

基于隐含狄利克雷模型的文献主题演化预测

茅利锋, 张 伟

(南京邮电大学 计算机学院, 江苏 南京 210023)

摘 要:利用隐含狄利克雷分配模型(LDA),根据科技文献往年的主题变化来分析科技文献主题的演化,是目前主题演化研究的热点。根据科技论文的主题演化具有无后效性的特点,使用马尔可夫链来预测主题的演化信息。该方法利用 LDA 模型获取不同时段的主题,使用相似度等方法对相邻时间窗口的主题进行关联,并根据主题的程度将主题分为热门主题、普通主题和冷门主题,最后利用马尔可夫链得到主题之间的强度转移概率矩阵,对主题的程度变化趋势进行分析和预测。对 NIPS 论文集进行实验表明,科技论文主题在长时间演化后,其状态占比趋于稳定,热门主题、普通主题和冷门主题占比将保持在 30%、60% 和 10% 左右。说明该方法能有效地根据现有的主题演化结果对主题在未来几年的演化信息进行预测。

关键词:隐含狄利克雷分配模型;主题演化预测;马尔可夫链;状态转移

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2016)09-0034-05

doi:10.3969/j.issn.1673-629X.2016.09.008

Topic Evolution and Prediction of Scientific Papers Based on Latent Dirichlet Allocation Model

MAO Li-feng, ZHANG Wei

(School of Computer Science & Technology, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China)

Abstract:According to the change of the topic of scientific papers in previous years, to analyze the evolution of scientific papers based on Latent Dirichlet Allocation (LDA) is the current research focus. Through the aftereffect for topic evolution of scientific paper, Markov Chain is used to predict the evolution information of topic. In this method, LDA is used first to obtain the topics in different time windows, then some calculation method like similarity is used to associate with topics in neighboring time window. According to the intensity of topics, these topics are divided into 3 states including popular, normal and cold. Finally, the state transition matrix which is gained by the Markov Chain is used to analyze and forecast the trend of topic evolution. The experiment on proceedings of NIPS shows that after a long period evolution, the state proportion of topics of scientific papers is stabilized, with hot 30%, normal 60% and cold 10% remained, which shows that this method can effectively predict the trend of topic evolution in the next few years according to the existing evolutionary information.

Key words:Latent Dirichlet Allocation; topic evolution and prediction; Markov Chain; state transition

0 引言

随着 Web2.0 的快速发展,如何从承载着科学研究各个领域最新成果的科技文献中挖掘出自己感兴趣的课题并研究其发展方向,成为了困扰科研工作者的一个难题。科技文献的主题演化主要研究一定时期内期刊主题的变化程度以及不同主题随时间变化的规律,这有助于揭示各领域在不同时期内的变化规律以

及重点研究对象,为科研工作者提供决策依据。

主题的发展和演化工作最早源于主题的探测和追踪 (Topic Detection and Tracking, TDT) 研究^[1]。TDT 主要是基于统计知识,对文本进行信息过滤,然后利用分类策略跟踪相关主题,但它不能反映不同文本间词汇的语义信息。为了解决这个问题,主题模型的方法被提出。隐含狄利克雷分配 (Latent Dirichlet Alloca-

收稿日期:2015-12-12

修回日期:2016-04-05

网络出版时间:2016-08-01

基金项目:国家自然科学基金资助项目(61272422)

作者简介:茅利锋(1991-),男,硕士,研究方向为社会计算;张 伟,博士,教授,研究方向为计算机病毒技术、网络数据行为分析、Web 应用与数据安全等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160801.0907.056.html>

tion, LDA) 模型^[2], 作为主题模型的一种, 因为其可以很好地模拟大规模语料的语义信息, 越来越受到研究者的重视。

LDA 模型在主题发现方面优势明显, 但是普通的 LDA 模型很难发现主题的演变过程^[3-6], 因此如何将 LDA 模型应用到主题演化研究中成为了科研工作者研究的重点。文中认为科技论文的主题演化具有无后效性的特点, 即一个主题将来处于什么状态、取什么值, 仅与它现在的状态相关, 与它过去处于什么状态、取什么值无关。马尔可夫链作为马尔可夫过程的一种特殊情况, 它表明事物的状态由过去转变到现在, 由现在转变到将来的过程, 同样具有无后效性^[7]。因此, 根据马尔可夫链的特性, 可以将它应用到科技文献的主题演化研究中, 预测主题的状态演变规律。

