

# 基于时空数据的用户社交联系强度研究

陈元娟 严建峰 刘晓升 杨 璐  
(苏州大学计算机科学与技术学院 苏州 215006)

**摘要** word2vec 是 Google 推出的一款将词表征为实数值的高效开源工具。采用该工具将时空数据中的每位用户表征为一个实数值向量并预测用户间社交联系的强度。提出了在 word2vec 学习过程中动态调整学习率的算法——Location-weight 算法,根据不同位置的不同用户数目在学习过程中加入位置权重,并探索其对用户社交联系强度预测的影响。实验结果表明,加入位置权重的学习算法提高了用户社交联系强度预测的准确性。

**关键词** word2vec, 位置权重, 用户社交联系强度

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.1.054

## Social Strength Learning between Users Based on Spatiotemporal Data

CHEN Yuan-juan YAN Jian-feng LIU Xiao-sheng YANG Lu  
(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

**Abstract** Word2vec is a high-efficiency open-source tool issued by Google, which represents a word with a float vector. We used this tool to process spatiotemporal data. Each user is represented with a float vector to predict social strength among users. To predict social strength among users more precisely, this paper proposed a location-weight algorithm that dynamically adjusts the learning rate in the process of learning with word2vec. According to the number of different users at different locations, we added location weight to the algorithm in the process of learning. Meanwhile, we explored the effect of location-weight on prediction of social strength among users. Experimental results validate performance of the proposed algorithm.

**Keywords** word2vec, Location weight, Social strength

## 1 引言

位置信息是时间与空间联合的时空数据,实现了网络和现实的连接,是现实世界和网络服务之间的一座桥梁。随着英特网的迅速发展,应运而生的空间定位技术不断成熟,使得毫不费力地获得位置信息成为可能。位置信息有显式获取和隐式获取两种方式,前者主要通过带有地理标签的内容(如在博客、Instagram、Flickr 或者 Facebook 上传照片)、基于位置的服务网络(如签到)或手机应用程序(如切客、嘀咕等)等直接获取;后者主要指从移动设备(如基站、GPS 或者 WiFi 热点)和信用卡交易记录等来源间接获取。随着时间推移的用户的位置信息(即时空数据)已成为研究各种社会行为的丰富的信息来源。本文主要研究在该类数据集中是否能够推测出用户的社交联系强度这一社会行为。即验证以下猜测:如果两个人在同一个时间同一个地方出现(即共现),则认为他们有很大可能是有社交联系的<sup>[1]</sup>。

2013年,Google 推出了一个用于获得单词向量(word vector)的开源工具包——word2vec,其简单、高效,引起了广

泛的关注。本文利用该工具包获得时空数据中每一位用户的单词向量作为用户的特征向量,并且根据特征向量来获得用户间的社交联系强度。然而,word2vec 工具包存在缺陷,其并没有考虑到时空数据中每个位置的特异性。针对该问题,本文考虑了位置权重对预测用户社交联系强度的影响,在 word2vec 算法中加入位置权重项,并探索其对预测准确性的影响。实验表明加入位置权重动态调整学习率的算法(Location-weight 算法)能提高用户社交联系强度预测的准确性。

## 2 相关工作

Gonzalez 等人通过研究 10 万多名用户的活动轨迹,发现用户的行为呈现“可重现(reproducible)”模式<sup>[2]</sup>。Zhiyuan 等人通过研究 Twitter 带有的位置信息发现影响用户进行签到的动机<sup>[3]</sup>。Lindqvist 和 Kostakos 等人研究了与位置相关的隐私问题<sup>[4,5]</sup>。李敏等人特别关注了用户在消费类地点的言行规律<sup>[10]</sup>。

在时空数据的用户社交联系强度的预测研究中,最主要的是运用时空数据的共现特征来进行实验。如:Crandall 等

到稿日期:2015-01-24 返修日期:2015-03-28 本文受国家自然科学基金(61373092,61033013,61272449,61202029),江苏省教育厅重大项目(12KJA520004),江苏省科技支撑计划重点项目(BE2014005),广东省重点实验室开放课题(SZU-GDPHPCL-2012-09)资助。

陈元娟(1990-),女,硕士生,主要研究方向为机器学习;严建峰(1978-),男,副教授,硕士生导师,主要研究方向为机器学习,E-mail:yanjf@suda.edu.cn(通信作者);刘晓升(1976-),男,博士生,主要研究方向为机器学习;杨璐(1982-),女,副教授,硕士生导师,主要研究方向为机器学习与软件工程。

