

基于上下文多维度的移动用户偏好动态分析方法

罗晓东

(江苏省流通现代化传感网工程技术研究开发中心 南京 211168)

(南京大学信息管理学院 南京 210023)

摘要 移动用户偏好的动态分析由于引入了上下文数据,使得原有的用户-项目二维矩阵将扩展为用户-项目-上下文的三维矩阵。根据多维矩阵中低秩分解理论,可以简化数据的分析,但是其移动用户偏好动态分析的自学习方法没有充分利用多维矩阵的低秩分解性质。针对此问题,提出了基于多维度上下文的张量低秩分解的自学习方法,此方法基于张量的平行因子分解性质,加快了算法的收敛速度,降低了数据分析的复杂度。仿真结果验证了算法在移动用户偏好估计精度方面的有效性。

关键词 移动互联网,多维矩阵,用户偏好,自学习

中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.02.038

Dynamic Analysis Method of Mobile User Preference Context Based on Multi-dimensional

LUO Xiao-dong

(Jiangsu Provincial Engineering Technology Research & Development Center for Distribution

Modernization Sensor Network, Nanjing 211168, China)

(School of Information Management, Nanjing University, Nanjing 210023, China)

Abstract Because contextual data is introduced into dynamic analysis of mobile Users preference, the original user-project two-dimensional matrix will be extended to users-Projects-Context three dimensional matrix decomposition theory based on multi-dimensional matrix of low rank, which can simplify the analysis of the data, but the low rank decomposition properties of self-learning method for a mobile user preferences dynamic analysis do not take full advantage of multi-dimensional matrix. To solve this problem, this paper presented a self-learning method which uses multi-dimensional matrix of low rank decomposition, promoting the convergence rate, reducing the data analysis complexity. The simulation results show the effectiveness of the proposed algorithm.

Keywords Mobile internet, Multi-dimensional matrix, User preference, Self-learning method

1 引言

随着移动互联网的发展,移动用户可以通过高传输速率的移动网络接入,特别是未来第五代移动通信的部署,会提供更高速率、更低时延的接入网络。尤其是移动互联网云计算的应用,使得移动终端可以通过低时延的网络得到更多的计算资源和存储资源,从而突破移动终端计算资源和存储资源受限的瓶颈,移动终端可以从事更加复杂和丰富多彩的业务。正是基于此背景,可以根据移动互联网中用户参与各个业务的数据,通过对数据的上下文进行分析,得到用户的偏好。然而由于移动用户数量和业务种类多以及上下文的引入,使得传统的数据结构产生了变化,即由传统的用户和项目二维数据矩阵变为了用户、项目和上下文组成的三维数据;由于上下文和用户偏好的客观存在,使得这些数据具有稀疏的特性。

这种数据的稀疏特性给移动用户偏好数据的表示、传输以及预测带来了不便,增加了数据传输的成本,消耗了大量的存储资源和计算资源;同时对移动用户偏好的预测时间也会增加,从而影响移动用户享受推荐业务的用户体验质量。传统的用户、项目和上下文三维数据模型均以二维数据矩阵方式存储,并且应用自适应学习的人工智能方法进行分析也是基于二维数据模型的。文献[1]提出了应用随机马尔科夫方法对相关的上下文信息进行分析和预测,其是将不同的上下文信息分成不同的随机状态,从而根据随机状态的转移进行数据的压缩和挖掘;文献[2]提出了一种基于音乐推荐的暂时性上下文分析方法,此方法能够在不同文本输入的条件下进行持续性概率时间分析,这是在时间的数据维度进行数据的压缩和挖掘,这种方法充分抓住了分析的业务的特点,即音乐推荐的输入是文本,而推荐的是音频,并且音频是音乐,会持

到稿日期:2016-09-06 返修日期:2016-11-06 本文受江苏省国内高级访问学者计划(2014FX024),江苏高校品牌专业建设工程一期 A 类项目(PZY2015A096),江苏省教育厅流通现代化传感网研发中心基金项目(2010-4)资助。