1 相关工作

Deerwester 等首先提出了潜在语义索引 (Latent Semantic Indexing, LSI)^[8], 但是该方法缺乏数学理论支撑。随后, Hofmann 等^[9]提出了概率潜在语义索引 (Probabilistic Latent Semantic Indexing, PLSI), 该模型能对文本及其隐含主题构建模型, 在性能上优于 LSI 模型, 但易导致过拟合问题。为了克服这个问题, Blei 等在 PLSI 的基础上, 引入了超参数, 提出了隐含狄利克雷分配模型 (Latent Dirichlet Allocation, LDA)。该模型克服了 PLSI 在理论上的缺陷, 并且继承了 PLSI 的降维优势, 提出后得到了广泛应用。

基于静态的主题模型方法很难发现主题的演变过程及趋势变化, 因此基于主题模型的演化模型被提出。Wang 等^[10]将时间作为主题模型的内在变量, 提出了 TOT (Topic Over Time) 模型。该模型可以表示主题在不同时刻的分布强度, 能很好地揭示主题强度的变化趋势。Xu 等^[11]又在此基础上, 加入了作者 (Author) 这一因素, 提出了 ATOT 模型。不同于 TOT 模型, Griffiths 等^[12]提出了先获取主题再离散到时间窗的方法。该方法首先用 LDA 模型获取文本集合所有的主题, 然后再利用文本的时间信息来衡量演化, 在科技会议论文测试中取得了较好的结果。以上两种方法都要求将文本集作为一个整体来处理, 这会导致不同时间窗主题个数固定以及在线处理功能的缺失。因此, 目前科研工作者主要集中于先离散到时间窗再获取主题的方法, 为以后每个时间窗内的分析提供充分的灵活性。Blei 等^[13]提出的 DTM (Dynamic Topic Model) 就是其中的代表。该模型每个时间窗口的文本都是由 LDA 模型生成, 然后使用 KL 距离 (Kullback-Leibler divergence) 计算不同时间窗内主题分布的相似程度。在 DTM 模型的基础上, Ahmed 等^[14]提出了 iDTM 模

型, 引入了 HDP 方法, 解决了各时间窗口内主题数固定的问题。Wei 等^[15]和 AlSumait 等^[16]分别提出了 DMM (Dynamic Mixture Model) 和 OLDA (Online LDA), 分别实现了连续时间模型和离散时间模型的在线化。Wang 等^[17]提出了 OHDP (Online HDP) 模型, 解决了 HDP 方法难以处理大规模以及实时性要求高的文本流的难题。

以上方法都集中于如何使用或改进 LDA 模型以实现对文本数据的动态挖掘, 没有对其演化规律的预测进行进一步的研究。文中根据科技论文的演化具有无后效性的特点, 引入马尔可夫链来对主题的状态进行演化预测。该方法采用先离散到时间窗再获取主题的方法, 然后依据主题的状态对主题进行状态划分, 根据马尔可夫链计算主题强度的状态转移矩阵, 凭借该矩阵可以总结出主题强度的演变规律并对未来几年的主题强度演化进行预测。

2 基础理论

2.1 LDA 模型

LDA 模型^[2]是一个三层变参数层次贝叶斯模型, 它假设每个文档都是由多个主题混合而成, 而每个主题都是由多个词的概率分布组成。通过这种隐变量模型, 可以获得观测数据后面隐藏的主题分布, 通过后验概率推断来获得这个隐藏的结构。其具体生成过程如下:

(1) 对于每个文档 $d \in D$, 根据 $\theta_d \sim \text{Dir}(\alpha)$, 得到文档 d 上主题的多项式分布参数 θ_d 。

(2) 对于每个主题 $z \in K$, 根据 $\varphi_z \sim \text{Dir}(\beta)$, 得到主题 z 上词汇的多项式分布参数 φ_z 。

(3) 对于文档 d 中的词汇 $w_{d,j}$, 根据多项分布 $z_{d,j} \sim \text{Mult}(\varphi_z)$, 得到主题 $z_{d,j}$, 根据多项分布 $w_{d,j} \sim \text{Mult}(\theta_z)$ 得到词汇 $w_{d,j}$ 。

为了得到 LDA 模型, 首先要对 θ 和 φ 进行推理。由于精确的推理很困难, 一般都采用近似的推理方法, 目前主要使用的有 Variational Inference 和 Gibbs Sampling 两种方法。Gibbs Sampling 因为实现简单, 被广泛采用。文中也采用 Gibbs Sampling 方法。

2.2 马尔可夫链模型

马尔可夫分析法^[5]主要分析随机事件未来发展变化的趋势, 即利用某一变量现在的状态去预测其未来的状态, 以预测某未来时刻可能产生的变化, 以便采用相应的对策。

马尔可夫过程是具有无后效性的随机过程。无后效性是指当过程在 t_m 时刻所处的状态为已知时, 过程在大于 t_m 的时刻所处的状态的概率特性与过程在 t_m 时刻所处的状态有关, 而与过程在 t_m 时刻以前的状态

无关。通常把时间和状态都离散的马尔可夫过程称为马尔可夫链。

马尔可夫链中需要使用转移概率矩阵,转移概率矩阵的定义如下:设系统的状态有 n 个,系统在 t_m 时刻处于状态 i ,在下一时间 t_{m+1} 转为状态 j 的概率为 p_{ij} ,则称 p_{ij} 为一步转移概率。将这些 p_{ij} 依序排列,即构成一步转移概率矩阵 \mathbf{P} 。若系统经过 k 次转移后在时刻 t_{m+k} 处于状态 j 的概率为 p_{ij} ,则 $p_{ij}(k)$ 称为 k 步转移概率。相应的构成 k 步转移概率矩阵 $\mathbf{P}^{(k)}$ 。

3 基于 LDA 模型的科技文献主题演化研究

3.1 算法体系结构

对于期刊主题而言,演化主要由三个方面组成。首先是主题在时间维度上的强度变化,也叫做主题的活跃度变化;然后是主题随时间在内容上的变化,即主题的内容迁移;最后是主题随时间的社会网络变化,比如作者的变化。文中主要讨论第一种变化,即主题的活跃度变化。

整体流程框架如图 1 所示。

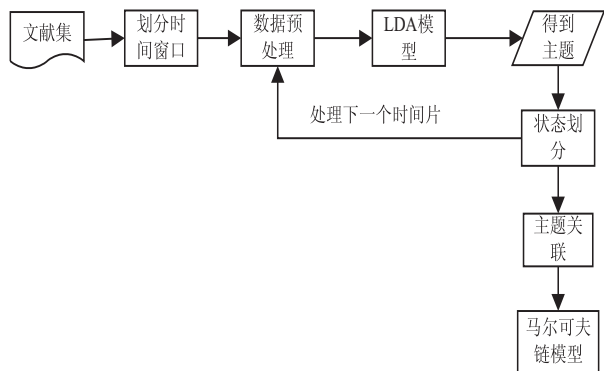


图 1 整体流程框架

首先,针对不同时间窗口的文献集,经 LDA 模型得到每个时间窗口的主题;然后,计算各个时间窗口的主题强度和相邻时间窗口的主题相似度,根据强度大小对主题状态进行划分,根据相似度对相邻时间窗口的主题进行关联;最后,根据关联信息统计计算得到状态转移概率矩阵,凭借状态转移概率矩阵对主题演化信息进行预测。