人提出了基于时间和空间上的共现的概率模型来预测两个人是朋友的概率<sup>[6]</sup>;Cranshaw 等人介绍了多种共现的特征并且运用这些特征于机器学习算法中,来判定用户间是否有联系<sup>[7]</sup>;Huy Pham 等人通过人们的共现的分布和共现的地点特征来计算用户间社会联系强度<sup>[1]</sup>;还有学者运用时空数据上的轨迹特征来预测他们是否有关联<sup>[8]</sup>。

本文主要研究的也是共现对预测用户社交联系强度的影响。不同于以上论文直接去寻找共现所产生的特征然后通过机器学习算法来预测<sup>[7]</sup>或者是直接通过计算共现带来的影响来预测用户间社交联系强度<sup>[1,6]</sup>,本文通过 word2vec 算法<sup>[9,11,12]</sup>将时空数据的共现转化为用户的特征向量,然后再预测用户间的社交联系强度,从而使得这种基于时空数据预测用户社交联系强度的方法与现有技术相比,操作更简单,且提高了精度。然而,word2vec 算法本身没有考虑到时空数据中地理位置的特异性。本文在 word2vec 算法产生用户特征的过程中提出了在 word2vec 学习过程中动态调整学习率的算法——Location-weight 算法。根据不同位置的不同用户数目,在学习过程中加入位置权重,并探索其对用户社交程度预测的影响。实验表明加入位置权重动态调整学习率的算法更能提高用户社交联系强度预测的准确率。

### 3 word2vec 算法介绍

本实验涉及 word2vec 算法。word2vec 拥有两个重要模型:CBOW 模型(Continue Bag-of-Words Model)和 Skip-gram 模型。且对于每个模型,word2vec 给出了两个算法,分别是: Hierarchical Soft 和 Negative Sample。

下面对 Skip-gram 模型中的 Hierarchical Soft 进行介绍。

#### 3.1 Skip-gram 模型中的 Hierarchical Soft 算法

Skip-gram 模型中的 Hierarchical Soft 算法的网络结构如图 1 所示,它包含输入层、投影层和输出层。其中,输入层只含当前样本的中心词  $w$  的词向量  $V(w)$ ;投影层是输入层的恒等投影,把  $V(w)$  投影到  $V(w)$ ;输出层是一棵 Huffman 树。输出层中的 Huffman 树是以语料中出现过的词为叶子结点,以各词在语料中出现的次数为权值构造出来的。

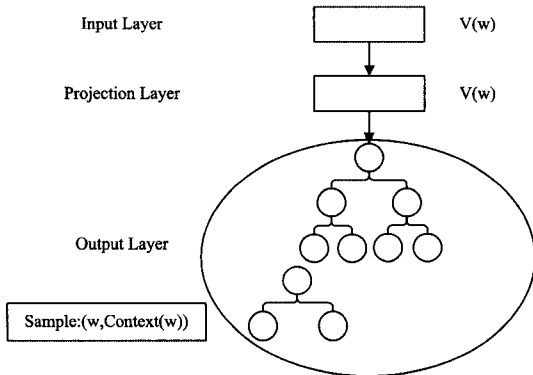


图 1 Skip-gram 模型中 Hierarchical Soft 算法的网络结构

对于 Skip-gram 模型,已知当前词  $w$ ,需要对其上下文  $Context(w)$  中的词进行预测,因此目标函数如式(1)所示,且关键是条件概率函数  $p(Context(w)|w)$  的构造,Skip-gram 将其定义为式(2)。以下公式中的符号标签如表 1 所列。

表 1 符号标签

符号	意义
$p^w$	从根节点出发到达单词 $w$ 对应叶子节点的路径
$l^w$	路径 $p^w$ 中包含结点的个数
$p_1^w, p_2^w, \dots, p_{l^w}^w$	路径 $p^w$ 中的 $l^w$ 个结点
$d_2^w, d_3^w, \dots, d_{l^w}^w \in \{0, 1\}$	词 $w$ 的 Huffman 编码
$\theta_1^w, \theta_2^w, \dots, \theta_{l^w-1}^w$	路径 $p^w$ 中非叶子结点对应的向量

