基于 LDA 模型的研究领域热点及趋势分析

杨 星,李保利,金明举
(河南工业大学 信息科学与工程学院,河南 郑州 450001)

摘 要:随着研究的不断深入以及信息传播手段的进步,与某个研究领域相关的科学文献越来越多,也越来越容易得到,然而要阅读和分析这些数以千计的文献,仅凭人力已经难于实现对该领域研究重点、研究热点以及趋势进行全面系统地分析。鉴于此,提出一种基于 LDA 模型对某研究领域在一定时期内的热点及趋势进行自动识别的方法。该方法利用 Gibbs 抽样计算模型参数,获取领域热点主题以及热点词语,通过按时间后离散的主题演化方法分析热点主题在时间轴上的强度演化。以中文信息处理领域为例,通过对《中文信息学报》2001—2010 十年间发表的学术论文进行分析,自动获取中文信息处理领域十年内的研究热点以及热点主题在时间轴上的演化趋势。实验结果初步证明了该方法的有效性。

关键词:研究热点;LDA 模型;Gibbs 抽样;主题数目;主题演化
中图分类号:TP31 **文献标识码**:A **文章编号**:1673-629X(2012)10-0066-04

LDA-based Research Domain Hotspots and Trend Analysis

YANG Xing, LI Bao-li, JIN Ming-ju
(College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: Along with continuing in-depth research and the advancement of modern information dissemination technologies, more and more papers in a research domain are becoming available. Obviously, it's quite difficult for researchers to read and analyze the huge amounts of papers for thoroughly detecting the research hotspots and trend of a domain. Targeting at solving the above problem, a LDA-based approach is proposed to automatically recognize the hotspots and trend of a research domain. Gibbs sampling is used to calculate the LDA model parameters and determine the research hotspots as well as their representative words. The trend analysis is achieved by post discretizing research topics over time. In the experiments, Chinese information processing is chosen as the target research domain. The research hotspots and trend over the ten year period from 2001 to 2010 were obtained by automatically analyzing all the papers published on the journal of Chinese information processing during that period. Preliminary experiments demonstrate the effectiveness of the proposed approach.

Key words: research hotspots; LDA model; Gibbs sampling; topic number; topic evolution

0 引 言

随着信息时代的来临,网络已经成为人们获取信息的重要渠道,大量以文本格式存储的科学文献信息出现在图书馆和相关主题网站上。这些数据的主要特点是海量且繁杂,如何利用一种有效的方法分析这些文本数据,从中识别出重要的研究热点信息,并且进一步分析研究热点的发展趋势,成为急需解决的问题。

研究领域热点及趋势分析,就是对某一科学领域的文献数据进行综合全面的分析,挖掘出该领域的研究热点,并分析热点随时间的演化趋势。其目的是找出人们所关注的热点问题、热点技术及发展状况等重要信息。研究领域热点挖掘可以帮助人们及时了解领域内的热点研究问题、获得该领域的热点知识和发展趋势,便于研究者对自己将要或正在从事的研究领域有一个全面的理解,以帮助他们发现现有研究的不足并确定个人的研究方向。

目前,对热点主题的识别研究主要有以下三种方法:

(1)以词语切分和噪声库为基础,采用三级滤噪方法对网络热点信息进行拼接,最后依靠适当的收录策略提取出热点信息串。该方法会得到很多相似的冗余词语,且没有建立热点词语与主题间的映射[1]。

(2)聚类方法。周亚东等人[2]采用密度聚类方法(DBSCAN)将相关联的热点词语聚合为簇,再结合文

收稿日期:2012-02-29;**修回日期**:2012-06-02
基金项目:河南省基础与前沿技术研究项目(112300410007);河南省教育自然科学研究计划(2011A120002)
作者简介:杨 星(1986-),男,硕士研究生,CCF 会员,研究方向为语言信息处理;李保利,博士,教授,研究方向为语言信息处理、机器学习等。

