

开源交通信息获取及分析系统

吕宜生^{1,2}, 陈圆圆^{1,3}, 赵学亮^{1,3}, 王晓^{1,2}, 陈虹宇¹, 刘玉强⁴, 王飞跃^{1,5}

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京, 100190

2. 青岛智能产业技术研究院, 青岛, 266109

3. 中国科学院大学, 北京, 100049

4. 中国铁道科学研究院, 北京, 100081

5. 国防科技大学军事计算实验与平行系统技术中心, 长沙, 410073)

摘要: 随着云计算、大数据、移动互联网的发展, 社交媒体、网站、在线地图等网络媒介将成为获取交通信息的重要媒介。网络中的交通信息与交通物理世界中采集到的交通信息互相补充、交叉验证, 丰富了交通信息获取内容。本文介绍了一个基于互联网内容的开源交通信息采集及分析系统。该系统可采集专题网站、论坛、微博等含有的指定区域内交通内容, 主要利用中文自然语言处理技术和机器学习方法从文本中挖掘交通事件信息, 分析交通舆情、交通行为, 预测未来交通状况。

关键词: 社会交通, 交通舆情, 社交媒体, 交通事件, 自然语言处理

An Online Open Source Traffic Information Acquisition and Analytic System

Yisheng Lv^{1,2}, Yuanyuan Chen^{1,3}, Xueliang Zhao^{1,3}, Xiao Wang^{1,2}, Hongyu Chen¹, Yuqiang Liu⁴, Fei-Yue Wang^{1,5}

(1. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Science, Beijing, 100190

2. Qingdao Academy of Intelligent Industries Qingdao, Shandong, 266109

3. University of Chinese Academy of Sciences, Beijing, 100049

4. China Academy of Railway Sciences, Beijing, 100081

5. Research Center of Military Computational Experiments and Parallel Systems, National University of Defense Technology, Changsha, Hunan 410073)

Abstract Internet sources like social media, websites, and online map service have become important traffic data sources with the dramatic development of cloud computing, big data, and mobile internets. Traffic information data collected with online sources can be complementary to physical world traffic information and help enrich traffic contents. In this paper, we introduce an online open source traffic information acquisition and analytic system. The proposed system can collect traffic information in specified area from specific websites, forums, and weibo. It mainly uses natural language processing techniques and machine learning methods to mine traffic events from texts, analyze traffic sentiment, infer traffic behavior, and make traffic prediction.

Keywords Social Transportation, Traffic Sentiment, Social Media, Traffic Event, Natural Language Processing

1. 引言

基金项目: 国家自然科学基金 71232006、61533019 和中国铁路总公司科技研究开发计划课题 J2016X011.

作者简介: 吕宜生, 研究方向为社会交通、平行交通、人工智能, Email: yisheng.lv@ia.ac.cn.

随着机动车保有量的持续增长, 城市交通拥堵、交通污染、交通事故问题日益加剧。智能交通系统被认为是解决城市交通问题、改善城市交通运行环境的有效手段, 而智能交通系统的运行需依赖实时、准确的交通信息。交通管理部门可以依据实时、准确的交通信息更合理提供交通服务, 如交通控制、

公交车调度、出行信息服务等，从而提高道路交通资源利用率，提高交通系统运行效率；交通出行者可以依据实时、准确的交通信息更合理的在出行前安排出行计划（包括出行时间、出行方式、出行路径和出行目的地）、在出行中调整行驶路线等，从而实现快捷、方便出行[1-3]。

我们知道，交通物理世界中所使用的交通信息采集手段主要包括感应线圈、微波、雷达、GPS 浮动车、摄像头等。伴随着社交网络、移动互联网的发展，人作为最灵敏的交通社会传感器，可提供大量用户产生的、自愿发布分享的实时交通信息。网络社交媒体（social media）正逐渐成为重要的交通信息来源[4-7]。网络社交媒体中蕴含的交通信息与交通物理世界中采集到的交通信息互相补充、交叉验证，丰富了交通信息获取内容。

事实上，除了社交媒体（weibo、twitter、微信）外，政府交通信息发布网站、地图服务提供商（bingmap、google map、baidu map、高德 map）、大型活动发布网站、交通天气预报网站等互联网站点也蕴含着与交通相关的开放数据信息。所获取的与交通相关的开放信息包括：交通事件信息、交通事故信息、交通拥挤信息、交通管制信息、大型活动信息、天气信息、假日信息等。

本文描述了一个基于互联网内容的开源交通信息采集及分析系统。通过网络爬虫或网站提供的 API 接口采集专题网站、论坛、微博、微信中含有的指定区域内交通内容；利用中文自然语言处理技术对抓取获得的文本进行处理；利用机器学习方法挖掘、分析交通信息，主要包括交通事件检测、交通舆情分析、交通行为推断以及交通预测。