罗晓东(1971—),女,硕士,副教授,CCF 会员,主要研究方向为信息智能处理、电子商务, E-mail: LuoxiaodongJX@qq.com。

续固定的时间;文献[3]提出了一种基于平均精度的优化算法,此方法从预测精度的目标出发,采用运筹学进行优化;而文献[4]则注重从快速自学习的角度去优化上下文用户偏好推荐算法;文献[5]提出了基于用户、项目和上下文的三维矩阵评价用户偏好,这是综合考虑了多种因素的评判标准。

根据所述的研究背景和研究现状,可以发现以下几个问题:1)所研究的基于上下文移动用户偏好的分析的数据是多维度的,包括时间、输入数据、项目和上下文种类等;2)其数据由于是某个用户的相关数据,其用户的偏好一定是客观存在的,因此其数据具有稀疏特性;3)数据的相互联系以及数据的分析均关系到后续用户偏好分析的准确程度、算法的计算复杂度以及收敛性,从而可以得到分析多维数据的方法是基于上下文移动用户偏好分析的关键;4)对于多维数据,没有相关的分析和研究,从而造成了算法的优化缺乏效率,并且推荐精度较低。

对多维矩阵进行分析是一个新兴的分析领域,其综合了统计学和矩阵分析理论。传统的统计学随着数据维度的增加会出现“维度灾难”,即随着维度的增加,其需要的分析样本需呈几何倍增长才能保证良好的统计结果。然而一般所获得的数据是有限的,这就会导致小样本的统计结果。故需要对所描述的数据进行多维表示,找到数据维度的联系,从而实现所得数据的有效性和关联性,提高自学习的效率,保证数据的有效利用和挖掘。利用多维矩阵的张量分析是行之有效的方法之一,具有以下好处:1)应用张量理论可以保持上下文数据的多维特性,从而可以利用多维数据的相关性进行分析和压缩;2)应用多维矩阵的分解特性,能够降低维度表示,特别是立体矩阵的分解唯一性,可以提高数据分析的有效性;3)张量积利于计算机处理,可以降低算法的复杂度。故张量理论应用在多个领域,已经取得了很好的效果。文献[6]将张量理论应用到人脸识别中,获得了较好的识别效果,文献[7]则将张量理论应用到雷达信号处理中,获得了较好的到达角估计。

平行因子张量分解理论是 20 世纪 70 年代提出的,最早应用于通信领域。由于平行因子张量分解具有低秩降维的特性,因此可以将其应用到数据挖掘领域,用于数据的筛选和压缩,这不仅减少了数据的传输,而且也提高了数据挖掘的速度。本文主要采用张量的平行因子分解对数据进行低秩分解,从而获得自适应学习的移动用户偏好动态分析方法,并给出针对移动用户的推荐结果。

2 系统模型

移动互联网中,每一个移动终端通过无线通信链路传输,经过核心网络上传到云数据中心,故移动用户的所有业务均在云数据中心存有相关上下文的数据。云计算中心根据业务需求的不同,对整个移动互联网的计算和存储资源进行调度,并且根据存储资源的传输和调度获取上下文数据,故网络拓扑结构适合移动用户偏好的分析与推荐。其具体拓扑结构如图 1 所示^[8-10]。

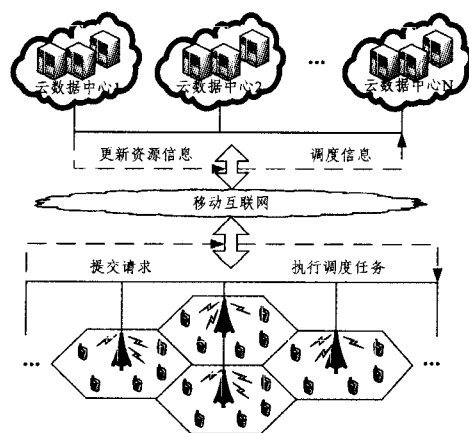


图 1 移动互联网用户上下文获取图

2.1 数据模型

在数据模型方面,考虑上下文参数,如用户的位置、业务应用的时间、用户资料、天气和其他数据参数。给定一个用户 u 的上下文标签,在项目 i 相关的条件下,用户的上下文参数集合可以表示为 c ,其中的参数可以用集合 $\{c_1, c_2, c_3, \dots, c_n\}$ 来表示。将用户 u 和上下文的参数一起存储到一个结构中,则可以表示为^[11-12]:

$$D = \{(u, i, c_1, c_2, c_3, \dots, c_n) | r_{uic_i}\} \quad (1)$$

其中, r_{uic_i} 代表根据上下文参数 c_i 所得到的移动用户的偏好, N 代表上下文未知的项目,若 $c_j = 0$ 则代表上下文未知。

数据模型中,每一个上下文参数 c_i 可以是标量也可以是矢量,随着上下文数据关联关系的不同,可以构造不同的数据结构,但是现实环境中数据结构较为复杂,数据多为矢量结构,这使得数据结构复杂,传统的基于二维矩阵的数据挖掘算法已经无法适用于这种复杂的数据结构源;同时由于不同业务带来的数据表现形式和存储形式也不同,这给异构数据的挖掘带来了困难。针对上述两个方面的问题,提出了基于张量积的表示形式,其可以灵活地表示较为复杂的数据,同时也为异构数据的存储和处理带来了方便。

2.2 数据张量模型

根据上下文参数条件,推荐的结果可以以评价的形式表现出来,其张量表现形式为:

$$p: U \times V \times C_1 \times C_2 \times \dots \times C_k \rightarrow D \quad (2)$$

可以看出,如果将单一的上下文参数组成的维度加载到数据中,则得到的张量评分是一个三维张量形式。通过简化,观测的稀疏张量矩阵可以表示为 $P \in p^{n \times m \times k}$,其中 n 代表用户数, m 代表业务数, k 代表评分的量化个数,如第 j 个上下文的变量 c_j 可以量化为 k 个值,从而具有 k 个上下文可能。

应用此三维矩阵可以得到三维矩阵的一些重要性质,其中最为重要而且在分解中应用较多的就是平行因子分解唯一性原理,其定理可以表述为:考虑一个三维张量形式,其可以表示为 $\underline{X} \in R^{I_1 \times I_2 \times I_3}$,其平行因子的分解标量形式为:

$$\underline{x}_{i_1, i_2, i_3} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} a_{i_1 r_1}^{(1)} a_{i_2 r_2}^{(2)} c_{r_1 r_2} b_{i_3 r_1}^{(1)} b_{i_3 r_2}^{(2)} \quad (3)$$

其中, $i_1 = 1, \dots, I_1$, $i_2 = 1, \dots, I_2$ 和 $i_3 = 1, \dots, I_3$ 。 $\underline{x}_{i_1, i_2, i_3}$ 是三维张量矩阵的元素,而 $a_{i_1 r_1}^{(1)}$, $a_{i_2 r_2}^{(2)}$, $c_{r_1 r_2}$, $b_{i_3 r_1}^{(1)}$ 以及 $b_{i_3 r_2}^{(2)}$ 是三维

张量投影到二维矩阵的元素,其对应的矩阵为 $\mathbf{A}^{(1)} \in R^{I_1 \times R_1}$, $\mathbf{A}^{(2)} \in R^{I_2 \times R_2}$, $\mathbf{C} \in R^{I_3 \times R_3}$, $\mathbf{B}^{(1)} \in R^{I_1 \times R_1}$ 和 $\mathbf{B}^{(2)} \in R^{I_2 \times R_2}$ 。 $\mathbf{A}^{(1)}$ 和 $\mathbf{A}^{(2)}$ 是三维张量的平行因子, $\mathbf{B}^{(1)}$ 和 $\mathbf{B}^{(2)}$ 为关联作用矩阵, $\mathbf{C}$ 是张量分解的核矩阵。式(3)可以表示为矩阵形式:

$$\mathbf{x}_{i_1, i_2, i_3} = \mathbf{A}^{(1)}(i_1, :) \mathbf{D}_{i_3}(\mathbf{B}^{(1)}) \mathbf{C} \mathbf{D}_{i_3}(\mathbf{B}^{(2)}) (\mathbf{A}^{(2)}(i_2, :))^T \quad (4)$$