3.2 主题识别及关联

在传统的 TDT 中,主题定义为由一个种子事件或活动以及与其直接相关的事件或活动的集合,而在 LDA 中,主题表示为分布在一组主题词上的向量,即

$$\mathbf{T} = \{(v_1, p_1), (v_2, p_2), \dots, (v_n, p_n)\} \quad (1)$$

其中, v_i 是与主题 \mathbf{T} 相关的词; p_i 是主题 \mathbf{T} 在该词上的概率分布。

同样的,文献则是这些主题上的多项分布,其表示如下: 万方数据

$$d = \{(T_1^*, p(T_1^* | d)), (T_2^*, p(T_2^* | d)), \dots, (T_k^*, p(T_k^* | d))\} \quad (2)$$

其中, $p(T_i^* | d)$ 为主题 T_i^* 在文献 d 中的概率。

经 LDA 识别的主题在不同时间窗口是互相独立的,为了分析主题的演化,需要对不同时间窗口的主题进行关联。研究者认为主题的演化过程中必然存在内容的延续与改变,相邻时间窗口相关联的主题必然存在着内容的相似性,因此文中使用基于相似度的主题关联方法,通过计算相邻时间窗口的主题相似度来衡量主题内容的延续性,并建立主题关联。因为不同时间窗口 LDA 的词汇分布不相同,所以首先要把相邻时间窗口的两个词汇表扩充为一个词汇表。为了方便起见,对于原本不属于当前时间窗口的词汇,其概率取为 0。

计算主题之间的相似度的方法常使用基于 KL 散度的方法和余弦相似度方法。文中使用余弦相似度方法。主题 T_i^* 与主题 T_j^* 之间的余弦相似度计算公式如下:

$$\text{Sim}(T_i^*, T_j^*) = \frac{\sum_{x=1}^n p_x^i \times p_x^j}{\|T_i^*\| \times \|T_j^*\|} \quad (3)$$

其中, $\|T_i^*\|$ 为 2 范数; n 为当前时间窗内词表 V 中词的总数。

在关联规则的选取中,采用最简单的关联规则,即认为当前时间窗口中的每个主题与下一时间窗口中相似度最高的主题之间存在关联。

3.3 主题强度度量及状态划分

主题强度主要描述了主题热门程度,在某一时刻关于某个主题的文章数量越多,说明该主题的主题强度越高。对时间窗口 t 内的主题 T_i^* ,其主题强度 $S(T_i^*)$ 计算公式如下:

$$S(T_i^*) = \frac{\sum_{j=1}^{D_t} p(T_i^* | d_j)}{D_t} \quad (4)$$

其中, D_t 为时间窗口 t 内文档的数目。

计算每个主题的主题强度后,根据强度的大小把主题的状态划分为热门主题、普通主题和冷门主题,其具体的划分方法如下所示:

(1) 首先计算主题的平均强度:

$$\bar{S}_t = \frac{1}{T}$$

其中, T 为时间窗口 t 内的主题数。

(2) 对数据进行分析,给出一个阈值 ε_t 。

(3) 按照以下规则对主题进行划分:如果 $S(T_i^*) \geq \bar{S}_t + \varepsilon_t$,则认为主题 T_i^* 为热门主题;如果 $S(T_i^*) \leq \bar{S}_t - \varepsilon_t$,则认为主题 T_i^* 为冷门主题;否则,认为主题 T_i^* 为普

通主题。

3.4 转移概率矩阵的建立及状态预测

在计算状态转移概率前,首先要对随机序列进行无后效性检验,若序列不具备无后效性,则不能当作马尔可夫链来对待。常用 χ^2 统计量对随机序列进行马氏性检验。

由于数据的随机性,准确地确定转移矩阵是马尔可夫过程求解中的难题,文中使用统计估算法,其方法如下:

首先对一组或者几组所需要研究的离散数据,按照3.3节提出的方法将状态分为3种,结合马尔可夫链的概念对这些离散数据进行统计,将从*i*状态转移到*j*状态出现的次数统计到矩阵 $M(i,j)$ 中,最后根据式(5)计算得到马尔可夫链转移矩阵。

$$p(i,j)=\frac{M(i,j)}{\sum_{n=1}^3M(i,n)} \tag{5}$$

在确定了马尔可夫状态转移矩阵后,就可以对未来几个时间窗口的状态进行预测,具体方法如下:

(1)确定初始状态概率向量。选取一个初始时间窗口,以该时间窗口各主题状态所占的比例为分向量,建立初始状态概率向量 $S^{(0)}$ 。

$$S^{(0)}=(s_1^{(0)},s_2^{(0)},s_3^{(0)}) \tag{6}$$

式中, $s_1^{(0)}$ 、 $s_2^{(0)}$ 、 $s_3^{(0)}$ 分别表示热门、普通、冷门三种状态在主题中的比重。

(2)利用马尔可夫链预测*k*年后的状态,公式为:

$$S^{(k)}=S^{(0)}\times p^k \tag{7}$$

4 实验结果与讨论

文中选取了英文数据集 NIPS 作为研究对象。NIPS 会议是神经计算和机器学习领域最好的会议之一,该数据集包含 1988~2001 年共 14 年会议 1 879 篇论文的全文数据,文中仅摘取了标题和摘要作为实验的研究对象,并去除了停用词和数词。将数据集按照年份分为 14 个时间窗口,其中前 10 个时间窗口用来进行主题发现和演化实验,后 4 个时间窗口用来进行预测分析。

4.1 文献主题挖掘及强度演化结果

对于文档的建模过程,使用 Gibbs Sampling 进行参数推演,迭代 1 000 次,模型参数 α 和 β 分别设置为 50/*K* 和 0.01, *K* 取值 20。实验使用了开源的 Gibbs Sampling 工具 GibbsLDA++。文中列举时间窗口 1 强度最高的 5 个主题包含的主题信息(主题序号由 LDA 算法产生),如表 1 所示,其强度从上往下依次降低。

从表 1 中可知,1988 年 NIPS 会议的重点集中于网络、模型与仿真数据化、模拟电路性能以及信息采集等

方面。

表 1 时间窗口 1 强度最高的 5 个主题

主题序号	主题内容
Topic 7th	network, result, rule, develop, theory, task, condition, communication, cortical
Topic 2th	model, simulation, behavior, adaptive, include, computer, structure, actual, length
Topic 5th	code, number, type, class, space, general, field, visual, found, convergence,
Topic 3th	network, analog, design, property, large, theory, serial, event, analyze, recurrent,
Topic 13th	information, process, map, state, fire, brain, classification, unit, epsp, potential,

根据 3.2 节的方法,计算主题之间的相似度,取相似度最高的主题作为关联主题,以表 1 所列的五个主题为例,分析这五个主题在时间窗口 1~10 的主题强度变化,如图 2 所示。