其中,

$$L = \sum_{w \in C} \log p(context(w) | w) \quad (1)$$

$$p(Context(w) | w) = \prod_{u \in Context(w)} p(u | w) \quad (2)$$

上式中的  $p(u|w)$  可以写成

$$p(u|w) = \prod_{j=2}^{l^w} p(d_j^u | V(w), \theta_{j-1}^u) \quad (3)$$

其中,

$$p(d_j^u | V(w), \theta_{j-1}^u) = [\sigma(V(w)^T \theta_{j-1}^u)]^{1-d_j^u} * [1 - \sigma(V(w)^T \theta_{j-1}^u)]^{d_j^u} \quad (4)$$

根据上面的公式得:

$$\begin{aligned} L &= \sum_{w \in C} \log \prod_{u \in Context(w)} \prod_{j=2}^{l^u} \{ [\sigma(V(w)^T \theta_{j-1}^u)]^{1-d_j^u} * [1 - \sigma(V(w)^T \theta_{j-1}^u)]^{d_j^u} \} \\ &= \sum_{w \in C} \sum_{u \in Context(w)} \sum_{j=2}^{l^u} \{ (1-d_j^u) * \log[\sigma(V(w)^T \theta_{j-1}^u)] + d_j^u * \log[1 - \sigma(V(w)^T \theta_{j-1}^u)] \} \quad (5) \end{aligned}$$

为方便梯度推导,令:

$$L(w, u, j) = (1-d_j^u) * \log[\sigma(V(w)^T \theta_{j-1}^u)] + d_j^u * \log[1 - \sigma(V(w)^T \theta_{j-1}^u)] \quad (6)$$

则考虑  $L(w, u, j)$  关于  $\theta_{j-1}^u$  的梯度计算为:

$$\frac{\partial L(w, u, j)}{\partial \theta_{j-1}^u} = [1-d_j^u - \sigma(V(w)^T \theta_{j-1}^u)] * V(w) \quad (7)$$

于是

$$\theta_{j-1}^u := \theta_{j-1}^u + \eta [1-d_j^u - \sigma(V(w)^T \theta_{j-1}^u)] * V(w) \quad (8)$$

同样得到:

$$\frac{\partial L(w, u, j)}{\partial V(w)} = [1-d_j^u - \sigma(V(w)^T \theta_{j-1}^u)] * \theta_{j-1}^u \quad (9)$$

于是,  $V(w)$  的更新公式可以写为:

$$V(w) := V(w) + \eta \sum_{u \in Context(w)} \sum_{j=2}^{l^u} \frac{\partial L(w, u, j)}{\partial V(w)} \quad (10)$$

结合以上数学公式的推导,word2vec 算法的伪代码如下。

**算法** word2vec 算法中 Skip-gram 模型中采用的 Hierarchical Soft 算法伪代码

FOR  $u \in Context(w)$  DO

{

$\Delta V(w) = 0$

FOR  $j = 2 : l^u$  DO

{

$f = \sigma(V(w)^T \theta_{j-1}^u)$

$g = \eta(1-d_j^u - f)$

$\Delta V(w) = \Delta V(w) + g * (\theta_{j-1}^u)$

$\theta_{j-1}^u := \theta_{j-1}^u + g * V(w)$

}

$V(w) := V(w) + \Delta V(w)$

}

#### 4 基于位置权重动态调整学习率的算法 (Location-weight 算法)

本文将 word2vec 工具应用于时空数据中,将每位用户表征为一个实数值向量,来预测用户间社交联系的强度。本文提出了在 word2vec 学习过程中动态调整学习率的算法,根据不同位置的不同用户数目,在学习过程中加入位置权重,并探索其对用户社交程度预测的影响。以下将介绍本文提出的动态调整学习率的算法——Location-weight 算法。

就时空数据而言,其具有共现的多样性和共现的偶然性。共现的多样性是指:在预测两个用户的社交联系强度时,我们认为这两个用户经常共现在不同的位置(即多样性高),暗示着这两个用户也许存在社交联系。共现的偶然性是指:当两个用户共现在人多的位置时,该位置很有可能是一个公共场所,暗示着这两个用户的共现也许只是偶然;当两个用户出现在人少的位置时,即使他们出现的次数不多,但是该位置很有可能是一个私人场所,该共现将会起到很重要的作用<sup>[1]</sup>。然而,word2vec 算法存在一个缺陷,其并没有考虑到时空数据中每个位置的特异性。针对这个问题,我们考虑了位置权重对预测用户社交联系强度的影响,在 word2vec 算法中加入位置权重项,并探索其对预测准确性的影响。即提出的 Location-weight 算法将位置权重带入 word2vec 学习过程中,动态地调整学习率。Location-weight 算法将每位用户表征为一个实数值向量,来预测用户间社交联系的强度。