其中, $\mathbf{A}^{(1)}(i_1, :)$ 代表矩阵的第 i_1 行。定义矩阵的剖面 $\mathbf{X}_{i_3}^{(3)}$ 是三维张量 $\mathbf{X}$ 的第三个维度的截面,故可以用矩阵表达式得到:

$$\mathbf{X}_{i_3}^{(3)} = \mathbf{A}^{(1)} \mathbf{D}_{i_3}(\mathbf{B}^{(1)}) \mathbf{C} \mathbf{D}_{i_3}(\mathbf{B}^{(2)}) (\mathbf{A}^{(2)})^T \quad (5)$$

3 基于张量分解的用户偏好分析

3.1 关联数据的平行因子分解

将三维矩阵 $\mathbf{P}$ 进行分解,分解后得到用户和推荐的分解二维截面矩阵,根据模型中的平行因子分解,三维张量矩阵 $\mathbf{P}$ 可以分解为:

$$\hat{\mathbf{P}} = \mathbf{G} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{C} \quad (6)$$

按照其元素的定义,可以表示为:

$$\hat{p} = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^{K_1} g_{ijk} u_i \circ v_j \circ c_k \quad (7)$$

可以表示为矩阵形式:

$$\hat{\mathbf{P}}_{i_3}^{(3)} = \mathbf{G}^{(1)} \mathbf{D}_{i_3}(\mathbf{U}^{(1)}) \mathbf{C} \mathbf{D}_{i_3}(\mathbf{V}^{(2)}) (\mathbf{G}^{(2)})^T \quad (8)$$

在矩阵形式中,矩阵 $\hat{\mathbf{P}}_{i_3}^{(3)}$ 是上下文参数和用户之间的关系,也是三维矩阵的映射在二维空间的体现。可以用矩阵之间的距离来衡量二维空间的准确性。对于两个矩阵,其估计矩阵和观测矩阵之间的距离采用文献[8]的距离公式,如式(9)所示:

$$L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \frac{\hat{\mathbf{P}}_{i_3}^{(3)} \otimes \mathbf{P}_{i_3}^{(3)}}{\text{tr}(\|\hat{\mathbf{P}}_{i_3}^{(3)}\|_F) \text{tr}(\|\mathbf{P}_{i_3}^{(3)}\|_F)} \quad (9)$$

其中, $\otimes$ 代表矩阵的张量积, $\|\cdot\|_F$ 代表矩阵的 Frobenius 范数, $\text{tr}(\cdot)$ 代表矩阵的迹。

3.2 基于平行因子分解的优化算法

为了达到最优的移动用户和上下文偏好的估计值,需要构造合理的优化模型,其优化问题可以表示为式(10):

$$\min \text{Line}(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \frac{\hat{\mathbf{P}}_{i_3}^{(3)} \otimes \mathbf{P}_{i_3}^{(3)}}{\text{tr}(\|\hat{\mathbf{P}}_{i_3}^{(3)}\|_F) \text{tr}(\|\mathbf{P}_{i_3}^{(3)}\|_F)} + \frac{\lambda}{2} \text{tr}(\mathbf{A}^{(1)} \mathbf{D}_{i_3}(\mathbf{B}^{(1)}) \mathbf{D}_{i_3}(\mathbf{B}^{(2)}) (\mathbf{A}^{(2)})^T) \quad (10)$$