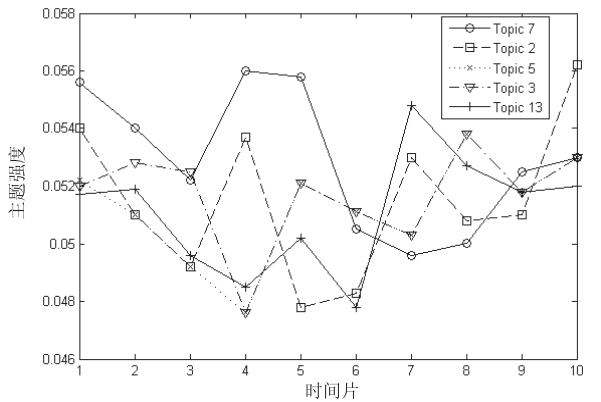


图 2 时间窗口 1 最热门主题随时间的强度变化

观察图 2,发现主题 3 和主题 5 的曲线在时间窗口 4 的曲线发生了合并,分析这两个主题在合并前的主题内容,发现主题 3 在时间窗口 1、2 和 3 中均为关于模拟电路的内容,而在时间窗口 4 其内容转移到了关于代码优化等方面的内容,而主题 5 内容随时间变化区别不大,因此分析主题 3 随着时间的进行而逐渐消亡的可能性较大。主题 7 在时间窗口 3 和时间窗口 5 主题强度变化程度较大,分析该主题在时间窗口 3、4、5 和 6 的主题内容,发现时间窗口 4 的主题内容相较时间窗口 3 关于硬件方面的词的比例显著上升,而时间窗口 6 的主题内容相较时间窗口 5 无显著变化。主题 2 在时间窗口 3、4 和 5 经历了强度显著下降又明显上升的过程,同样分析其在这三个时间窗口的主题内容,发现在时间窗口 3 和时间窗口 4 主题内容无明显变化,而在时间窗口 5 主题内容关于图像方面的内容比例显著上升。同理,观察其他主题,发现当主题强度显著上升时,其主题内容均发生了比较明显的变化,而主题强度明显下降时,主题内容变化不大。说明当该主题产生了新的技术或者有新的发现时,其主题强度

会在短时间提升,而当主题在长时间一成不变的时候,其主题强度自然而然会下降。这说明一个主题要想长时间保持热度,必须要随时保持创新。

4.2 马尔可夫链预测

$\bar{S} = 0.05$,取 $\varepsilon_i = 0.002$ 。那么主题的状态划分如下:如果 $S(T_i^t) \geq 0.052$,那么该主题为热门主题;如果 $S(T_i^t) \leq 0.048$,那么该主题为冷门主题;否则,该主题为普通主题。以时间窗口 1~20 个主题为基准,分析这 20 个主题随时间的状态变化规律。其在时间窗口 1~10 的状态变化如表 2 所示,其中 1 表示热门,2 表示普通,3 表示冷门。

表 2 时间窗口 1 中主题的状态变化

主题序号	状态变化(1988-1997)
Topic 0	3->2->2->1->2->2->2->2->2->2
Topic 1	2->2->2->2->2->2->1->1->1->2->2->1
Topic 2	1->2->2->1->3->2->1->2->2->2->1
Topic 3	1->1->1->3->2->2->2->2->1->2->2->1
Topic 4	2->2->1->2->1->1->1->2->2->2->2->2
Topic 5	1->2->2->3->2->2->2->2->1->2->2->1
Topic 6	2->3->2->3->2->2->2->2->1->2->2->1
Topic 7	1->1->1->1->1->1->2->2->2->2->1->1->1
Topic 8	2->2->3->3->3->3->2->3->3->1->2->2->1
Topic 9	3->2->1->2->2->2->1->1->1->2->2->2->1
Topic 10	2->3->2->2->2->2->2->1->1->1->2->2->2->2
Topic 11	2->2->1->2->2->3->1->1->2->2->2->2->3
Topic 12	2->2->2->2->2->2->2->1->1->1->2->2->2->1
Topic 13	2->2->2->2->2->2->2->3->1->1->1->2->2->1
Topic 14	2->2->2->2->1->3->3->2->1->1->2->2->2->1
Topic 15	2->2->2->2->3->2->2->2->2->2->1->2->2->1
Topic 16	3->2->2->2->2->2->2->3->1->1->1->2->2->1
Topic 17	2->2->2->2->2->2->2->2->1->1->1->2->2->2->2
Topic 18	2->2->2->2->2->2->2->2->1->1->1->2->2->2->1
Topic 19	3->2->2->1->2->2->3->3->1->1->2->2->2->2->3

根据表 2,计算一步转移概率矩阵 P ,得到:

$$P = \begin{bmatrix} 0.30 & 0.57 & 0.13 \\ 0.29 & 0.60 & 0.11 \\ 0.21 & 0.68 & 0.11 \end{bmatrix}$$

