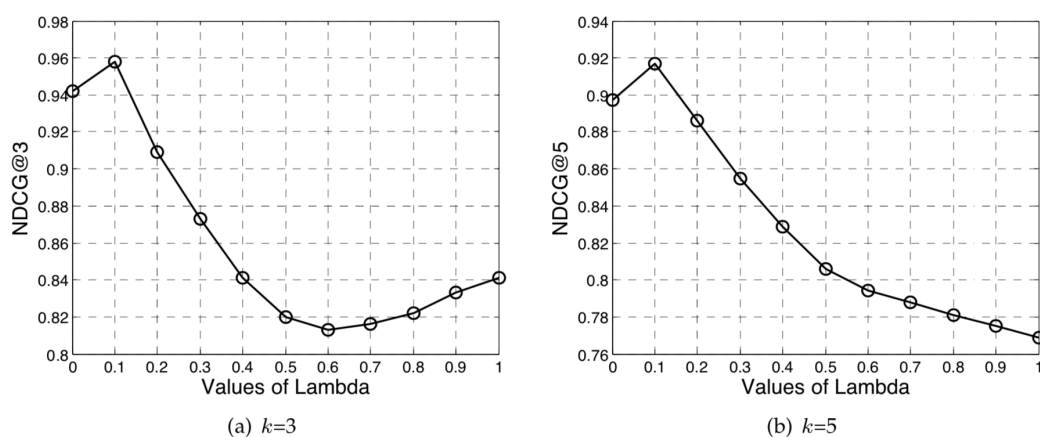


Fig. 4. Impact of λ to The Performance of WS-TRM

图2.2 不同 Λ 取值的排序精确度走势图

2.2 实际应用

作者提出WS-TRM模型可以应用于1、网络服务的聚类^[9]，主要解决由于关键词匹配失准问题^{[6][7]}；2、；3、基于标签数据的网络服务检索^[8]。我主要针对标签推荐部分做了深入的了解，另外两部分作者没有做详细的阐述。

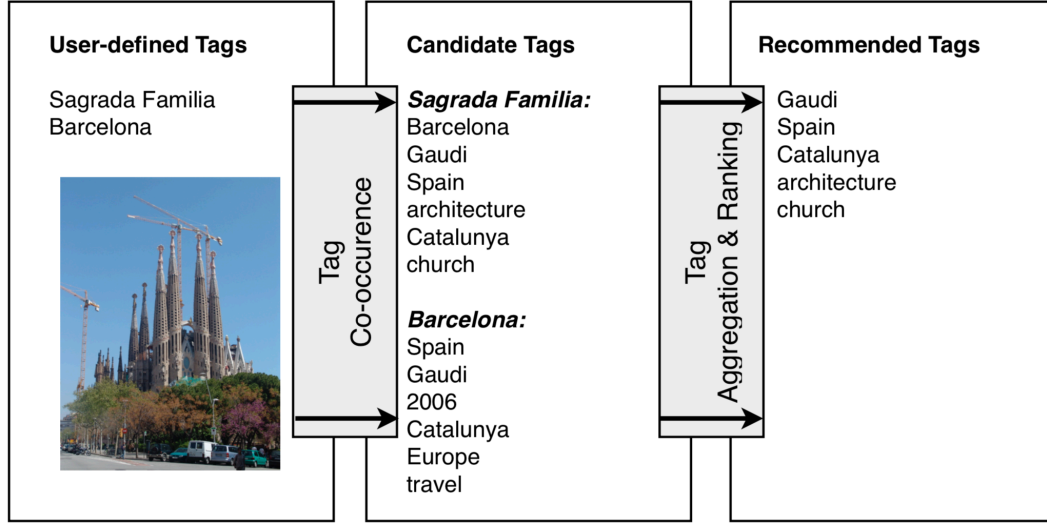


Figure 4: System overview of the tag recommendation process.

图2.3 推荐标签的生成流程

图2.3是文献[3]中生成推荐标签的主要流程，该文献主要针对网络中的图片，引用到本篇论文里就是Web Service，首先获取到Web Service的原始标签，也就是user-defined tags，然后在实验数据集中，分别计算user-defined tag与其他标签的共现系数，计算方式分为对称算法（公式2-3）和非对称算法（公式2-4），其中t分别表示对应标签在多少个网络服务中出现。

$$J(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \quad (2-3)$$

$$P(t_j | t_i) = \frac{|t_i \cap t_j|}{|t_i|} \quad (2-4)$$

以上步骤得到了每一个原始标签的共现标签，选取得分最高的前若干位形成候选标签列表，然后通过聚合得到最终的推荐标签。原文献中采用了vote和sum两种算法，本篇论文则在这一基础上引入该研究中的标签相关度得分，加权求和后进行排名，进行与上述实验类似的测试对排序精确度做评估，测试指标为S@k和P@k，分别表示序列中评分为最高两档的标签比例，以及与Web Service相关度高的比例。最后图2.4显示，基于WS-TRM的标签推荐有更高的准确率，图2.5显示，准确率与k值、标签总个数成反比。