章标题及网站 IP 等信息得到热点话题描述。该方法避免了热点词语的相互独立,使其与相应主题结合起来,但该方法容易将大量不重要的信息也当作热点词语识别出来,准确率较低。罗亚平等人[3]采用单遍聚类算法(Single Pass)建立热点话题发现模型,提出基于用户浏览行为的热点话题发现方法,该方法可以得到话题在一段时间内的演化趋势。但由于很多参数,如相关报道数目、话题数目等,需要人工设置或统计得到,所以缺乏自动计算相关参数的能力。李翔等人[4,5]将基于密度的聚类思想引入 K—Means 算法,克服了单纯采用 K—Means 算法依赖初始聚类数和初始聚类中心点的缺陷,同时回避了基于密度聚类算法速度慢的缺点,可以有效地发现互联网媒体信息热点。

(3)主题模型。Li 和 Yamanishi 等人[6]采用有限混合模型(Finite Mixture Model)实现主题分析,该方法不依靠任何统计假设来计算文本中的词汇分布,而是直接利用 EM(Expectation Maximization)对语料进行分析,这样容易导致出现局部最大值,并且收敛速度过慢。PLSA(Probabilistic Latent Semantic Analysis)模型[7]也被用在主题识别与分析,但该模型无法利用已有的文本集合模型对新的文本集合进行建模,即不能进行增量处理。此外,文本数目的增加也会引起隐含参数数量的变化,容易产生过度拟合。

主题模型(Topic Model)[8]起步于概率语义索引(PLSA),该模型是第一个完整意义的主题模型,其他主题模型都是在 PLSA 的基础上发展而来。主题模型的特点是将主题看成是词项的概率分布,文本又由主题随机混合而成。它具有基于聚类等方法不具备的识别大规模文本集中潜藏的主题信息的能力,利用词分布将文本信息转化为易于建模的数字信息,直观地表现主题,大大地简化了问题的复杂性,所以应用主题模型进行主题分析已成为热点主题识别的重要研究方法。

LDA(Latent Dirichlet Allocation)模型[9]是 Blei 等人在 PLSA 基础上改进得到的完全生成主题模型,由于其参数简单、数量不变,且不易产生过度拟合现象等优点,已经成为主题模型的研究热点之一。文中基于 LDA 模型和 Gibbs 抽样为文本建模,并计算模型参数,以后离散演化方法分析研究热点随时间的变化趋势。实验结果表明,该方法能正确获得研究领域的热点主题及在时间轴上的变化趋势。

1　LDA 模型

LDA 模型是一个三层贝叶斯具有文本主题表示能力的非监督产生式概率模型(一元模型)。三层是指:文本,主题,词汇。拓扑结构图如图 1 所示。

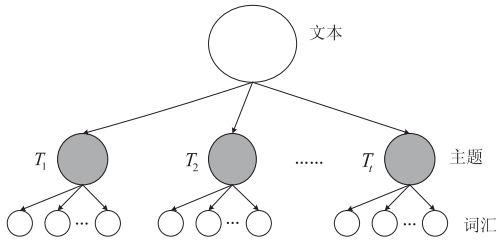

图 1　*LDA* 三层拓扑结构

若给定一个文本语料库 C,包括 M 个文本 $\{m_1, m_2 \cdots\}$,T 个主题,N 个唯一性词汇 $w_i, i \in (1,N)$。则,文本中的第 i 个单词 w_i 的概率可用下式表示:

$$p(w_i) = \sum_{j=1}^{T} p(w_i \mid z_i = j) p(z_i = j) \tag{1}$$

其中,z_i 是潜在变量,表示第 i 个单词 w_i 取自该主题,$p(w_i \mid z_i = j)$ 表示第 i 个单词 w_i 属于主题 j 的概率,$p(z_i = j)$ 表示主题 j 属于文本的概率。文本 m 中的词汇 w 的概率可以表示为:

$$p(w \mid m) = \sum_{j=1}^{T} p(w \mid z = j) p(z = j \mid m) \tag{2}$$

另外,LDA 模型假定文本由主题随机混合而成,记为:

$z_i \mid \boldsymbol{\theta}^{(m_i)} \sim \text{Discert}(\boldsymbol{\theta}^{(m_i)})$

主题又是词汇表中所有单词的随机组合,记为:

$w_i \mid z_i, \boldsymbol{\beta}^{(z_i)} \sim \text{Disceret}(\boldsymbol{\beta}^{(z_i)})$

其中,$\boldsymbol{\theta}$ 是一个列向量,它表示文本中每个主题发生的概率,不同的文本对应不同的 $\boldsymbol{\theta}$。$\boldsymbol{\beta}$ 是一个 $K \times V$ 的矩阵,行表示单词,列表示主题,矩阵单元表示某个主题生成某个单词的概率。

为了使 LDA 模型具备处理训练语料之外的新文本,方便模型参数的估计,需在参数 $\boldsymbol{\theta}$ 和 $\boldsymbol{\beta}$ 上做对称的 Dirichlet 先验假设,分布如下:

$\boldsymbol{\beta}^{(s_i)} \sim \text{Dirichlet}(\varphi), \boldsymbol{\theta}^{(m_i)} \sim \text{Dirichlet}(\alpha)$

这里假定上面的两个分布为对称的 Dirichlet 分布,即所有的 α 取相同的值,所有的 β 也取相同的值。LDA 模型图如图 2 所示:

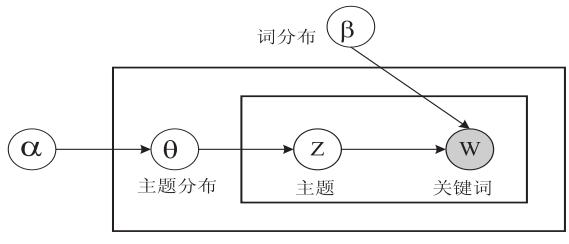

图 2　LDA 模型图

模型图解:图中只有关键词 w 是可观察变量,其他的变量都是隐藏变量。箭头方向表示条件概率方向,矩形表示可重复过程,大矩形表示从 Dirichlet 分布中为文本集中的每个文本反复抽取主题分布,小矩形表示从主题分布中反复抽取产生文本的词。

2 Gibbs 抽样

计算 LDA 模型参数的方法有多种,通常采用 EM 算法[9]、Gibbs 抽样[10]和 Expectation-Propagation[11]方法。由于 Gibbs 抽样速度快且易于实现,所以文中采用 Gibbs 抽样算法间接计算模型参数 $\boldsymbol{\theta}$ 和 $\boldsymbol{\beta}$。Gibbs 抽样是 MCMC(Markov Chain Monte Carlo)[10]的一种实现形式,其目的是从收敛于目标函数的马尔科夫链中抽取接近某概率分布的样本。对于文中的 LDA,仅需要对主题的词汇分布,也就是变量 z_i 进行抽样。记后验概率为:$p(z_i = j \mid z_{-i}, w_i)$,计算公式如下:

$$p(z_i = j \mid z_{-i}, w_i) = \frac{\frac{n_{-i,j}^{(w_i)} + \varphi}{n_{-i,j}^{(\cdot)} + W\varphi} \cdot \frac{n_{-i,j}^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}}{\sum_{j=1}^{T} \frac{n_{-i,j}^{(w_i)} + \varphi}{n_{-i,j}^{(\cdot)} + W\varphi} \cdot \frac{n_{-i,j}^{(d_i)} + \alpha}{n_{-i,\cdot}^{(d_i)} + T\alpha}} \tag{3}$$

其中,$z_i = j$ 表示将记号为 i 的单词分配给主题 j;z_{-i} 表示所有 z_k 的分配($k \neq i$);$n_{-i,j}^{(w_i)}$ 表示将属于 w_i 的词数分配给主题 j;$n_{-i,j}^{(\cdot)}$ 表示主题 j 需要分配的所有词数;$n_{-i,j}^{(d_i)}$ 表示主题 j 需要从记号为 i 的文本中分配的词数;$n_{-i,\cdot}^{(d_i)}$ 表示记号为 i 的文本中所有被分配了主题的词数。另外所有词数均应去掉 $z_i = j$ 的这次分配。

