基于隐主题分析和文本聚类的微博客新闻话题发现研究

路荣，项亮，刘明荣，杨青

中国科学院自动化研究所模式识别国家重点实验室，北京，100190

E-mail: (rlu, lxiang,mrliu, qyang)@nlpr.ia.ac.cn

摘　要：本文研究在大规模微博客文本集上的话题发现问题。微博客与传统博客不同，首先，它的长度比传统博客短，往往只有只言片语。其次，它可以通过手机，即时通讯软件等，实时发布，从而会在短时间内产生大量数据。对于微博客的短文本数据，传统使用词作为特征来表示文本的方法，会由于同一个词共现在两篇不同短文本中的概率较小，而无法度量它们之间的相似度。本文使用隐主题模型，充分挖掘短文本的隐主题信息，并在隐主题空间上度量短文本之间的相似度，从而有效解决了短文本的数据稀疏性问题。另一方面，对于大规模的数据，传统直接利用聚类方法聚合出新闻话题的方法，很难快速得到理想结果。而本文则首先根据新闻的特点，选择出最有可能谈论新闻事件的微博客，然后用一种两层的 K 均值和层次聚类的混合聚类方法，将选择出的微博客快速准确地聚合成不同的新闻话题。实验结果表明，本文的方法能有效地从大规模微博客短文本数据集中，挖掘出新闻话题。

关键词：微博客，短文本，隐主题模型，话题发现，混合聚类

Extracting News Topics from Microblogs based on Hidden Topics Analysis and Text Clustering

Rong Lu, Liang Xiang, M.R. Liu, Qing Yang

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences Beijing 100190

E-mail: (rlu, lxiang, mrliu, qyang)@nlpr.ia.ac.cn

Abstract: This paper focused on the task of news topics extraction from large-scale short posts of microblogging service. Microblog is very different from traditional blog. First, it is consisted of from several words to dozens of sentences. That is much shorter than the traditional blog. Second, it can be posted in real-time by mobile phone, instant messaging software and so on, which results in a huge number of posts in a very short period. For the shortness and sparseness of the microblog text, traditional VSM(Vector Space Model) using words or terms as characters can not reach desired accuracy. So, in this work, hidden topics discovering was performed on the whole dataset to reduce the sparseness and make the data more topic-focused. For the large-scale of posts, we first selected the microblogs which are most likely to talk about news events. Then a two-level K-means-hierarchical hybrid clustering method was chosen to cluster all the selected tweets to different news topics. Experimental studies show our method works well on large-scale microblog dataset.

Keywords: microblog, short text, hidden topics model, news topics extraction, hybrid clustering

1　引言

微博客（简称微博），是一个基于用户关系的信息分享、传播以及获取平台，用户可以通过 WEB、WAP 以及各种客户端组件个人社区，以很短的文字更新信息，并实现即时分享。最早也是最著名的微博是美国的 Twitter，根据相关公开数据，截至 2010 年 1 月份，Twitter 在全球已经拥有 7500 万注册用户。而在国内，门户网站新浪网于 2009 年 8 月份首先推出“新浪微博”，至此微博正式进入中文上网主流人群视野。

路荣，男，1985 年生，博士研究生，主要研究方向为信息检索、数据挖掘。E-mail：rlu@nlpr.ia.ac.cn， 电话：01062632273

微博的内容简单发布便捷，使得新闻事件可以在微博中快速传递。因此从微博中检测出这些新闻话题，对舆情监控，信息安全，金融证券，行业调研都有十分重要的意义。

而传统的从新闻网页和博客中检测新闻话题的研究，通常使用网页和博客文本中的词作为特征，将文本表示成为一个特征向量，并使用 tf-idf 办法来衡量每个特征（即向量每一维）的权重。再使用聚类的方法，将描述同一新闻事件的网页和博客聚到同一个类中。

但对于微博来说，由于它的文本内容非常短，那么同一个词出现在不同短文本中的概率会远小于普通长度的文本，这种数据的稀疏性，会使得传统以词或短语为特征的向量表示方法，很难准确计算文本间的相似度。为此，本文使用隐主题建模的方法，挖掘出每个微博客短文本的隐主题信息，再将每个短文本表示成隐主题空间上的一个向量，从而有效减少数据稀疏性对度量文本之间相似度的影响。