在本文算法中将式(10)稍作修改,为:

$$V(w) := V(w) + \eta \sum_{u \in \text{Context}(w)} \sum_{j=2}^m \frac{\partial L(w, u, j)}{\partial V(w)} * \text{LocationWeightValue} \quad (11)$$

其中,  $L(w, u, j)$  是该模型的目标函数,  $V(w)$  是当前样本的中心词  $w$  的词向量,  $\eta$  是算法的学习率,  $\frac{\partial L(w, u, j)}{\partial V(w)}$  是  $L(w, u, j)$  关于  $V(w)$  的梯度,  $\text{LocationWeightValue}$  是一个关于地点上出现的不同的人的数目的一个递减函数。这里的  $\text{LocationWeightValue}$  起到动态调整学习率的作用。

实验中用到的  $\text{LocationWeightValue}$  的计算公式为:

$$f(x, \alpha) = \alpha(1 + e^{-x}) \quad (12)$$

式中,  $\text{LocationWeightValue}$  有两个参数分别为:  $\alpha$  和  $x$ , 其中  $x$  表示某一地点上出现的不同的人的数目,  $\alpha$  是一个大于 0 的实数。根据式(12)可知, 当某一个地点上出现不同的人的数目较少时,  $\text{LocationWeightValue}$  是一个较大的值; 当一个地点上出现不同的人的数目较多时,  $\text{LocationWeightValue}$  却是一个较小的值。结合式(11)、式(12)可知, Location-weight 算法结合了地点特征, 通过动态调整学习率, 避免了共现的偶然性。

值得注意的是这里的递减函数是关于地点上的不同人数的数目的一个递减函数, 而不是关于该地点的签到数目的递减函数。因为一个地点是否为一个公共场所是由不同的人的数目决定, 而不是由签到的数目决定。所以, 我们的算法首先求得一个地点上出现不同的人的数目, 然后据此给  $\text{LocationWeightValue}$  赋予不同的值。

本实验的流程如图 2 所示。首先处理时空数据, 将时空

数据处理成只有用户编号但同样具有时空性的数据; 然后将只有用户编号的数据输入到 Location-weight 算法中, 得到用户的特征向量; 最后, 将两两用户的特征向量连接起来, 输入到分类器中进行分类, 得出相关评价标准的值。

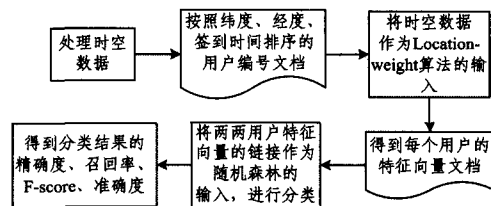


图 2 用户社交联系强度预测的流程

本实验中, Location-weight 算法的伪代码如下。

**算法** Location-weight 算法中 Skip-gram 模型采用的

Hierarchical Soft 算法伪代码

FOR  $u \in \text{Context}(w)$  DO

{

Number = Different\_word(Context(w))

Location\_weight =  $f(\alpha, \text{Number})$

$\Delta V(w) = 0$

FOR  $j = 2 : l^w$  DO

{

$f = \sigma(V(w)^T \theta_{j-1}^u)$

$g = \eta(1 - d_j^u - f)$

$\Delta V(w) = \Delta V(w) + g * (\theta_{j-1}^u)$

$\theta_{j-1}^u := \theta_{j-1}^u + g * V(w)$

}

$V(w) := V(w) + \Delta V(w) * \text{LocationWeightValue}$

}

#### 5 实验分析

本节中, 我们在通用的时空数据集 Gowalla 和 Brightkite 上分别使用精确度 (precision)、召回率 (recall)、F-score 和准确度 (accuracy) 4 个标准来评估 word2vec 和基于位置权重动态调整学习率的 Location-weight 算法。