LDA 中 Gibbs 抽样过程:

1) 初始化。对于每一个主题 z_i 被初始化为 1 ~ T 之间的随机数(N 是单词总数,i 需从 1 循环到 N),此时得到初始马尔科夫链。

2) 迭代。i 从 1 循环到 N,按照后验概率公式(3)将单词分配给主题,此时可得到马尔科夫链的下一个状态。

3) 计算 $\boldsymbol{\beta}$ 和 $\boldsymbol{\theta}$ 的值。重复执行第二步到一定的次数,即马尔科夫链逐渐接近目标分布时,记录 z_i 的当前值。可按照公式(4)分别计算 $\boldsymbol{\beta}$ 和 $\boldsymbol{\theta}$ 的值,其中,$n_j^{(w)}$ 表示主题 j 得到某一单词 W 的频数;$n_j^{(\cdot)}$ 表示主题 j 得到的所有单词数;$n_j^{(d)}$ 表示主题 j 从某一文本 d 中得到的单词数;$n_{\cdot}^{(d)}$ 表示得到了主题的文本所包含的单词数。

$$\hat{\boldsymbol{\beta}}_w = \frac{n_j^{(w)} + \varphi}{n_j^{(\cdot)} + W}, \quad \hat{\boldsymbol{\theta}}_{z=j} = \frac{n_j^{(d)} + \alpha}{n_{\cdot}^{(d)} + T\alpha} \tag{4}$$

3 实验及分析

3.1 实验数据

本实验以中文信息处理为目标研究领域,通过分析 2001 ~ 2010 年《中文信息学报》上发表的 848 篇文献来获取该领域的研究热点及趋势。在实验中,使用每篇文章的标题和摘要而不是全文,这是因为:一方面,标题和摘要已经能够充分反映文章的主要内容;另一方面,受版权以及处理技术等限制全文并不总是能得到。

对于中文文本数据,首先需要进行分词处理,并过滤掉停用词等意义不大的词。使用中科院计算所汉语分词系统 ICTCLAS 来处理 848 篇文献的标题和摘要。经过处理后得到的有效词语有 7087 个。

3.2 模型参数求解

由 LDA 模型分析可知,该模型的可变量包括超参数 α,φ 以及主题数目 T。α 和 φ 的取值应与主题的数目和单词表的大小相关,由经验值可取 $\alpha = 50/t$(α 根据主题数目的变化而变化),$\varphi = 0.01$(这种取值在实验中有较好的表现)。

主题数目影响着 LDA 的性能,确定主题数目的方法有很多种,文中采用困惑度(Perplexity)[12]来对主题数目进行选取,困惑度为文本集中包含的各句子相似性几何均值的倒数,随句子相似性的增加而逐步递减。困惑度的计算公式为:

$$\text{Perplexity}(D) = \exp\left\{ -\frac{\sum_{m=1}^{M} \log p(w_m)}{\sum_{m=1}^{M} N_m} \right\} \tag{5}$$

式中:N_m 表示第 m 个文本的长度,M 为文本集,$p(w_m)$ 表示 LDA 产生文本的概率:

$$p(w_m) = \sum_{d} \prod_{n=1}^{N} \sum_{j=1}^{T} p(w_i \mid z_i = j)\, p(z_i = j \mid w_m)\, p(d) \tag{6}$$

从图 3 可以看出困惑度随着主题数目的增加而逐渐减小,当 $T = 90$ 时困惑度达到最小值。由于困惑度越低模型的泛化能力越强,性能越优,因此对文中的语料库来说当 $T = 90$ 时模型最优。