另一方面，每时每刻都有大量微博数据产生。根据相关公开数据，目前 Twitter 上，每天更新的微博总数已超过 5 千万。如此大规模的数据，如果直接使用聚类算法，效率和质量都难以得到保证。并且，也不是所有的微博都是描述新闻事件的，很多微博只是描述用户心情，状态，工作情况等等。

所以，本文根据新闻的特点，首先将描述新闻事件的微博选取出来，再使用聚类的方法。但尽管如此，选取出来的微博，数量仍可能十分巨大，常用的层次聚类的方法时间上将无法忍受，K 均值聚类的方法，在类中心个数 K 远小于微博总数时，聚类速度将大幅提升，但提前指定类中心个数（新闻话题的个数），这显然并不容易做到。所以，本文采用一种两层的 K 均值和层次聚类的混合聚类方法来克服这两种聚类方法的缺点。

实验结果表明，本文提出的方法，可以较好地完成从大规模微博客数据中快速提取新闻话题这一目标。

2 相关工作

目前，有很多方法尝试解决短文本的数据稀疏性问题，一种办法是通过搜索引擎来扩展短文本的上下文［7,14］，这种办法的缺点是非常耗时，尤其对即时系统非常不适合。另一种办法是通过隐主题模型来给短文本建模，将短文本表示成隐主题按一定比例的混合，它的好处是能够充分挖掘文本集合的内在信息，从而减少短文本的数据稀疏性的影响。隐主题建模研究有很多［6,11,10,9］。其中，Latent Semantic Indexing(LSI)［9］是通过构造文本特征向量矩阵，然后对该矩阵进行奇异值分解来实现的。Probabilistic Latent Semantic Indexing(PLSI)［12］则在 LSI 的基础上提出一个有坚实统计理论基础的生成模型。而 Latent Dirichlet Allocation(LDA)［6］则是一个完全的文本生成模型，其基本思想是，每个文本都是由多个隐主题混合而成，而每个隐主题又由多个词混合而成。

本文的目的旨在从大量的微博数据中发现新闻话题，并将描述同一新闻事件的微博聚合到一起，这和话题发现与跟踪领域(Topic Detection and Tracking, TDT)的研究十分相似。话题发现与跟踪的作用在于自动的将相关话题的信息汇总，以供人查阅［1,2,3］。目前，它的主要研究对象是新闻报道和博客，关注点较多包括报道切分、话题跟踪、话题发现、新事件发现和报道关联发现，研究数据多采用 TREC 会议提供的 TDT 语料［5,8,4］，规模相对本文的研究对象较小。

3 基于隐主题挖掘和文本聚类的微博客新闻话题发现

3.1 数据准备

为了本文的研究工作，我们从国外最著名的微博网站 Twitter，选取了 21302 个用户，

抓取他们从 2010 年 2 月 24 日至 3 月 17 日，共 22 天，发表的所有微博数据。在做过去除停用词处理后，选取长度为 4 个单词以上的微博客文本共 3,079,860 条。

3.2 方法思想和基本框架

对于大规模的微博客短文本，从中发现新闻话题，需要克服两个难点。一是，如何表示短文本进行有效地相似度度量；二是，如何快速准确的处理大规模的微博数据。

本文的办法是，首先通过隐主题模型，将微博客短文本集的隐主题充分挖掘出来，已减少短文本的数据稀疏性，对文本表示的影响。然后，根据新闻的特点，从大量数据中选取出最有可能描述新闻话题的微博。最后使用一种快速准确的混合聚类的方法，将选取出来的微博，通过聚类聚合成不同的新闻话题。

本文提出的方法框架可以用图 1 来表示。

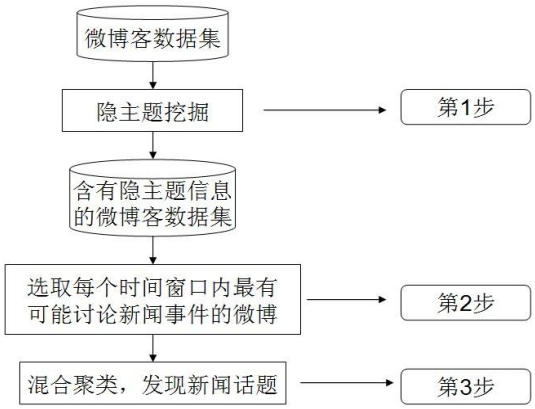

图 1 整体框架图

Fig. 1 General framework

下面将依次介绍上面提出方法的三个步骤。

3.3 隐主题挖掘

前面提到过，目前常用的隐主题建模的有 LSI、PLSI、LDA 等方法。LDA 模型相对于 LSI 和 PLSI 模型具有清晰的层次结构，是一个完全的生成模型。所以本文选用 LDA 模型来对收集的 Twitter 数据集进行隐主题建模，