##### 5.1 数据集

本文主要采用的数据集是基于位置的在线社交网络数据集 (Gowalla, Brightkite)。这两个数据集公布在斯坦福大型网络数据集收集的网站上<sup>1)</sup>。

Gowalla 是一个基于地理位置的社交网站, 其用户通过签到来分享他们的地理位置。其朋友圈是不确定的, 但是通过他们公共的 API 可以获得其朋友圈。该数据集中包含 196591 个结点 (即用户)、950527 条边 (即社交联系) 和 6442890 条签到记录。这些数据是从 2009 年 2 月—2010 年 10 月收集到的记录。其中, 每一条记录的格式是 (用户编号, 签到时间, 纬度, 经度, 位置 ID)。

Brightkite 是一个基于位置的社交网络, 其用户可以通过签到来分享他们的地理位置。同样, 其朋友圈也是通过其公共的 API 来获得的。Brightkite 包含 58228 个结点 (即用户)、214078 条边 (即社交联系) 和 4491143 条签到记录。这些数据是从 2008 年 4 月—2010 年 10 月收集到的记录。其中,

<sup>1)</sup> <http://snap.stanford.edu/data/index.html#socnets>

Brightkite的每一条记录的格式与 Gowalla 的记录格式是一样的。

但是,word2vec 只是一个将单词表征为单词向量的工具。不管是 Gowalla 数据集还是 Brightkite 数据集,其中都包含一些无用的单词如签到时间、纬度、经度、位置 ID。在这个实验中,我们仅仅需要用户的特征向量来预测用户间社交强度即可。因此,仅仅需要一个用户编号元组数据集。为了使得获得的用户编号元组数据集仍具有时空性质,我们将用户的签到数据集按照纬度、经度、签到时间排序。然后,将属于同一个纬度和经度的用户编号分到一个小集合中。将 Gowalla 和 Brightkite 数据集转化为一个仅有用户编号的数据集。该数据集中每一条记录记载了同一个位置上用户签到的前后顺序,即每个数据集是一个在某同一个位置按签到时间排序的用户编号的有序集合。由此可知,该数据集合同样具有时空性质。

### 5.2 评价标准

本文采用检索分类领域通用的性能评价指标:精确度(precision)、召回率(recall)、F-score 和准确度(accuracy)来评价算法的性能。预测结果信息如表 2 所列。

表 2 预测结果信息

	正样例	负样例
预测出来的	预测出来的正确的正样例数 (True Positive, TP)	预测出来的非正样例数 (True Negative, TN)
未预测出来的	未预测出来的正样例 (False Positive, FP)	未预测出来的非正样例数 (False Negative, FN)

$$\text{其中,精确度 (precision)} = \frac{TP}{TP+FP}, \text{召回率 (recall)} = \frac{TP}{TP+FN}, F\text{-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}, \text{准确度 (accuracy)} = \frac{TP+TN}{TP+TN+FP+FN}$$

由以上等式可知,精确度是指被分类器判定的正例中真正的正例样本的比重;召回率反映了被正确判定的正例占总的正例的比重;F-score 则是精确度和召回率的加权调和平均;准确率反映了分类器对整个样本的判定能力。由此可知,以上 4 个评估标准的值越高越好。

### 5.3 实验结果及讨论

为了公平地比较,我们将算法(word2vec 算法和 Location-weight 算法)的参数都初始化为固定值。设置算法都使用 Skip-gram 模型中的 Hierarchical Soft 的算法,窗口大小为 5 等。分别比较 Word2vec 算法、Location-weight 算法在 Gowalla 和 Brightkite 两个数据集上对预测用户间社交联系强度的效果。

图 3—图 6 分别显示了在 Gowalla 数据集上 4 个评价标准的分布情况。其中,每张图的横坐标表示取其中社交强度排在前面  $4.5 * 100000, 4.55 * 100000, \dots, 4.85 * 100000$  对用户为正样例,来求取算法的精确度、召回率、F-score 和准确度。由图可知,在取社交联系强度排在前  $4.85 * 100000$  对用户为正样例时,word2vec 的准确率为 83.33%; Location-weight 的准确率为 83.94%。Location-weight 的效果比 word2vec 的效果提高了大约 0.6%。通过计算,Location-weight 算法比 word2vec 算法将会多预测正确 5482 对社交联

