



文章编号:1001-9081(2017)11-3053-04

DOI:10.11772/j.issn.1001-9081.2017.11.3053

基于卷积神经网络的谣言检测

刘政^{1,2}, 卫志华^{1,2*}, 张韧弦^{1,2}

(1. 同济大学 计算机科学与技术系, 上海 201804; 2. 嵌入式系统与服务计算教育部重点实验室(同济大学), 上海 201804)

(*通信作者电子邮箱 zhihua_wei@tongji.edu.cn)

摘要:人工检测谣言通常需要耗费大量的人力物力,并且会有很长的检测延迟。目前现存的谣言检测模型一般根据谣言的内容、用户属性、传播方式人工地构造特征,而人工构建特征存在考虑片面、浪费人力等现象。为了解决这个问题,提出了基于卷积神经网络(CNN)的谣言检测模型。将微博中的谣言事件向量化,通过卷积神经网络隐含层的学习训练来挖掘表示文本深层的特征,避免了特征构建的问题,并能发现那些不容易被人发现的特征,从而产生更好的效果。实验结果表明,所提方法能够准确识别谣言事件,在准确率、精确率与F1值指标上优于支持向量机(SVM)与循环神经网络(RNN)等对比算法。

关键词:微博;谣言检测;谣言事件;卷积神经网络

中图分类号:TP391.41 **文献标志码:**A

Rumor detection based on convolutional neural network

LIU Zheng^{1,2}, WEI Zhihua^{1,2*}, ZHANG Renxian^{1,2}

(1. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China;

2. Key Laboratory of Embedded System and Service Computing of Ministry of Education (Tongji University), Shanghai 201804, China)

Abstract: Manual rumor detection often consumes a lot of manpower and material resources, and there will be a long detection delay. At present, the existing rumor detection models construct features manually according to the content, user attributes, and pattern of the rumor transmission, which can not avoid one-sided consideration, waste of human and other phenomena. To solve this problem, a rumor detection model based on Convolutional Neural Network (CNN) was presented. The rumor events in microblog were vectorized. The deep features of text were mined through the learning and training in hidden layer of CNN to avoid the problem of feature construction, and those features that were not easily found could be found to produce better results. The experimental results show that the proposed method can accurately identify rumor events, and it is better than Support Vector Machine (SVM), Recurrent Neural Network (RNN) and other contrast algorithms in accuracy rate, precision rate and F1 score.

Key words: microblog; rumor detection; rumor event; Convolution Neural Network (CNN)

0 引言

随着在线社交媒体的迅速发展,大量不可靠的信息得以快速和广泛地在人群中传播。社交媒体上谣言泛滥可能导致人们难以从纷繁的信息中甄别得到可信的信息,进而影响人们正常的生活秩序。特别是在面临突发公共事件(自然灾害、事故灾难、公共卫生事件、社会安全事件、经济危机等)时,广泛传播的谣言可能会具有极大的破坏性。在微博中,信息内容主要通过人与人之间建立的“关注-被关注”网络进行传播。人与人之间的互联、人与信息之间的互联高度融合,人人参与到信息的产生与传播过程,这种传播方式使得一条信息能够在短时间内传播到数百万计的用户^[1]。因此,自动高效地识别社交媒体中的谣言意义重大。

社交媒体的自动谣言检测是基于推特信息的可靠性检测而来的^[2],Kwon等^[3]介绍了一个基于时间属性的时间序列适应模型;Ma等^[4]使用动态时间序列扩展了该模型;Zhao

等^[5]通过使用线索词如“not true”“unconfirmed”等来进行早期的谣言检测;Ma等^[6]提出了利用循环神经网络(Recurrent Neural Network, RNN)进行谣言检测,实验结果表明RNN方法优于现有的基于人工构造特征的谣言检测模型。

近年来,卷积神经网络(Convolutional Neural Network, CNN)在图像处理、目标检测、图像语义分割等领域取得了一系列突破性的研究成果,其强大的特征学习与分类能力引起了广泛关注,具有重要的分析与研究价值。早期的CNN结构相对简单,主要应用在手写字符识别、图像分类等相对单一的计算机视觉应用领域中^[7]。随着研究的不断深入,CNN的结构不断优化,其应用领域也逐渐得到延伸。