$$\text{s.t. } \mathbf{A}^{(1)} (\mathbf{A}^{(1)})^T = \mathbf{I}$$

$$\mathbf{A}^{(2)} (\mathbf{A}^{(2)})^T = \mathbf{I}$$

同时在考虑贝叶斯先验知识概率约束的条件下,得到优化的条件函数,从而提高动态移动用户偏好分析的准确性。可以考虑用户的上下文的概率分布模型,根据观测的概率模型,即

$$p(C_{j_i}) = N(c_{j_i} | \mu_{j_i}, \Sigma) \quad (11)$$

也就是观测到的上下文数据取值符合实数的正态分布。其均

值和方差分别为 μ_{j_i} 和 Σ 。采用稀疏贝叶斯的平行因子分解的学习原理,联合概率分布为:

$$p(C_{j_i}, \mathbf{G}^{(m)} |_{m=1}^2, \mathbf{U}^{(1)}, \mathbf{V}^{(2)}, \hat{\mathbf{P}}_{i_3}^{(3)}) = p(\hat{\mathbf{P}}_{i_3}^{(3)} | C_{j_i}) \prod_{m=1}^2 p(\mathbf{G}^{(m)} | \mathbf{U}^{(1)}) p(\mathbf{V}^{(2)}) \quad (12)$$

在求解最优的上下文和用户之间的关系矩阵 $\hat{\mathbf{P}}_{i_3}^{(3)}$ 的估计值时采用拉格朗日乘数法,其下降梯度最大的方法是保证算法的收敛性。首先计算距离函数的梯度,然后采用最优梯度下降法更新 $\mathbf{U}^{(1)}$, $\mathbf{V}^{(2)}$, $\mathbf{C}_{i_3}$ 和 $\mathbf{G}^{(m)}$ 。距离函数的梯度可以表示为:

$$\partial_{U_i} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \partial_{c_{ijk}} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) \mathbf{G}^{(1)} \times_{U^{(1)}} \mathbf{U}^{(1)} \times_{V^{(2)}} \mathbf{V}^{(2)} \times_C \mathbf{C}_{i_3} \quad (13)$$

同时

$$\partial_{V_i} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \partial_{c_{ijk}} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) \mathbf{G}^{(1)} \times_{U^{(1)}} \mathbf{U}^{(1)} \times_C \mathbf{C}_{i_3} \quad (14)$$

$$\partial_{C_i} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \partial_{c_{ijk}} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) \mathbf{G}^{(1)} \times_{U^{(1)}} \mathbf{U}^{(1)} \times_{V^{(2)}} \mathbf{V}^{(2)} \quad (15)$$

$$\partial_G L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) = \partial_{c_{ijk}} L(\hat{\mathbf{P}}_{i_3}^{(3)}, \mathbf{P}_{i_3}^{(3)}) \otimes \mathbf{U}^{(1)} \otimes \mathbf{V}^{(2)} \otimes \mathbf{C}_{i_3} \quad (16)$$

根据以上的推导,可以得到张量分解后的梯度最大化迭代方案,此方案步骤如下。

(1)输入:在不同的时间对用户的上下文进行采集,并根据不同用户、不同上下文和不同时间的数据建立三维张量矩阵模型,并记作 $\mathbf{P}$ 。

对 $\mathbf{P}$ 进行分析,考虑其是否符合分解唯一性条件。如果符合,则进行分解;如果不符合,则继续收集数据建立三维张量矩阵模型 $\mathbf{P}$ 。

(2)分解张量数据模型 $\mathbf{P}$ 。

1)根据三维张量模型的分解唯一性,利用平行因子分解,得到式(6)。

2)抽取式(6)的映射二维矩阵,得到用户与上下文参数之间的关系,如式(8)。

(3)从 $i=0$ 到 $i=n$

1)利用式(12)计算相关的先验概率;

2)利用式(13)计算关于 $\mathbf{U}^{(1)}$ 的梯度;

3)利用式(14)计算关于 $\mathbf{V}^{(2)}$ 的梯度;

4)利用式(15)计算关于 $\mathbf{C}_{i_3}$ 的梯度;