系。由图可知,Location-weight 算法比 word2vec 算法效果好。Huy Pham 等人也在该数据集上得出了用户的社交联系强度,他们最好的是(精确度=80%,召回率=70%)或者(精确度=70%,召回率=82%)。而我们的实验结果最好的是(精确度=84%,召回率=84%)。综上所述,所提用户联系强度的预测模型可以显著提高预测准确率。

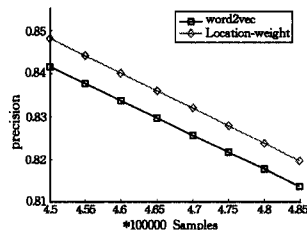


图 3 Gowalla 数据集上的精确度

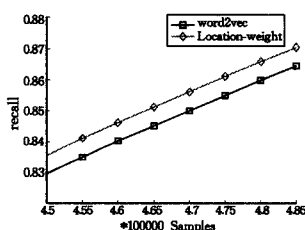


图 4 Gowalla 数据集上的召回率

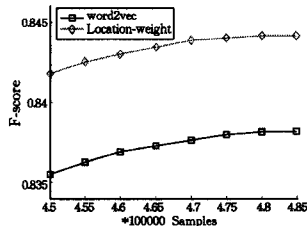


图 5 Gowalla 数据集上的 F-score

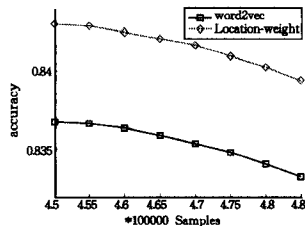


图 6 Gowalla 数据集上的准确度

图 7—图 10 分别显示了在 Brightkite 数据集上 4 个评价标准的分布情况。其中,每张图的横坐标表示取其中社交强度排在前面  $1.9 * 100000, 1.92 * 100000, \dots, 2.04 * 100000$  对用户为正样例,来求取算法的精确度、召回率、F-score 和准确度。据图可知,Location-weight 算法效果比 word2vec 算法效果好。



图 7 Brightkite 数据集上的精确度

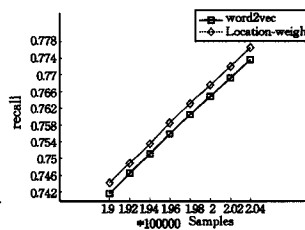


图 8 Brightkite 数据集上的召回率

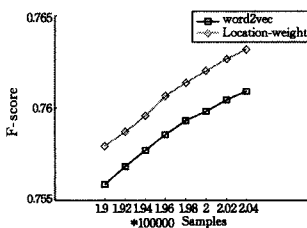


图 9 Brightkite 数据集上的 F-score

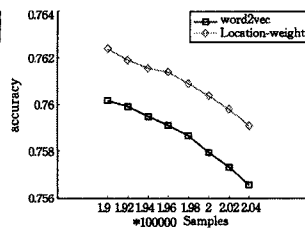


图 10 Brightkite 数据集上的准确度

通过以上实验,可以充分地认为,我们提出的 Location-weight 算法,即在 word2vec 学习过程中加入位置权重动态调整学习率的算法提高了用户社交联系强度预测的准确度。

**结束语** 本文分析了使用时空数据来预测用户间社交联系强度的方法。通过算法产生用户的特征向量,再用分类器

(下转第 274 页)