2012年,Krizhevsky等^[8]提出的AlexNet在大型图像数据库ImageNet^[9]的图像分类竞赛中以准确度超越第二名11%的巨大优势夺得了冠军,使得CNN成为了学术界的焦点;并且,CNN不断与一些传统算法相融合,加上迁移学习方法的

收稿日期:2017-05-16;修回日期:2017-06-05。基金项目:国家自然科学基金资助项目(61573259, 61673301, 61573255, 61673299);上海市中医药三年行动计划重点项目(ZY3-CCCX-3-6002);上海自然科学基金资助项目(15ZR1443800)。

作者简介:刘政(1992—),男,山东济南人,硕士研究生,CCF会员,主要研究方向:自然语言处理;卫志华(1979—),女,山西晋中人,副教授,博士,CCF会员,主要研究方向:机器学习、文本挖掘、图像内容分析;张韧弦(1976—),男,浙江鄞县人,副教授,博士,主要研究方向:自然语言处理、社交网络、数据挖掘。



引入,使得 CNN 的应用领域获得了快速的扩展。一些典型的应用包括:CNN 与 RNN 结合图像的摘要生成^[10-11];自然语言处理方面,卷积神经网络模型在语义分析^[12]、搜索结果提取^[13]、句子分类^[14]、句子建模^[15]、句子预测^[16]和其他的传统自然语言处理任务^[17]中都取得了很好的结果。

由于 CNN 的良好表现,所以本文应用于微博自动谣言检测当中。本文对 Kim 等^[14]提出的句子分类模型进行改进,使其适应于微博谣言检测,实验表明本文模型的有效性。

本文的主要贡献如下:

- 1) 提出了一种基于卷积神经网络模型的谣言检测方法,将谣言事件向量化,通过自动构建谣言的特征来进行模型训练,从而达到判断谣言的目的;
- 2) 提出了一种基于微博的谣言检测框架,实现微博中的谣言检测。

1 本文模型

1.1 基本定义

本文的研究对象为谣言事件。所谓谣言事件,就新浪微博而言,本文需要关心的是某个最初始的不实微博(即源头微博)是否是谣言,并不关心其传播过程中所衍生出来的微博是否存在不实,或者是否是谣言。最初始微博一经判断,则与其相关的微博也自然得到判断。例如,“传说中麦当劳全国通用无线上网密码,真的吗?”这条谣言以及其相关微博“真的么?”“假的”“存了多少次都找不到了”等就构成了一个谣言事件,而本文判断的仅是“传说中麦当劳全国通用无线上网密码,真的吗?”这条微博是否是谣言,而其相关微博是否是谣言并不关心,或者说其相关微博是作为判断源头微博是否是谣言的证据存在。因此,在详细描述本文的模型之前,本文先给出一些基本的符号解释和定义。

定义 1 谣言事件^[6]。定义一个所有事件的集合 $E = \{E_i\}$,

传说中麦当劳全国通用
无线上网密码,真的吗?
真的么?
假的
存了多少次都找不到了
转发能用
试试看
你们再转 该换密码了
有机会去M试试
真的假的??肯德基的呢?