5)利用式(16)计算关于 $\mathbf{G}^{(m)}$ 的梯度;

6)按照 $\mathbf{U}^{(1)}$, $\mathbf{V}^{(2)}$, $\mathbf{C}_{i_3}$ 和 $\mathbf{G}^{(m)}$ 的顺序进行迭代,从而不断优化收敛,当 $i=n$ 时迭代结束。

(4)计算推荐结束。

4 实验仿真结果

本文所采用的实验数据为麻省理工学院的现实挖掘的数据集合[8],此集合里面具有 203 个移动互联网用户应用移动终端开展业务的数据记录。在本次实验仿真过程中,实验硬件配置是一个工作站,基本配置为双 Intel CPU 6 核处理器,

主频 2GHz, 内存 32GB, 硬盘 2TB。在该机器上创建 5 台虚拟机, 虚拟机系统环境为 64 位, 每台虚拟机配置系统, 分配 4GB 内存, 双核 CPU, 硬盘空间 160GB。实验平台采用分布式环境进行搭建, 其中一台作为 Master, 其余 4 台为 Slave。同时搭建 Hadoop 环境, 数据以 HDFS 文件的形式存储于此机器。考虑的上下文信息包括移动用户终端的位置信息、移动用户开展业务的数据信息、周围人员与此用户的关系以及使用设备和周围人员等。本文从张量矩阵分解和映射得到的结果出发, 分别对用户中上下文条件下用户的偏好行为的波动率和用户偏好预测的准确度、响应时间进行了比较。如图 2 所示。

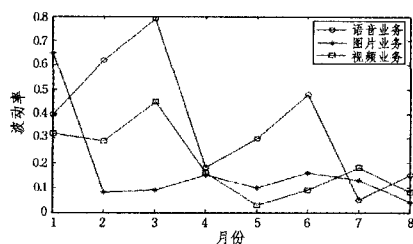


图2 偏好行为波动率对比

图2说明了不同时间下不同业务类型的波动率的变化情况。可以看出, 随着时间的推移, 业务类型不同, 变化规律也不相同, 可以通过此统计得到移动用户在时间维度的偏好动态。

为监测所提算法的精确度和相应时间, 将所提算法和其他4种代表性学习方法进行对比, 结果如图3和图4所示。图中横坐标中1代表支持向量机的学习方法; 2代表神经网络方法; 3代表本文所提方法; 4代表加权遗传算法; 5代表蚁群算法。

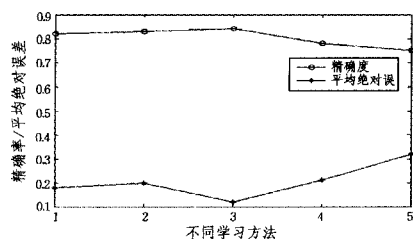


图3 精确度/平均绝对误差

图3显示了不同的学习方法所对应的不同精确度和平均绝对误差。从图中可以看出基于张量分解的学习方法的精确度最高, 平均绝对误差最低, 说明了基于张量的分析因利用了多维度的联系, 学习效率较高, 得到了较高的精确度, 因此所提的算法具有较好的准确度。

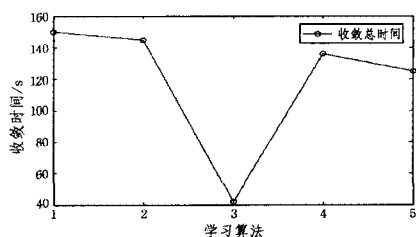


图4 响应时间对比

图4显示了在5种不同学习方法下得到的检测时间对比结果。从此图可以看出, 基于张量分解的学习方法的响应时间最短, 说明所提算法的计算复杂度较低, 在较短的时间内就可以得到准确的移动用户偏好。