综合以上多个实验，可以看出，算法在k值不同时，精确度还是存在一定的局限性，因为一个Web Service的相关标签是有限的，随着k越大，列表就会包含越多的无关标签，所以精确度必然会降低。

图2.4 标签推荐优化算法的精确度对比

TABLE 2
S@K Comparison of Four Tag Recommendation Strategies

Given Tag	Method	K=1	K=2	K=3	K=4	K=5
1-2	Sum	0.8132	0.7081	0.6738	0.7087	0.7181
	Sum ⁺	0.8331	0.7129	0.7033	0.7221	0.7318
	Vote	0.6392	0.5949	0.6737	0.7005	0.6972
	Vote ⁺	0.6875	0.6112	0.6745	0.7143	0.7384
3-5	Sum	0.7534	0.7143	0.7380	0.6852	0.6720
	Sum ⁺	0.7745	0.7322	0.7449	0.7208	0.6775
	Vote	0.7867	0.6646	0.7042	0.7022	0.7103
	Vote ⁺	0.7958	0.7436	0.7323	0.7128	0.7219
>5	Sum	0.7632	0.7211	0.6944	0.6975	0.6647
	Sum ⁺	0.7822	0.7318	0.7098	0.7145	0.6897
	Vote	0.8136	0.7769	0.7749	0.7282	0.6973
	Vote ⁺	0.8364	0.8012	0.7943	0.7438	0.7012

TABLE 3
P@K Comparison of Four Tag Recommendation Strategies

Given Tag	Method	K=1	K=2	K=3	K=4	K=5
1-2	Sum	0.6933	0.5083	0.4277	0.3788	0.3562
	Sum ⁺	0.7612	0.5329	0.4879	0.4374	0.4038
	Vote	0.7879	0.5495	0.4503	0.3947	0.3689
	Vote ⁺	0.7945	0.5983	0.4832	0.4329	0.3925
3-5	Sum	0.6512	0.4857	0.4171	0.3654	0.3345
	Sum ⁺	0.6856	0.5134	0.4658	0.3765	0.3564
	Vote	0.7415	0.5414	0.4496	0.3925	0.3494
	Vote ⁺	0.7667	0.5934	0.5092	0.4333	0.3764
>5	Sum	0.5894	0.4656	0.4365	0.3451	0.3508
	Sum ⁺	0.6219	0.5043	0.4754	0.3922	0.3657
	Vote	0.7148	0.5478	0.4105	0.4026	0.3638
	Vote ⁺	0.7443	0.5874	0.4459	0.4322	0.3745

3 改进方案的思考

论文中有提到实验中人工数量的不足，有可能导致测试的准确性，的确这种相关度的评估无法用机器进行，只能大量运用人力，其准确性是与参与人数呈正比的。除了人力，网络数据资源也是一大局限，论文中出现的数据来源太窄，后期改进可以增加更多专业类别的Web Service，建立更大的训练集。

另外，上文提到精确度与k值呈反比，也是可以优化的一个方向，目前没有想到很好的优化方案。

参考文献

- [1] W. Liu and W. Wong, “Web service clustering using text mining techniques,” *International Journal of Agent-Oriented Software Engineering*, vol. 3, no. 1, pp. 6 – 26, 2009.
- [2] K. Elgazzar, A. E. Hassan, and P. Martin, “Clustering wsdl documents to bootstrap the discovery of web services,” *International Conference on Web Services*, pp. 147 – 154, 2009.
- [3] B. Sigurbjörnsson and R. van Zwol, “Flickr tag recommendation based on collective knowledge,” *Proc. of the 17th international conference on World Wide Web (WWW)*, pp. 327 – 336, 2008.
- [4] Y. Wang et al., “A theoretical analysis of NDCG ranking measures,” in *Proc. 26th Annu. Conf. Learn. Theory*, 2013, pp. 1 – 26.
- [5] E. Kanoulas and J. A. Aslam. “ Empirical justification of the gain and discount function for ndcg. In To appear in CIKM’ 09: Proceedings of the 18th ACM international conference on Information and knowledge management, 2009.
- [6] W. Liu and W. Wong, “Web service clustering using text mining techniques,” *International Journal of Agent-Oriented Software Engineering*, vol. 3, no. 1, pp. 6 – 26, 2009.
- [7] K. Elgazzar, A. E. Hassan, and P. Martin, “Clustering wsdl documents to bootstrap the discovery of web services,” *International Conference on Web Services*, pp. 147 – 154, 2009.
- [8] L. Chen, L. Hu, Z. Zheng, and J. Wu, “Wtcluster: Utilizing tags for web services clustering”, *Proc. of the 9th International Conference on Service Oriented Computing (ICSOC)*, pp. 204 – 218, 2011.
- [9] 黄媛.一种基于标签推荐的服务聚类方法[J].计算机与数字工程, 2017,45(6): 1141-1144.
- [10] X Li , CGM Snoek , M Worring, “Learning Social Tag Relevance by Neighbor Voting”, *IEEE Press* , 2009 , 11 (7) :1310-1322