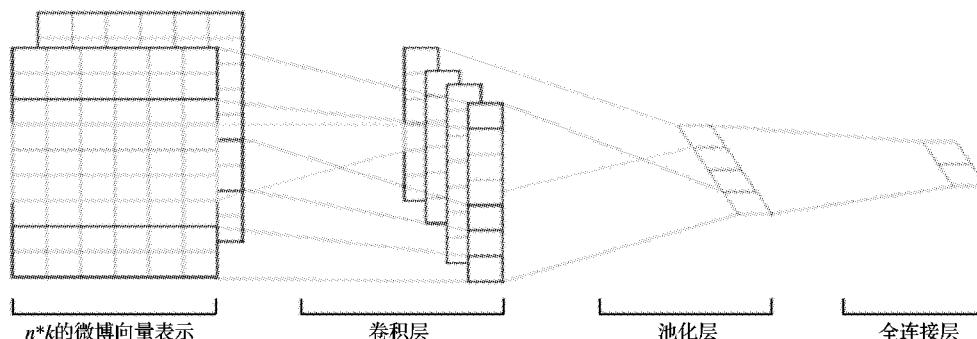


图 2 基于卷积神经网络的谣言检测模型

Fig. 2 Rumor detection model based on convolutional neural network

令 $m_{ij} \in \mathbf{R}^k$,其中 m_{ij} 表示某个谣言事件 E_i 的某条相关微博,用 k 维的句子向量表示。一个包含 n 条相关微博的谣言事件可以表示为:

$$E_i = m_{i1} \oplus m_{i2} \oplus \cdots \oplus m_{in}$$

其中 \oplus 表示串联操作。

卷积层进行卷积操作,卷积操作就是利用 filter $w \in \mathbf{R}^{h \times k}$ 来产生新的特征。例如事件 i 中第 j 条相关微博到第 $j+h-1$ 条微博所产生的特征 a_i 的计算公式如下:

$$a_i = f(w \cdot m_{j:(j+h-1)} + w_b)$$

其中: w_b 是 filter 的偏置项, f 是一个非线性函数,比如 ReLU

$\{E_i\}$,其中每一个 $E_i = \{m_{ij}\}$ 包含与其相关的所有微博, m_{ij} 表示谣言事件的某一条微博,本文的任务就是判断 E_i 是不是谣言事件。

表 1 标记与定义

Tab. 1 Notations and definitions

标记	定义
E	所有事件,包括谣言事件与非谣言事件
E_i	事件 i
m_{ij}	事件 i 的某一条相关微博 j
w	过滤器 filter
w_b	filter 的偏置项
a	特征矩阵

1.2 模型结构

本文的谣言检测的流程框图如图 1 所示。首先,通过新浪微博官方的辟谣平台——微博社区管理中心取得某条具体的谣言;然后,对与其相关的微博进行聚类,得到 N 个谣言事件,相应地,为了便于分类训练,取得与谣言事件数目近似相同的 M 个非谣言事件,将每一个谣言事件作为一个整体,对其中的每一条微博向量化,组成一个输入矩阵,依此类推,向量化完成之后,作为输入矩阵,进行卷积神经网络模型的训练。

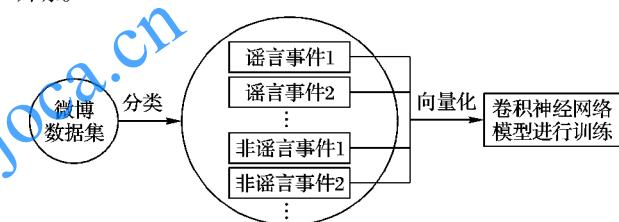


图 1 微博谣言检测框架

Fig. 1 Microblog rumor detection framework

用于谣言检测的卷积神经网络如图 2 所示。

激活函数。filter 应用于事件 i 的每一个窗口 $\{m_{i1:ih}, m_{i2:(i+1)h}, \dots, m_{i(n-h+1):nh}\}$ 来生成一个特征矩阵:

$$a = [a_1, a_2, \dots, a_{n-h+1}]$$

池化层进行池化操作。通常在卷积层之后会得到维度很大的特征,将特征切成几个区域,取其最大值或平均值,得到新的、维度较小的特征。主要作用是下采样,通过去掉特征矩阵中不重要的样本来减少参数的数量。本文采用的是 max pooling, 即对特征矩阵 a 中的每一个 a_i 取最大值,作为采样后的样本值。

最后将池化后的特征矩阵传入全连接层,进行 Softmax 操作。



作,最后输出分类的概率。

本文模型采用了多个 filter 来获得多种特征。值得注意的是,传统的卷积神经网络模型 filter 遍历每一种图片的所有局部块如图 3(a)所示。而在本文模型之中,filter 遍历每个谣言事件的长度为 h 的所有窗口,而 filter 宽度通常等于输入矩阵的宽度,如图 3(b)所示。这样设置的理由如下:因为本文模型的输入矩阵每一行表示事件的某一条微博的微博向量,将微博向量当作最小单位,有助于更好地挖掘相关微博之间的特征,以及谣言的传播特征;反之,若是 filter 扫描局部块,则更注重的是谣言事件中所有词之间的关系,不能从整体上把握谣言传播的时序特点。因为 Kim^[14]的工作表明,对于文本分类任务,含有一层卷积层的卷积神经网络已经表现得很好,所以本文模型仅采用一个卷积层与一个池化层。