结束语 本文根据移动互联网的网络结构和用户业务的特点, 提出了一种基于多维度上下文的张量分解自学习算法, 此算法应用了平行因子分解方法, 根据贝叶斯概率构造搜索方案。仿真实验结果表明, 所提的方案在移动用户偏好估计的精度和算法复杂度方面具有优势。此算法具有广泛的应用场景, 例如利用在线的移动互联网商业信息, 对消费产品的品牌偏好、种类偏好等进行预测和推荐, 具有潜在的商业价值。

参考文献

- [1] NIARAKI A S, KIM K. Ontology based personalized route planning system using a multi-criteria decision making approach[J]. Expert Systems with Applications, 2009, 36(2): 2250-2259.
- [2] DU C J, SUN D W. Learning techniques used in computer vision for food quality evaluation: a review[J]. Journal of food engineering, 2006, 72(1): 39-55.
- [3] LUO X, LV Y, LI R, et al. Web Service QoS Prediction Based on Adaptive Dynamic Programming Using Fuzzy Neural Networks for Cloud Services[J]. IEEE Access, 2015, 3: 2260-2269.
- [4] GHOBADI S. What drives knowledge sharing in software development teams: A literature review and classification framework[J]. Information & Management, 2015, 52(1): 82-97.
- [5] CANUTO C, SIMONCINI V, VERANI M. Contraction and optimality properties of an adaptive Legendre-Galerkin method: the multi-dimensional case[J]. Journal of Scientific Computing, 2015, 63(3): 769-798.
- [6] CEPEDA-GOMEZ R, MICHIELS W. Some special cases in the stability analysis of multi-dimensional time-delay systems using the matrix Lambert W function[J]. Automatica, 2015, 53: 339-345.
- [7] FORSYTHE P, JUPP J, SAWHEY A. Building information modelling in tertiary construction project management education: A programme-wide implementation strategy[J]. Journal for Education in the Built Environment, 2013, 8(1): 16-34.
- [8] RODRIGUEZ-SANCHEZ M C, MARTINEZ-ROMO J, BORROMEO S, et al. GAT: Platform for automatic context-aware mobile services for m-tourism[J]. Expert Systems with Applications, 2013, 40(10): 4154-4163.
- [9] MEHROTRA A, HENDLEY R, MUSOLESI M. PrefMiner: Mining User's Preferences for Intelligent Mobile Notification Management[C] // The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016). 2016: 1223-1234.
- [10] CHOU C K, LIN C C, CHEN M S. Context-aware daily activity summarization with adaptive transmission[C] // Proceedings of the 31st Annual ACM Symposium on Applied Computing. ACM, 2016: 560-565.