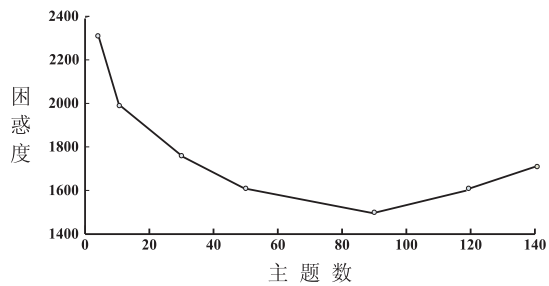

图 3 困惑度随主题数目的变化

3.3 实验结果分析

3.3.1 主题及关键词分析

求得模型参数后,得到主题在每个文本上的概率分布及单词在每个主题上的概率分布。此时需要统计每个主题在文本中出现的次数,通过设定合适的阈值来判定该主题是否为要寻找的领域热点。文中选取排在前 4 位的主题作为领域热点,并选择每个主题下的 10 个关键词作为热点词语,如表 1 所示。

表 1 基本说明了《中文信息学报》上所反映近十

年内中文信息处理领域内的研究热点。分析表 1 结果可得出以下结论:

表 1　主题与关键词

热点主题	与各主题相关的关键词
机器翻译	机器翻译 汉英 模型 统计 对齐 语料 消岐 HMM 语音 过滤
	翻译 性能 岐义 自动机 识别 语料 统计 中文 算法 匹配
分词	汉语 分词 信息 切分 中文 算法 词典 信息 自动 未登录
	匹配 标注 统计 分词 交集 岐义 切分 算法 中文 准确率
识别	汉语 算法 人名 汉字 编码 识别 语音 模型 精确率 自动
	短语 识别 耦合度 抽样 编码 说话人 图 自动 技术 口语
信息检索	信息 查询 检索 系统 编码 压缩 性能 模型 中文 算法
	检索 引擎 信息 语义 模型 搜索 查询 中文 排序 索引
	自动 信息 抽取 句法 语义 过滤 消岐 提取 岐义 模板

(1) 通过 LDA 产生主题的粒度对语料的描述非常准确和详细。一个主题可能包含很多的关键词序列,例如机器翻译这个主题,LDA 自动识别出两个与主题相关的关键词序列。对其他主题而言也是如此。

(2) 通过 LDA 可以识别种子词以及后续词。例如分词这个主题,可以识别出种子词(分词,切分)和后续词(歧义,标注等)。

(3) LDA 可以假定一篇文章包含多个主题。这种假设更加符合客观现实,因为一篇论文很多时候不是由一个主题构成,它可能包含很多主题词。

3.3.2　后离散主题演化分析

后离散分析是指先在文本集合上运用 LDA 模型获取主题,然后利用文本的时间信息检查主题在离散时间上的分布来衡量演化的强度[13]。进而按照文本的发布时间,将文档离散到相应的时间窗口。对于某个主题 Z_i,依次考虑它在每个时间窗口的强度[14]:

$$\delta_i^t = \frac{1}{D_t}\sum_{t_d \in t}\theta_{di} \tag{7}$$

其中,D_t 表示属于时间窗口 t 的文本数。

文中选择了"信息检索"、"识别"、"机器翻译"这三个热点主题,按文本的时间信息,采用按时间后离散的方法,时间粒度选择以年为单位,分析了主题随时间的演化,给出了不同的主题在时间上的分布走向(见图 4)。

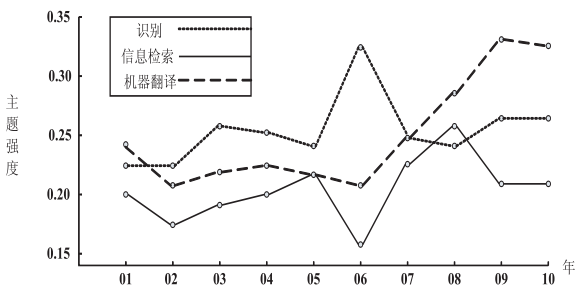

图 4　热点主题随时间的变化

从上图中可以看到三个热点主题在每一年的热度及其在十年间的变化趋势(由于语料规模有限,所以实验结果仅限于在该期刊所包含论文范围内进行分析,可能会与现实出现误差),其中这一主题在 06 年以后热度呈现逐年增长的趋势,且一直比信息检索机器翻译更受研究人员的关注;由于识别这一主题包含了很多相关的方向,所以一直处于一个较高的关注度水平,在 06 年的时候其关注度甚至比机器翻译高出约 75% 左右。图示热点走向与每一年所刊登的相关主题文章数量基本吻合,真实反映了中文信息处理领域的热点强度变化。