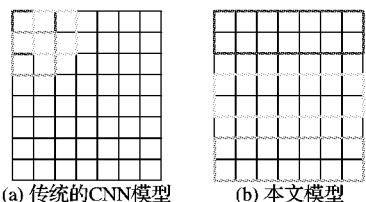


图 3 两种模型过滤器遍历输入矩阵的方式

Fig. 3 Way of filter traversing input matrix in two models

2 实验与分析

2.1 数据收集

通过研究新浪微博发现数据获取的方式可以通过以下步骤:

1)首先在微博社区管理中心,找得到不实信息的微博,如图 4 所示。



图 4 微博社区管理中心

Fig. 4 Microblog community management center

2)根据不实微博的公示信息,获得原文,如图 5 所示。

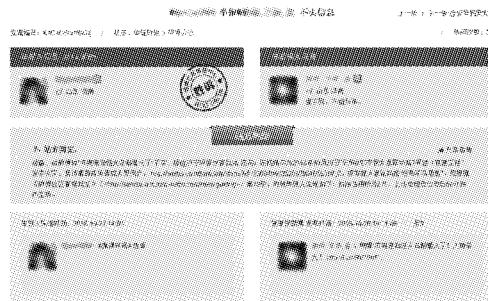


图 5 某条谣言微博的公示信息

Fig. 5 Public information of a rumor microblog

3)根据原文信息,进行具体谣言事件的相关微博信息的抓取。

为了便于本文模型与其他模型的比较,本文采用 Ma

等^[6]公开的数据集。该数据集包含 2313 个谣言事件与 2351 个非谣言事件,共包含 3805 656 条微博。由于微博本身含有很大的噪声,包含@ 某某人、超链接等无关信息,所以本文通过正则表达式匹配的方法对该数据集进行了降噪等预处理。

2.2 实验设置

本文将本文模型与 Ma 等^[6]的工作进行了对比:

Ma 等的模型 Ma 等提出了采用循环神经网络模型来进行微博谣言的检测,分别实现了 tanh-RNN (tanh-Recurrent Neural Network)、LSTM (Long Short-Term Memory)、GRU (Gated Recurrent Unit) 等模型,并取得了良好的效果。此外, Ma 等基于本文所采用的数据集,重现了之前较经典的谣言检测的工作,便于进行模型的比较。

本文模型 本文的卷积神经网络模型,包含一个卷积层和一个池化层,且输入矩阵是使用 Doc2Vec 训练好的向量矩阵。

超参数设置 本文采用 ReLU 激活函数,filter 窗口高度分别采用 3、4、5,dropout rate 为 0.5。每一条微博的维度 k 设置为 50,由于不同的谣言事件或者非谣言事件含有的微博数目不同,所以本文对于微博数目小于最大微博数的事件,采用补充零向量的方式,使输入矩阵的大小一致。

预处理的句子向量 对于每一条微博,本文进行预处理之后,进行分词,之后利用 Doc2Vec 方法,将微博转化为句子向量。

评价指标 模型的评价指标采用自然语言处理中常用的准确率(accuracy)、召回率(recall)、精确率(precision),以及 F1 值,它们针对谣言的定义如下:

$$\text{准确率}(\text{accuracy}) = \frac{\text{系统正确识别的谣言与非谣言数目}}{\text{标准测试集中谣言与非谣言数目}}$$

$$\text{召回率}(\text{recall}) = \frac{\text{系统正确识别的谣言的数目}}{\text{标准测试集中谣言的数目}}$$

$$\text{精确率}(\text{precision}) = \frac{\text{系统正确识别的谣言的数目}}{\text{系统识别的谣言的数目}}$$

$$F_1 = \frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$$

本文的卷积神经网络是通过 TensorFlow 实现。为了便于与 Ma 等的模型作对比,本文同样地分别选取 10% 的谣言事件与非谣言事件用于模型的调整;将剩下的数据的 30% 作为训练集,10% 作为测试集。