- Empirical Methods in Natural Language Processing (EMNLP 2009). East Stroudsburg, United States, 2009; 1446-1454.
- [8] XU J, XU R, LU Q, et al. Coarse-to-fine Sentence-level Emotion Classification Based on the Intra-sentence Features and Sentential Context [C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012). United States, 2012; 2455-2458.
- [9] LIN K, YANG C, CHEN H. Emotion Classification of Online News Articles from the Reader's Perspective [C]// Proceedings of IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops (WI-IAT 2008). Sydney, NSW, Australia, 2008; 220-226.
- [10] ALM C, ROTH D, SPROAT R. Emotions from Text: Machine Learning for Text-based Emotion Prediction [C]// Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2009). East Stroudsburg, United States, 2009; 579-586.
- [11] TOKUHISA R, INJI K, MATSUMOTO Y. Emotion Classification Using Massive Examples Extracted from the Web [C]// Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008). 2008; 881-888.
- [12] BHOWMICK P, BASU A, MITRA P, et al. Multi-label Text Classification Approach for Sentence Level News Emotion Analysis [C]// Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science, Berlin, Germany, 2009; 261-266.
- [13] LI S, HUANG L, WANG R, et al. Sentence-level Emotion Classification with Label and Context Dependence [C]// Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL 2015). Beijing, China, 2015; 1045-1053.
- [14] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment Classification Using Machine Learning Techniques [C]// Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2010). 2002; 79-86.
- [15] DAVIDOV D, TSUR O, RAPPOPORT A. Enhanced Sentiment Learning Using Twitter Hastags and Smileys [C]// Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008). 2010; 241-249.
- [16] LIU H, LI S, ZHOU G, et al. Joint Modeling of News Reader's and Comment Writer's Emotions [C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013). Sofia, Bulgaria, 2013; 511-515.
- [17] LI C, WU H, JIN Q. Emotion Classification of Chinese Microblog Text via Fusion of BoW and eVector Feature Representations [C]// Proceedings of the 3rd CCF Conference on Natural Language Processing and Chinese Computing (NLP&CC 2014). 2014; 217-228.
- [18] QUAN C, REN F. Sentence Emotion Analysis and Recognition Based on Emotion Words Using Ren-CECPs [J]. International Journal of Advanced Intelligence, 2010, 2(1): 105-117.
- [19] YAO Y L, WANG S W, XU R F, et al. The Construction of an Emotion Annotated Corpus on Microblog Text [J]. Journal of Chinese Information Processing, 2014, 28(5): 83-91. (in Chinese)
- 姚源林, 王树伟, 徐睿峰, 等. 面向微博文本的情绪标注语料库构建 [J]. 中文信息学报, 2014, 28(5): 83-91.
- [20] MAEDA H, SHIMADA K, ENDO K. Twitter Sentiment Analysis Based on Writing Style [C]// Proceedings of the 8th International Conference on NLP (NLP 2012). Kanazawa, Japan, 2012; 278-288.
- [21] ZHANG J, ZHU B, LIANG L L, et al. Recognition and Classification of Emotions in the Chinese Microblog Based on Emotional Factor [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2014, 50(1): 79-84. (in Chinese)
- 张晶, 朱波, 梁琳琳, 等. 基于情绪因子的中文微博情绪识别与分类 [J]. 北京大学学报(自然科学版), 2014, 50(1): 79-84.
- [22] HUANG L, LI S, ZHOU G. Emotion Corpus Construction on Microblog Text [C]// Proceedings of the 16th Workshop on Chinese Lexical Semantics Workshop (CLSW 2015). Beijing, China, 2015; 204-212.
- [23] HIRST G, FEIGUINA O. Bigrams of Syntactic Labels for Authorship Discrimination of Short Texts [J]. Literary & Linguistic Computing, 2007, 22(4): 405-417.
- [24] CHEN F, CHAO W H, ZHOU Q, et al. Convolution Tree Kernel Based Sentiment Element Recognition Approach for Chinese Microblog [J]. Computer Science, 2014, 41(12): 133-137, 142. (in Chinese)
- 陈锋, 巢文涵, 周庆, 等. 基于卷积树核的中文微博情感要素识别 [J]. 计算机科学, 2014, 41(12): 133-137, 142.

(上接第238页)

- [11] GARZA P, QUINTARELLI E, RABOSIO E, et al. Reducing Big Data by Means of Context-Aware Tailoring [M]// New Trends in Databases and Information Systems. Springer International Publishing, 2016; 115-127.
- [12] OH K J, KIM Z, OH H, et al. Travel intention-based attraction network for recommending travel destinations [C]// 2016 International Conference on Big Data and Smart Computing (Big-Comp). IEEE, 2016; 277-280.
- [13] ZHANG J W, YANG Z. Collaborative Filtering Recommendation Algorithm Based on Improved User Clustering [J]. Computer Science, 2014, 41(12): 176-178. (in Chinese)
- 张峻玮, 杨洲. 一种基于改进的层次聚类的协同过滤用户推荐算法研究 [J]. 计算机科学, 2014, 41(12): 176-178.
- [14] YU X S, SHAN H. Research on Personalized Recommendation Model Based on Network Users' Information Behavior [J]. Journal of Chongqing Institute Technology, 2013, 27(1): 47-50. (in Chinese)
- 余肖生, 孙珊. 基于网络用户信息行为的个性化推荐模型 [J]. 重庆理工大学学报(自然科学版), 2013, 27(1): 47-50.