- [3] Hobbs J R. Granularity[C]//Proc. of the IJCAI, 1985;432-435
- [4] Zadeh L A. Fuzzy logic-computing with words[J]. IEEE Trans. on Fuzzy Systems, 1996, 4(2):103-111
- [5] Zadeh L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets and Systems, 1997, 90(2):111-127
- [6] Lin T Y. Granular computing on binary relations II: Rough set representations and belief functions[M]//Skowron A, Polkowski L, eds. Rough Sets in Knowledge Discovery. Heidelberg: Physica-Verlag, 1998;121-140
- [7] Gacek A. Signal processing and time series description: A Perspective of Computational Intelligence and Granular Computing [J]. Applied Soft Computing, 2015, 27:590-601
- [8] Li Jin-hai, Mei Chang-lin, Xu Wei-hua, et al. Concept learning via granular computing: A cognitive viewpoint[J]. Information Sciences, 2015, 298:447-467
- [9] Qian Yu-hua, Zhang Hu, Li Fei-jiang, et al. Set-based granular computing: A lattice model[J]. International Journal of Approximate Reasoning, 2014, 55(3):834-852
- [10] Chen Y M, Miao D Q, Wang R Z. A rough set approach to feature selection based on ant colony optimization[J]. Pattern Recognition Letters, 2010, 31(3):226-233
- [11] An J J, Wang G Y, et al. A rule generation algorithm based on granular computing[C]//2005 IEEE International Conference on Granular Computing. 2005;102-107
- [12] Zhang L, Zhang B. Fuzzy reasoning model under quotient space structure[J]. Information Sciences, 2005, 173(4):353-364
- [13] Zhu W, Wang F. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences, 2003, 152(1):217-230
- [14] Zadeh L A. Fuzzy sets and information granularity[M]//Gupta M, Ragade R, Yager R, eds. Advances in Fuzzy Set Theory and Applications. North-Holland, Amsterdam, 1979:3-18
- [15] Hu Q H, Yu D R, Xie Z X. Neighborhood classifiers[J]. Expert Systems with Applications, 2008, 34(2):866-876
- [16] Hu Q H, Pedrycz W, Yu D R, et al. Selecting discrete and continuous features based on neighborhood decision error minimization[J]. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 2010, 40(1):137-50
- [17] Liu Qing. Rough sets and rough reasoning [M]. Beijing: Science Press, 2001(in Chinese)  
刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001
- [18] Miao Duo-qian, Fan Shi-dong. The Calculation of Knowledge Granulation and Its Application [J]. Systems Engineering-theory & Practice, 2002, 22(1):48-56(in Chinese)  
苗夺谦, 范世栋. 知识的粒度计算及其应用[J]. 系统工程理论与实践, 2002, 22(1):48-56
- [19] Gou Guang-lei, Huang Li-feng, Ni wei. Conceptual Clustering Algorithm Based on Granular Computing [J]. Journal of Chongqing University of Technology (Natural Science), 2013, 27(6):76-79(in Chinese)  
苟光磊, 黄丽丰, 倪伟. 基于粒计算的概念聚类算法[J]. 重庆理工大学学报(自然科学), 2013, 27(6):76-79

(上接第 254 页)

(Random Forest)进行分类。由于时空数据的特殊性(共现的多样性和共现的偶然性),我们提出了在算法中加入位置权重动态调整学习率。实验结果表明 Location-weight 算法提高了用户社交联系强度预测的结果。进一步的研究工作将会围绕时空数据中共现的多样性对用户间社交联系强度预测的影响来展开。

## 参 考 文 献

- [1] Pham H, Shahabi C, Liu Y. Ebm: an entropy-based model to infer social strength from spatiotemporal data[C]//Proceedings of the 2013 International Conference on Management of Data. ACM, 2013:265-276
- [2] Gonzalez M C, Hidalgo C A, Barabasi A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196):779-782
- [3] Cheng Z, Caverlee J, Lee K, et al. Exploring Millions of Footprints in Location Sharing Services[C]//ICWSM. 2011:81-88
- [4] Lindqvist J, Cranshaw J, Wiese J, et al. I'm the mayor of my house: examining why people use foursquare—a social-driven location sharing application[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011:2409-2418
- [5] Kostakos V, Venkatanathan J, Reynolds B, et al. Who's your best friend-targeted privacy attacks In location-sharing social networks[C]//Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, 2011:177-186
- [6] Crandall D J, Backstrom L, Cosley D, et al. Inferring social ties from geographic coincidences[J]. Proceedings of the National Academy of Sciences, 2010, 107(52):22436-22441
- [7] Cranshaw J, Toch E, Hong J, et al. Bridging the gap between physical location and online social networks[C]//Proceedings of the 12th ACM International Conference on Ubiquitous Computing. ACM, 2010:119-128
- [8] Li Q, Zheng Y, Xie X, et al. Mining user similarity based on location history[C]//Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2008:34
- [9] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013
- [10] Li Min, Wang Xiao-cong, Zhang Jun, et al. Study on Check-in and Related Behaviors of Location-based Social Network Users [J]. Computer Science, 2013, 40(10):72-76(in Chinese)  
李敏, 王晓聪, 张军, 等. 基于位置的社交网络用户签到及相关行为研究[J]. 计算机科学, 2013, 40(10):72-76
- [11] Mikolov T, Yih W, Zweig G. Linguistic Regularities in Continuous Space Word Representations [C] // HLT-NAACL. 2013:746-751
- [12] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011:5528-5531