4　结束语

文中提出了一种基于 LDA 模型的研究领域热点及趋势分析方法:用 LDA 对文本建模,通过 Gibbs 抽样计算模型参数,自动获得研究领域内的热点词语,并将热点词语归类到相应的隐含主题,以后离散演化分析得到研究热点的变化趋势。结合中文信息处理研究领域,在 2001 ~ 2010 年《中文信息学报》数据集合上所做的实验初步证明了该方法的有效性。

尽管如此,在利用 LDA 模型对研究领域热点及趋势进行分析时仍存在一些问题:首先 LDA 模型主题数目需要事先确定,且在一定时间内是固定不变的,这样就无法在时间上观测新引入的研究热点和将要消亡的研究热点;其次,仅依靠各主题下的热点词语无法对研究热点进行很好的解释;最后,LDA 模型无法将各时间间隔内相关的研究热点进行关联,无法得到热点内容的演化。所以,如何动态确定主题数目、如何使得到的热点主题更加容易理解、如何分析研究领域热点内容的演化趋势,将是下一步研究的重点。

参考文献:

[1] 曾依灵,许洪波. 网络热点信息发现研究[J]. 通信学报,2007,28(12):141-146.

[2] 周亚东,孙钦东,管晓宏,等. 流量内容词语相关度的网络热点话题提取[J]. 西安交通大学学报,2007,41(10):1-5.

[3] 罗亚平,王　枞,周延泉. 基于关注度的热点话题发现模型[C]//中文计算技术与语言问题研究-中文信息处理国际会议. 北京:电子工业出版社,2007:402-408.

[4] 李若朋,李　翔,林　祥,等. 基于 DK 算法的互联网热点主动发现研究与实现[J]. 计算机技术与发展,2008,18(9):1-4.

[5] 黄宇栋,李　翔,林　祥. 互联网媒体信息热点主动发现技术研究与应用[J]. 计算机技术与发展,2009,19(5):1-4.

[6] Li H, Yamanishi K. Topic analysis using a finite mixture model [J]. Information Processing and Management, 2003, 39(4):

(下转第 74 页)

了很多,主要是应用层的移动节点可以从已经建立的协调层接口中快速、及时获得需要的资源节点路由信息,并迅速地查找到所需要的资源位置。TBRPF 路由协议也可以根据网络协调层,及时地获得应用层所提供的节点信息,这样就消除层间模块冗余信息,减少不必要相同的操作。

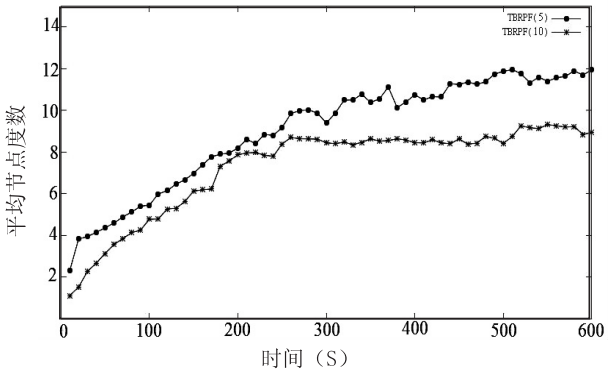

图 6 平均节点度数与时间关系图

图 6 是改进后的移动 Ad-hoc 网络下的平均连接度数与时间的关系图,可以看出,使用跨层技术后的 Gnutella 网络中的节点很快就达到最低连接数目的要求,与邻居节点建立连接的速度明显比改进之前的网络要快,而且在 200 秒之后就开始达到最大连接数目,有利于网络拓扑快速恢复,迅速重建路由,实现资源更快的查找。