3.3.1 LDA 模型简介

LDA 模型是一种完全的文本生成模型，它可以将单个文本表示为所有隐主题的特定比例的混合。如图 2 所示，为 LDA 模型的贝叶斯网络图。

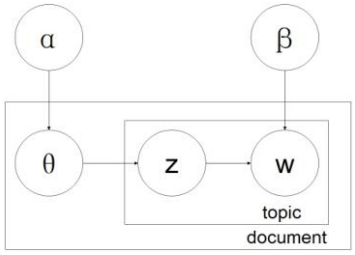

图 2 LDA 图模型

Fig.2 Graphical model of LDA

在文档的生成过程中，LDA 首先从 Dirichlet 分布中抽样产生一个文本特定的主体多项式分布；然后对这些主题反复抽样，产生文本中的每个词。如图 4 所示，α 和 β 是文本集合的参数，α 反映了文本集合中隐主题的相对强弱，β 刻画了所有隐主题自身的概率分布。随机变量 θ 其分量表示目标文档中个隐含主题的比重。z 表示目标文档分配在每个词上的隐主题比重，w 是目标文档的词向量表示。更详细的 LDA 模型的描述请参见［8］。

3.3.2 LDA 建模结果

对 LDA 模型的参数进行估计的办法也有很多，常用的有 EM 算法和吉布斯采样（Gibbs sampling）。本文选用吉布斯采样方法推断 LDA 模型的参数，使用 GibbsLDA++［13］对收集的 Twitter 数据集建模。根据经验，隐主题数目取 200，超参数 α 取 0.25，β 取 0.1。

结果如下：

1. Φ，一个 $K \times V$ 的矩阵，K 表示隐主题的个数，V 文本集合中所有不同词的个数。它表示了每个隐主题生成每个词的概率。
2. Θ，一个 $M \times K$ 的矩阵，M 是数据集中文本总数，K 是隐主题个数。它表示了数据集中每个文本生成隐主题的概率。
3. Z，它的每一行都是原数据集中的一个文本，但和原数据集中不同的是，所有文本中的每个词都被标记到某一个隐主题中了。

3.3.3 单义词单元

如果一个词 w 在文本集中两个不同的位置（不同文本中，或同一文本的不同位置）出现，并被分配到两个不同的隐主题 t_i 和 t_j 下，因为 LDA 假设各个隐主题之间是相互独立的，所以 $w:t_i$ 和 $w:t_j$，可以看作 w 的两个不同词义，即 w 为多义词。我们将 $w:t$ 的形式定义为一个单义词单元。

例如“jobs”这个词。在隐主题建模之后，有两个隐主题产生该词的概率较大，表 1 列出了这两个隐主题及其中发生概率较大的词。

表 1 两个产生“jobs” 概率较大的隐主题

Tab.1 Two hidden topics which are most possible to generate “jobs”

隐主题序号	隐主题生成概率最大的若干个词
Topic 24	job pay work works bills customers jobs stuff creative tools workers
Topic 56	Iphone app ipad apple touch store ipod software artist itunes jobs

在一些位置上，“jobs”表示和“工作”相关的意义被分配到隐主题“Topic 24”上,另一些位置上“jobs”表示美国苹果公司总裁的姓名，被分配到了隐主题“Topic 56”上。本文把“jobs：Topic 24”和“jobs：Topic 56”称为一个单义词单元。这样每篇 tweet 都可以表示成一些单义词单元组成的向量，如 $d = (w_1:t_1, w_2:t_2, w_3:t_3, ..., w_n:t_n)$。3.3.2 节的结果 3 就是这种形式的文本向量集合。

3.4 选取新闻微博

在获取每个微博文本的隐主题信息后，我们将利用新闻的一些特点，从大规模的微博客数据中选取出那些最有可能讨论新闻话题的微博。本文采用的办法是，首先给数据集中出现的所有单义词单元 $w:t$ 打分，然后每个 tweet 的得分就是其包含的所有单义词单元得分之和。得分最高的 tweets 被认为最有可能谈论新闻话题。