2.3 实验结果

本文的模型 CNN 训练集和测试集比例为 3:1,与其他模型的实验结果比较如表 2 所示。其中,“-1”“-2”表示隐含层数,“-1”表示一层,“-2”表示两层。

表 2 CNN 模型与其他模型的实验结果比较

Tab. 2 Result comparison of CNN model with other models

模型	accuracy	precision	recall	F_1
DT-Rank	0.732	0.738	0.715	0.726
SVM-TS	0.857	0.839	0.885	0.861
tanh-RNN	0.873	0.816	0.964	0.884
LSTM-1	0.896	0.846	0.968	0.913
GRU-1	0.908	0.871	0.958	0.913
GRU-2	0.910	0.876	0.956	0.914
CNN(3:1)	0.966	0.978	0.951	0.964

从表 2 中可以看出:就人工构造特征与模型自动构造特征而言,人工构造特征中基于时间序列的支持向量机模型



(Support Vector Machine based on Time-Series, SVM-TS) 优于决策树模型 DT-Rank, 而本文模型与支持向量机模型相比, 准确率提高了 10.9 个百分点, 精确率提高了 13.9 个百分点, 召回率提高了 6.6 个百分点, F_1 值提高了 10.3 个百分点。可以得出模型自动构造特征明显优于人工构造的特征。

就卷积神经网络模型与循环神经网络模型而言, 可以看出, 循环神经网络模型中, GRU-2 模型优于其他模型, 而本文的模型与 GRU-2 模型相比, 准确率提高了 5.6 个百分点, 精确率提高了 10.2 个百分点, 召回率与 LSTM-1 模型相比降低了 1.7 个百分点, F_1 值提高了 5 个百分点。可以看出除了召回率略低以外, 准确率、精确率和 F_1 值都优于循环神经网络 (RNN) 模型。所以本文认为在谣言检测当中, 卷积神经网络模型是通过谣言事件中发现微博之间的关系来构造特征, 而循环神经网络则是通过谣言事件中所有词之间的关系来构造特征, 对于谣言来说, 由于其传播方式等的特点, 本文认为按照相关微博来构造特征比按照词来构造特征更合理, 所以认为卷积神经网络模型优于循环神经网络模型。

综上所述, 本文认为就谣言检测而言, 模型自构建特征优于人工构建特征, 卷积神经网络模型优于循环神经网络模型。

此外, 本文将训练集与测试集按照 9:1 的分割比例进行了同样的实验, 如表 3 所示。从表 3 中可以看出按照 9:1 的分割比例时, 训练的模型实验结果略好于按照 3:1 进行分割的模型。

表 3 CNN 模型不同分割比例之间的结果比较

方法	accuracy	precision	recall	F_1
CNN(3:1)	0.966	0.978	0.951	0.964
CNN(9:1)	0.970	0.983	0.958	0.970

本文通过刘知远等^[18]的工作了解到, 新浪微博中一般人工检测并确认谣言的时间周期大约在一周左右, 图 6 是本文的 CNN 模型训练过程中的收敛情况, 其中实线表示在训练集中的收敛情况, 点线表示在测试集上的收敛情况。从图中可以看出模型收敛的速度非常快, 当训练到达 10000 步时, 模型的准确率就已经达到了 94.47%, 这也正好弥补了人工检测微博谣言存在较长时间的延迟问题。

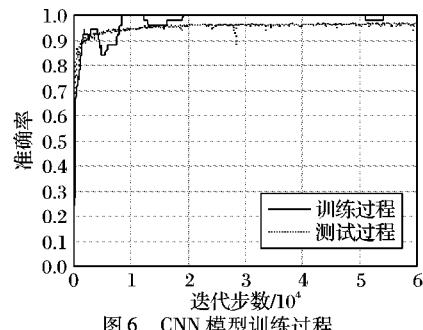


图 6 CNN 模型训练过程

Fig. 6 CNN model training process

3 结语

目前在大多数谣言检测任务中, 都是采用人工构建特征的方式, 本文提出将卷积神经网络模型应用于谣言检测中, 从实验中得出本文的模型优于目前最好的谣言检测方法, 准确率及 F_1 值均高出 5% 左右。如何将谣言事件下的用户评价信