5 结束语

文中首先阐述了目前移动 Ad-hoc 网络的基本概念,并简述了应用层的 Gnutella 协议与网络层的 TBRPF 协议。论述了跨层机制的种类及特点。然后分别针对 Gnutella 协议下采用 AODV 和 TBRPF 路由机制的移动 Ad-hoc 网络进行试验模拟,对其中的问题进行相关的描述。最后提出了基于 Gnutella 协议的无线Ad-hoc网络跨层优化模型,并对该模型进行模拟仿真。

仿真实验表明,该模型能够更好地容忍 Ad-hoc 的动态性特点,更好地减少移动网络的总体开销,并且提高资源查询效率。

参考文献:

[1] Radunovic B. A Cross-layer Design of Wireless Ad-Hoc Networks[J]. IEEE Computer,2005,6(12):78-84.

[2] 朱梅丽,李万磊,谢 波,等. Ad Hoc 网络协议栈跨层自适应设计[J]. 计算机系统应用,2010,19(11):45-49.

[3] Conti M,Gregori E. A cross-layer optimization of Gnutella for mobile ad hoc networks[J]. IEEE Computer,2004,18(2):114-119.

[4] Ogier R,Templin F. Topology Dissemination Based on Reverse Path Forwarding[R]. [s. l.]:SRI International,2004.

[5] Ogier R. A Simulation Comparison of TBRPF, OLSR and AODV[C]//SRI International. [s. l.]:[s. n.],2002.

[6] Bouckaert S. Cross-layer Architecture and Optimizations in Hybrid Wireless Mesh Networks [R]. [s. l.]:Sch. of Electron. & Commun. Eng. Commun,2009.

[7] 商西达,王启国,杨 平. 移动自组网跨层设计:方法与难点[J]. 舰船电子工程,2010,30(8):110-114.

[8] Raisinghani T,Iyer S. ECLAIR:Efficient cross layer architecture for wireless protocol stacks[M]. [s. l.]:TataInfotech Ltd,2004:370-389.

[9] Raisinghani T,Iyer S. Cross-layer design optimization in wireless protocol stacks[J]. Computer Communications,2004,27(8):720-724.

[10] 李春秀,刘方爱. 一种有效的非结构化 P2P 网络资源搜索策略[J]. 计算机技术与发展,2010,20(11):117-121.

[11] Conan V,Levy M. Cross-layer Interface for Wireless Ad hoc Networks[C]//Thales Communications. [s. l.]:[s. n.],2006.

[12] 徐伟强,汪亚明,俞成海. 移动 Ad Hoc 网络的跨层优化拥塞控制[J]. 软件学报,2010,21(7):16-21.

(上接第 69 页)

521-541.

[7] Hofmann T. Probabilistic latent semantic analysis[C]//Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. Stockholm,Sweden:Morgan Kaufmann,1999:289-296.

[8] Blei D M,Lafferty J. Topic models[M]//Text Mining:Theory and Applications. London,UK:Taylor and Francis,2009.

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

[10] Steyvers M, Griffiths T. Probabilistic topic models[M]// Handbook of Latent Semantic Analysis. New Jersey:Springer, 2007.

[11] Minka T,Lafferty J. Expectation-propagation for the Generative Aspect Model[C]//Proc of the 18th Conf on Uncertainty in Artificial Intelligence(UAI). [s. l.]:[s. n.],2002.

[12] Cao Juan,Xia Tian,Li Jintao,et al. A density based method for adaptive LDA model selection[J]. Neurocomputing,2009,72(7-9):1775-1781.

[13] 单 斌,李 芳. 基于 LDA 话题演化研究方法综述[J]. 中文信息学报,2010,24(6):43-49.

[14] Griffiths T L,Steyvers M. Finding scientific topics[C]//Proceeding of the National Academy of Science of United States of America. [s. l.]:[s. n.],2004:5228-5235.

基于LDA模型的研究领域热点及趋势分析

作者： 杨星， 李保利， 金明举
作者单位： 河南工业大学 信息科学与工程学院,河南 郑州 450001
刊名： 计算机技术与发展
英文刊名： Computer Technology and Development
年，卷(期)： 2012(10)