根据 P 对时间窗口 11~14 的主题状态进行预测,结果如表 3 所示。

由表 3 可知,平均误差在 16.6% 左右,因主题总数只有 20,误差在可接受范围之内。在主题窗口 11~14,热门主题、普通主题和冷门主题的数量变化不大,说明时间窗口 1 的主题随着长时间的演变后,其状态占比趋于稳定,热门主题、普通主题和冷门主题占比将

保持在 30%、60% 和 10% 左右。

表 3 1998-2001 年主题状态预测

时间窗口	预测情况			实际情况			相对误差
	热门	普通	冷门	热门	普通	冷门	
11	5.78	11.74	2.48	5	13	2	16.5
12	5.65	12.02	2.33	4	14	2	24.0
13	5.68	12.01	2.31	6	11	3	12.8
14	5.67	12.01	2.32	5	11	4	13.2

5 结束语

文中选用 LDA 模型,采用马尔可夫链预测方法对科技论文主题的演化进行了探索,并以 NIPS 会议的论文集为例对该方法的有效性进行了验证。实验结果表明,该方法在主题强度的演化中有一定的准确度,能够有效地反映出主题的热度演变规律。

但是仍存在以下不足:首先,设置每个时间窗口内主题的数量相同,这与实际并不相符;其次,关联规则设置过于简单,不能有效地显示出主题演化过程中的主题消亡、合并、分裂等现象;最后,马尔可夫链状态转移计算方法需要进一步改进。

因此,下一步研究中,将针对以上几点不足,对基于 LDA 模型的科技论文演化方法进行演化和改进。

参考文献:

[1] Allan J, Carbonell J, Doddington G, et al. Topic detection and tracking pilot study final report [R]. Virginia: Lansdowne, 1998.

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

[3] 赵迎光,洪娜,安新颖. 主题模型在主题演化方法中的应用研究进展[J]. 现代图书情报技术, 2014, 30(10): 63-69.

[4] 楚克明,李芳. 基于 LDA 模型的新闻话题的演化[J]. 计算机应用与软件, 2011, 28(4): 4-7.

[5] 赵旭剑,杨春明,李波,等. 一种基于特征演变的新闻话题演化挖掘方法[J]. 计算机学报, 2014, 37(4): 819-832.

[6] 胡艳丽,白亮,张维明. 一种话题演化建模与分析方法[J]. 自动化学报, 2012, 38(10): 1690-1697.

[7] Karlin S. A first course in stochastic processes[M]. [s. l.]: Academic Press, 2014.

[8] Deerwester S, Dumais S, Landauer T, et al. Indexing by latent semantic analysis[J]. Journal of the American Society of Information Science, 1990, 41(6): 391-407.

[9] Hofmann T. Probabilistic Latent semantic indexing[C]//Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval.

后,通过 Dispatcher 来更新 UI,WPF 会根据 Dispatcher 的优先顺序来更新 UI。

4 仿真结果显示及分析

在 WPF 的理论基础上,实现了基于 WPF 的 3D 仿真监测系统。WPF 能够很好地实现三维仿真。基于 WPF 的 3D 仿真监测系统的分析有以下几点:

(1)WPF 仿真计算的模拟和图形显示分离的策略便于开发人员对项目的开发,提高了系统的交互性。

(2)结果的显示在实体应变上以不同颜色呈现,形象直观,画面清晰。

(3)模型显示操作支持多种视图效果,可以全方位地预览和设计模型,同时支持上、下、左、右、前、后等运动操作,有利于实体的改进与优化。

(4)对于同样的 3D 模型不需要进行重新构建和导入,可以把之前导入的模型添加到模板,从而实现模型复用,有利于快速开发。

5 结束语

文中着重论述了在 Windows 环境下基于 WPF 的 3D 模型仿真的研究与实现,介绍了通过 XAML 技术实现界面与逻辑完全分离,可以使用少量的代码实现,开发过程简单,分析了构建 3D 仿真监测系统中涉及到的关键技术。该方法已应用于三维仿真监测系统中。结果表明,基于 WPF 的三维仿真系统的研究与实现具有很好的效果。