此外，之所以选择单义词单元，而不只仅仅是给词打分，是因为单义词单元可以有效地解决多义词问题，对准确选取新闻微博有帮助。后面的实验证明这样做确实可以取得更

好的效果。

3.4.1 给单义词单元打分

一般新闻有两个重要特性，首先新闻讨论的内容非常新，之前很少出现过相似内容，而在某个时段忽然出现。其次新闻的内容往往十分重要，且具有重大影响力或是极具争议性，因而在出现后的短期内，会引起大量关注，并出现大量相关内容的讨论。

根据这两个特性，先将前面得到所有数据按时间顺序分配到若干个时间窗口中，可以有以下结论：如果一个单义词单元在某个时间窗口内相比前一个时间窗口内出现的次数明显增多，可以认为它和一些新的话题相联系；如果一个单义词单元在某个时间窗口内，出现的次数比该时间窗口内其他单义词单元明显多，那么认为它和一些重大，热门话题关联。

为此，我们作如下定义：

单义词单元 $w:t$ 的文档频率 df ，和用户频率 uf ：

$$df(w:t)=\left|tweets:(w:t)\in tweet\right|$$

$$uf(w:t)=\left|users:(w:t)\in tweet\in users\right|$$

df 描述了讨论某个单义词单元 $w:t$ 的 tweets 数量，uf 描述了讨论某个单义词单元 $w:t$ 的用户数量。那么，在当前时间窗口内单义词单元 $w:t$ 的文档频率可以表示为 $df_c(w:t)$，用户频率为 $uf_c(w:t)$。前一时间窗口中则为 $df_h(w:t)$ 和 $uf_h(w:t)$。

那么单义词单元文档频率和用户频率在两个时间窗口间的变化则为：

$$\Delta(df)=df_c(w:t)-df_h(w:t)$$

$$\Delta(uf)=uf_c(w:t)-uf_h(w:t)$$

最后给出打分规则如下：

如果 $df(w:t)>\lambda_1$，并且 $uf(w:t)>\lambda_2$，那么单义词单元 $w:t$ 的得分可以使用下面的公式：

$$score(w:t)=\frac{\Delta(df(w:t))}{1+df_h(w:t)}+\frac{\Delta(uf(w:t))}{1+uf_h(w:t)} \tag{1}$$

否则，$score(w:t)=0$。

$df(w:t)>\lambda_1$，并且 $uf(w:t)>\lambda_2$ 表明了该单义词单元 $w:t$ 在当前时间窗口内，讨论的 tweets 和用户都比较多，反映其在当前时间内的重要性；公式(1)表示了该单义词单元 $w:t$ 的文档频率和用户频率在当前时间窗口和相比前一时间窗口中的增加率，同时反映了其“新”和重要的程度。

3.4.2 选取最有可能谈论新闻话题的微博

最后给所有当前时间窗口内的 tweet 打分，每个 tweet 的得分使用公式(2)计算：

$$score(tweet)=\sum_{n=1}^{N}tfidf(w_i:t_i)\times score(w_i:t_i) \tag{2}$$

给打完分的 tweet，按照分由高到低排序，选取前面若干个，视为最有可能讨论新闻事件的微博，作为下一步聚类的数据。

3.5 对选取的新闻微博聚类

3.5.1 微博短文本的隐主题空间向量表示和相似度度量

在 4.4.2 章提到的 LDA 建模结果中，结果 2 是一个 $M\times K$ 的矩阵，M 是数据集中文本总数，K 是隐主题个数。它表示了数据集中每个文本生成隐主题的概率，也可以看作每个文本在 K 维隐主题空间上的分量值。

本文中隐主题个数 K 选取了 200，那么每个 tweet 都可以表示成一个在 200 维隐主题空间中的向量。这样有效地避免了短文本数据稀疏性，导致的文本见相似度无法准确度量的问题。

因为每篇文本在 K 个隐主题上的分布都是从同一分布中采样出来的［6］，所以，我们可以用 KL 散度来度量两个文本之间的距离。

对于随机变量 X 和 Y，其 KL 散度定义为：

$$D_{kl}(X \| Y) = \sum_{n=1}^{N} p(x=n) \log \frac{p(x=n)}{p(y=n)} \tag{3}$$

由于 KL 散度并不具有对称性，即 $D_{kl}(X \| Y) \neq D_{kl}(Y \| X)$，所以它并不是很好的距离度量，本文使用它的一个更平滑的具有对称性的变形形式，即 Jensen-Shannon 距离：

$$D_{js}(X \| Y) = \frac{1}{2}\left[D_{kl}(X \| M) + D_{kl}(Y \| M)\right] \tag{4}$$