息加入, 以及如何进行卷积神经网络的层数的选取, 将是接下来进一步的研究工作。

参考文献 (References)

- [1] 李洋, 陈毅恒, 刘挺. 微博信息传播预测研究综述 [J]. 软件学报, 2016, 27(2): 247–263. (LI Y, CHEN Y H, LIU T. Survey on predicting information propagation in microblogs [J]. Journal of Software, 2016, 27(2): 247–263)
- [2] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675–684.
- [3] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media [C]// Proceedings of the 2013 IEEE 13th International Conference on Data Mining. Piscataway, NJ: IEEE, 2013: 1103–1108.
- [4] MA J, GAO W, WEI Z, et al. Detect rumors using time series of social context information on microblogging websites [C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York: ACM, 2015: 1751–1754.
- [5] ZHAO Z, RESNICK P, MEI Q. Enquiring minds: early detection of rumors in social media from enquiry posts [C]// Proceedings of the 24th International Conference on World Wide Web. New York: ACM, 2015: 1395–1405.
- [6] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2016: 3818–3824.
- [7] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述 [J]. 计算机应用, 2016, 36(9): 2508–2515. (LI Y D, HAO Z B, LEI H. Survey of convolutional neural network [J]. Journal of Computer Applications, 2016, 36(9): 2508–2515.)
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1097–1105.
- [9] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248–255.
- [10] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 2625–2634.
- [11] VINYALS O, TOSHEV A, BENCI S, et al. Show and tell: a neural image caption generator [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3156–3164.
- [12] YIH W, HE X, MEEK C. Semantic parsing for single-relation question answering [C]// Proceedings of the 2014 Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2014: 643–648.
- [13] SHEN Y, HE X, GAO J, et al. Learning semantic representations using convolutional neural networks for Web search [C]// Proceedings of the 23rd International Conference on World Wide Web. New York: ACM, 2014: 373–374.

(下转第 3100 页)



- tions, 2012, 32(5): 1439 – 1442.)
- [14] 郭崇. 基于大数据分析的网购客户忠诚度建模仿真[J]. 计算机仿真, 2015, 32(10): 239 – 242. (GUO C. Modeling and simulation of online customer loyalty based on analysis of big data [J]. Computer Simulation, 2015, 32(10): 239 – 242.)
- [15] 谷斌, 徐菁. 基于知识共享的专业虚拟社区用户忠诚度挖掘[J]. 情报科学, 2015, 33(1): 105 – 110. (GU B, XU J. Mining of professional virtual community user' loyalty based on knowledge sharing [J]. Information Science, 2015, 33(1): 105 – 110.)
- [16] 唐胡鑫. 电子商务客户忠诚度模型仿真研究[J]. 计算机仿真, 2016, 33(1): 413 – 415. (TANG H X. Simulation research on loyalty model on e-commerce customer [J]. Computer Simulation, 2016, 33(1): 413 – 415.)
- [17] KREUZER S, HOEBEL N. Fuzzy clustering of Web user profiles for analyzing their behavior and interests[J]. Fuzzy Methods for Customer Relationship Management and Marketing: Applications and Classifications, 2012, 49: 307 – 313.
- [18] SANCHEZ-FRANCO M J, RONDAN-CATALUNA F J. Virtual travel communities and customer loyalty: customer purchase involvement and Web site design[J]. Electronic Commerce Research and Applications, 2009, 9(2): 16 – 22.
- [19] GUO Y, NIU D. An analysis model of power customer satisfaction based on the decision tree[J]. International Journal of Business and Management, 2009, 2(3): 30 – 35.
- [20] MIGLAUTSCH J R. Thoughts on RFM scoring[J]. Journal of Database Marketing & Customer Strategy Management, 2000, 8(1): 67 – 72.
- [21] GRIFFIN J. Customer Loyalty: How to Earn and How to Keep it [M]. New York: Jossey-Bass Inc., 1995: 32 – 38.
- [22] GOLDIN D G, KANELLAKIS P C. On similarity queries for time-series data: constraint specification and implementation[C]// Proceedings of the 1st International Conference on Principles and Practice of Constraint Programming. London: Springer-Verlag, 1995: 137 – 153.
- [23] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics Theory and Experiment, 2008, 10: 155 – 168.
- [24] AL-GARADI M, VARATHAN K D, RAVANA S D. Identification of influential spreaders in online social networks using interaction weighted K-core decomposition method[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 468: 80 – 89.
- [25] LI Q, ZHOU T, LYU L, et al. Identifying influential spreaders by weighted LeaderRank[J]. Physica A: Statistical Mechanics and its Applications, 2014, 404(24): 50 – 55.
- [26] 马寅. 社会网络影响力最大化算法及传播模型的研究[D]. 兰州: 兰州大学, 2012: 20 – 23. (MA Y. Research on the maximum algorithm and propagation model of social network influence [D]. Lanzhou: Lanzhou University, 2012: 20 – 23.)
- [27] 黎雷. 社会网络影响力模型及其算法研究[D]. 北京: 北京交通大学, 2010: 30 – 35. (LI L. Research on social network influence model and its algorithm [D]. Beijing: Beijing Jiaotong University, 2010: 30 – 35.)