参考文献:

- [1] 曲 啸. 基于 OpenGL 的空间仿真可视化技术研究[D]. 西安:西安电子科技大学,2013.
- [2] 彭晓军,李 焱,贺汉根. 3Ds 模型在虚拟现实几何建模中的应用[J]. 计算机仿真,2003,20(3):52-55.
- [3] 郭成操. 图形仿真加工系统的研究与开发[D]. 成都:四川大学,2003.
- [4] MacVittie L A. XAML in a Nutshell[M]. [s. l.]:O'Reilly Media,2006.
- [5] 张景仁. WPF 模拟三维消防场景的技术实现[J]. 武警学院学报,2009(6):87-89.
- [6] Anderson C. WPF 核心技术[M]. 朱永光,译. 北京:人民邮电出版社,2009.
- [7] 李 响. 葵花宝典:WPF 自学手册[M]. 北京:电子工业出版社,2010.
- [8] 徐 宁,杜维刚. 对 Windows Vista 中新一代标记语言 XAML 的研究[J]. 电脑知识与技术,2007,1(3):774-775.
- [9] Evjen B, Hanselman S, Rader D. ASP. NET 2.0 高级编程特别版[M]. 北京:清华大学出版社,2007.
- [10] Sells C, Griffiths I. Programming WPF[M]. [s. l.]:O'Reilly Media,2007.
- [11] Eisenberg R, Bennage C. Teach yourself WPF in 24 hours[M]. American:SASM,2008.
- [12] MacDonald M. Pro WPF in C# 2010: windows presentation foundation in .NET 4[M]. [s. l.]:Apress,2014.
- [13] Andrade C, Livermore S, Meyers M, et al. Professional WPF programming:NET developmentwith the windows presentation foundation[M]. [s. l.]:Wiley Publishing,2007.
- [14] 钟红梅,简兴祥,彭玛丽,等. 三维可视化物探数据分析系统研究[J]. 硅谷,2014(22):41-42.

(上接第 38 页)

- New York:ACM,1999:50-57.
- [10] Wang X, McCallum A. Topics over time:a non-Markov continuous-time model of topical trends[C]//Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. [s. l.]:ACM,2006:424-433.
 - [11] Xu S, Shi Q, Qiao X, et al. Author-Topic over Time (AToT): a dynamic users' interest model[M]//Mobile, ubiquitous, and intelligent computing. Berlin:Springer,2014:239-245.
 - [12] Griffiths T L, Steyvers M. Finding scientific topics[J]. Proceeding of the National Academy of Science,2004,101(S1):5228-5235.
 - [13] Blei D M, Lafferty J D. Dynamic topic models[C]//Proceedings of the 23rd international conference on machine learning. [s. l.]:[s. n.],2006:113-120.
 - [14] Ahmed A, Xing E P. Timeline: a dynamic hierarchical

- Dirichlet process model for recovering birth/death and evolution of topics in text stream[C]//Proceedings of the 26th conference on uncertainty in artificial intelligence. [s. l.]:AUAI Press,2010.
- [15] Wei X, Sun J, Wang X. Dynamic mixture models for multiple time-series[C]//Proceedings of the 20th international joint conference on artificial intelligence. Hyderabad, India: [s. n.],2007:2909-2914.
 - [16] Al Sumait L, Barabur D, Domeniconi C. On-line lda: adaptive topic models for mining text streams with applications to topic detection and tracking[C]//Proc of eighth IEEE international conference on data mining. [s. l.]:IEEE,2008:3-12.
 - [17] Wang C, Paisley J W, Blei D M. Online variational inference for the hierarchical Dirichlet process[C]//Proc of international conference on artificial intelligence and statistics. [s. l.]:[s. n.],2011:752-760.