其中 $M = \frac{1}{2}(X+Y)$

3.5.2 两层的 K 均值和层次聚类的混合聚类

如前所述，微博客如 Twitter 的数据量非常大，即使我们选出了最有可能谈论新闻事件的 tweets，数据仍非常多。传统的聚类方法，如层次聚类速度非常慢，根本不适用于大规模数据集；K 均值聚类，则难以提前指定类的数目。K 值太大了，讨论同一新闻话题的 tweets 可能无法聚入同一个类；K 值太小，讨论不同新闻话题的 tweets 势必聚入同一个类中。

所以，本文采用一种两层的混合聚类方法。首先用 K 均值聚类方法做第一层聚类，选取一个适当大的类数目 K（K 仍应远小于选取出来的微博数量），这样可以充分发挥 K 均值聚类速度快的优点。然后，对 K 均值聚类的结果，给定合适的阈值，再使用层次聚类，直到所有类之间的距离大于该阈值。

4 实验与结论

本文的针对大规模的微博客数据集，提出了基于隐主题挖掘和文本聚类的方法，来实现新闻话题的发现。主要工作有三步，见图 1。实验中，我们设定时间窗口的长度为一天。

第一步的实验结果在 3.3.2 中，做过说明。第二步和第三步的评测和实验结果在下面两小节阐述。

4.1 评测新闻微博的选择

在 Twitter 中，转发的 tweet 称为 retweet。经验告诉我们，通常无论是普通网页还是博客，有关重要新闻话题的往往转载或转发概率较高。所以，如果本文的方法确实能将那些更有可能讨论新闻话题的微博客选取出来，那么在 Twitter 数据集上选取出来的 tweets 中，retweet 的比率应该高于平均值。实验证明了这一点，如表 2 所示。

表 2 在 3 月 8 日聚类结果中最大的一个类中得分最高的 5 条 tweets

Tab.2 Top scored 5 tweets in the biggest cluster on March 8

方法	Retweet 平均比例
本文选取 tweets 的方法	33.14%
不使用单义词单元	28.09%
全部 tweets 中	19.11%

可以看出，本文的方法确实能选出那些更有可能谈论新闻话题的微博，并且本文统计单义词单元的频率和变化率的方法比仅仅统计词是更有效的。

4.2 混合聚类的结果

首先，给出一个已知新闻事件的实例，表明本文的方法确实能有效发现该新闻话题。在我们的数据集收集期间，第 82 届奥斯卡颁奖典礼于 3 月 7 日晚举行了。对于这样一个重要的新闻事件，本文的方法确实能够非常有效地检测出该新闻话题。

因为本文的时间窗口是按天设置的。所以在 3 月 8 日的聚类结果中，最大的一个类就是谈论奥斯卡颁奖典礼这一新闻话题的。其中得分最高的 5 条 tweets 如表 3 所示

表 3 在 3 月 8 日聚类结果中最大的一个类中得分最高的 5 条 tweets

Tab.3 Top scored 5 tweets in the biggest cluster on March 8

作者	tweet 文本
SilkCharm	Trivia: Kathryn Bigelow won Best Movie/Best Director for Hurt Locker. Her ex Husband James Cameron, director of Avatar, did not #oscars
LAmovieexaminer	'Precious'star Mo'Nique wins Best Supporting Actress Oscar at the 82nd Academy Awards: http://bit.ly/bc5joW #Oscars
Cocacy	My predictions in the six big categories: Jeff Bridges, Meryl Streep, Mo'Nique, Christoph Waltz, Avatar; kathryn bigelow #theoscars
skynewsbreak	Kathryn Bigelow wins Best Director Oscar for The Hurt Locker - the first woman in history of the Academy Awards to win the accolade
prayoonko	RT @TwitBreakinNews: Christoph Waltz and Mo'Nique have taken the best supporting actor and actress Oscars at the 82nd Academy Awards in Hollywood.