This work is partially supported by the National High Technology Research and Development Program (863 Program) of China (2009AA062802), the National Natural Science Foundation of China (60473125, 71572015), the CNPC Innovation Foundation (05E7013), the National Science and Technology Major Project (G5800-08-ZS-WX), the Science Foundation of China University of Petroleum-Beijing at Karamay (RCYJ2015B-03-001).

XUE Yun, born in 1977, Ph. D. candidate, associate professor. Her research interests include data mining, knowledge discovery.

LI Guohe, born in 1965, Ph. D., professor. His research interests include artificial intelligence, machine learning, knowledge discovery.

WU Weijiang, born in 1971, Ph. D. candidate, associate professor. His research interests include artificial intelligence, knowledge discovery.

HONG Yunfeng, born in 1965. His research interests include enterprise resource planning, data management.

ZHOU Xiaoming, born in 1965, senior engineer. His research interests include information management system, decision support.

(上接第 3056 页)

- [14] KIM Y. Convolutional neural networks for sentence classification [EB/OL]. [2016-11-20]. <http://www.aclweb.org/anthology/D14-1181.pdf>.
- [15] KALCHBRENNER N, GREFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences [EB/OL]. [2016-11-20]. <http://anthology.aclweb.org/P/P14/P14-1062.pdf>.
- [16] COLLOBERT R, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning [C]// Proceedings of the 25th International Conference on Machine Learning. New York: ACM, 2008: 160 – 167.
- [17] COLLOBERT R, WESTON J, KARLEN M, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12(1): 2493 – 2537.
- [18] 刘知远, 张乐, 涂存超, 等. 中文社交媒体谣言统计语义分析 [J]. 中国科学: 信息科学, 2015, 45(12): 1536 – 1546. (LIU Z Y, ZHANG L, TU C C, et al. Statistical and semantic analysis

of rumors in Chinese social media [J]. Scientia China: Informationis Sciences, 2015, 45(12): 1536 – 1546.)

This work is partially supported by the National Natural Science Foundation of China (61573259, 61673301, 61573255, 61673299), the Program of Further Accelerating the Development of Chinese Medicine Three Year Action of Shanghai (ZY3-CCCX-3-6002), the Natural Science Foundation of Shanghai (15ZR1443800).

LIU Zheng, born in 1992, M. S. candidate. His research interests include natural language processing.

WEI Zhihua, born in 1979, Ph. D., associate professor. Her research interests include machine learning, text mining, image content analysis.

ZHANG Renxian, born in 1976, Ph. D., associate professor. His research interests include natural language processing, social networking, data mining.