最后列出聚类的一些实验结果。选取每天聚类得到的最大的一个类，并把这个类中得分最高的 tweet 作为该类的新闻话题代表列出。表 4 是其中若干天的结果，可以看出，这些 tweets 谈论的都是重要的新闻话题。

表 4 聚类结果中最大的类中得分最高的 tweet

Tab.4 The top scored tweet of the biggest cluster in some days

日期	作者	tweet 文本
Feb. 25	avivao	Dog fight. Lamar claims Obamacare will raise premium prices. Obama claims prices will go down 14-20%, per CBO. #HCSummit #HCS #HCR
Feb. 27	Infidel007	RT @BreakingNews: 54th major aftershock, magnitude 5.0, centered off coast of Bio-Bio, #Chile - U.S. Geological Survey
Mar. 03	NewTechBooks	FAA puts two air traffic control employees on admin leave after teen directs aircraft over JFK: CBS News has repor... http://bit.ly/b8pUwH
Mar. 07	vhernandezcnn	Major development RT @CNNworldgirl: Sr Pakistan officials tell CNN they have arrested Adam Gadahn, the American-born spokesman for al Qaeda.
Mar. 11	vivekmadan	Suicide blasts kill 45 in Pakistan's Lahore: Two suicide bombers targeting the Pakistani military killed at least... http://bit.ly/bbofZA
Mar. 15	troyjensen	Plane kills jogger in SC beach emergency landing; http://j.mp/dzTwWb - Damn, that is one series of terrible events!

5 总结与展望

本文在从大规模的微博客短文本数据集中，检测出新闻话题方面首先做了尝试。虽然我们的数据集从 Twitter 获得，但我们的方法同样可以应用到其他任何微博客系统中。本文中利用隐主题建模的方法，有效地解决了短文本集数据稀疏性的问题。选取最有可能讨论新闻话题的微博客的方法，将关注的数据对象规模大大缩小，并在一定程度上排除了非新闻博客的干扰。最后，使用一个两层的 K 均值和层次聚类的混合聚类方法，可以快速准确的将所有微博聚集到不同的新闻话题之下。但是对于本文的工作，仍有可以改进之处。

一方面，本文的方法并不是实时的。但可以通过缩短时间窗口，引入大规模外部数据集，在后台做隐主题挖掘的办法[13]，来实现一个真正实时的微博客新闻发现系统。

另一方面，微博由于自身长度的原因，很难全面的描述一个新闻事件，如何选择几篇微博将一个新闻事件完整的描述出来，也是将来工作的方向。

参 考 文 献

[1] 骆卫华，刘群，程学旗. 话题检测与跟踪技术的发展与研究[A]. 全国计算语言学联合学术会议(JSCL-2003)论文集[C] . 北京：清华大学出版社，2003，560-566.

[2] 骆卫华，于满泉，许洪波，王斌，程学旗. 基于多策略优化的分治多层聚类算法的话题发现研究[J]. 中文信息学报. 2006，20（1）:29-36

[3] 洪宇，张宇，刘挺，李生. 话题检测与跟踪的评测及研究综述[J]. 中文信息学报. 2007，21（6）：71-87

[4] J. Allan, J. Carbonell, G. Doddington, J. Yamron, et al. Topic detection and tracking pilot study final report[C]. In Proc of the DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[5] J. Allan, R. Papka, and V. Lavrenko. On-line new event detection and tracking[C]. In Proc of SIGIR'98，pp 37-45.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3:993-1022.

[7] D. Bollegala, Y. Matsuo, and M. Ishizuka. Measuring emantic similarity between words using web search ngines. In Proc of WWW'07, 2007, pp 757-66,.

[8] M. Connell, A. Feng, G. Kumaran, et al. Umass at tdt 2004[C]. In Proc of TDT 2004, 2004.

[9] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent sementic analysis[J]. Journal of the American Society of Information Science, 41(6).

[10] T. L. Gri_ths and M. Steyvers. Finding scientific topics[C]. In Proc of the National Academy of Sciences of the United States of America, 2004, pp 5228-5235.

[11] G. Heinrich. Parameter estimation for text analysis. Technical report, 2005.

[12] T. Hofmann. Probabilistic latent semantic analysis[C]. In Proc of UAI-99, 1999, pages 289-296.

[13] X.-H. Phan, L.-M. Nguyen, and S. Horiguchi. Learning to classify short and sparse text & web with hidden topics from large-scale data collections[C]. In Proc of WWW '08, 2008, pp 91-100.

[14] M. Sahami and T. D. Heilman. A web-based kernel function for measuring the similarity of short text snippets[C]. In Proc of WWW '06, 2006, pp 377-386.