2. 国内外相关工作介绍

快速发展的互联网和社交网络数据已被应用于多个领域，比如健康领域、消费领域、公共安全领域等[8-11]。在交通领域，社交媒体数据主要用于交通信息采集、交通预测、交通行为分析、交通舆情分析[12-14]。

当前，互联网已成为人们发布、交流、获取信息的重要渠道。互联网上交通信息的来源包括交通信息发布网站（官方、非官方）、

社交媒体（weibo、twitter）（包括政府部门官方账号、个体账号、其他公众账号等）、专业论坛、在线地图服务提供商（bing map、高德 map）、活动发布/售票网站、天气预报网站等。

Morgul et al. 提出了虚拟检测器获取交通数据方法，即：利用互联网地图服务商 Bing Maps 和 MapQuest 提供的 API 接口，通过设置虚拟检测器，提取实时交通数据；通过对比此方法获得的交通数据和利用线圈检测器及电子标签得到的数据，表明虚拟检测方法得到的交通数据是有效的[15]。进而，Morgul et al. 对虚拟检测器方法进行了扩展，利用 tweet 数据实时获取事件信息[16]。

He et al. 利用 tweet 数据预测未来较长时交通流量。首先，他们进行了交通流量和 tweet 数之间的相关性分析，进而基于 tweet 语义提取了交通指标，构建了以交通流量和 tweet 语义为自变量的线性回归模型预测未来交通流量[17]。Ni et al. 利用 tweet 数据研究了特殊事件下的短时交通流预测问题。考察了在同时考虑交通特征和 tweet 特征、仅考虑交通特征下自回归模型、神经网络、支持向量回归、k-近邻方法四种预测方法的性能，实验表明当同时考虑交通特征和 tweet 特征下的预测效果明显好于仅考虑交通特征时[18]。Grosenick 利用 tweet 数据挖掘交通事件，同时结合实际检测数据，利用人工神经网络预测道路交通速度[19]。Zhou et al. 利用分层贝叶斯网预测道路交通状态，该方法将交通事件、天气、交通状态整合在一起，其中交通事件和天气信息由网络获取，交通状态通过由 RFID 采集[20]。考虑到可以容易地从社交网络中提取人类活动和社会事件信息，Pathania 和 Karlapalem 提出在模拟城市交通系统时，要考虑人类活动和社会事件，以刻画现实交通行为[21]。Chen 等通过收集网络开源数据，利用 LSTM 模型预测交通拥挤状况[22]。

从微博内容中检测交通事件，是社交媒体在交通领域中研究较早和较广泛的一个方向。在微博中搜索城市交通有关的内容，很容易的会搜到诸如堵车、交通事故、道路施工等交通事件信息，比如：“上班的路上

又堵了, ……”, “去单位得(的)路上他妹的又堵了”等。熊佳茜利用基于条件随机场概率模型 CRF 和基于规则的正则表达式相结合的办法提取微博文本中的交通事件要素[23]。崔健等开发了 1 套基于微博的节假日突发交通事件提取与分析系统,旨在抓取并提取出交通主题相关微博[24]。Gutierrez et al. 从英国交通管理部门发布的 tweet 中提取交通事件[25]。Eleonora D’Andrea 等提出了一个利用 tweet 数据的交通事件实时监控系统,并声称该系统检测到的交通事件时间通常早于交通网站发布的时间[26]。

Maghrebi 等分析了人们在墨尔本都市区发布的 Twitter 数据内容,从中提取出行者的出行方式,该信息可作为传统交通调查的补充[27]。Zhang 等利用 Twitter 数据分析了出行行为特征,并给出了异常移动背后的社会出行动机[28]。Hasan 等利用从社交媒体应用中获取的位置数据,分析了人类活动和移动模式[29]。

交通与人们的生活息息相关,交通行业一直是群众舆论关注的热点。Zeng et al. 分析了国内黄金周期间的交通舆情[4]。Cao et al. 通过获取网站和社交媒体信息,进行交通舆情分析[5]。Semwal et al. 利用社交媒体进行交通事件检测、舆情分析、交通问题根源分析以及解决建议[30]。

在系统应用集成及展示方面, Freddy Lécué et al. 提出了 STAR-CITY (Semantic Traffic Analytics and Reasoning for CITY), 该系统集成了大量结构化、非结构化数据,静态和流式数据,可以对城市交通情景进行分析、诊断、推理和预测如:交通状况时空分析、公交车拥挤诊断、道路拥挤程度预测[31-34]。Ben Sharon Rose Ben Jacob Singh 开发了利用开放数据和 tweet 数据的伦敦交通状况展示系统,该系统所利用的数据包括 Twitter 中的 Twitter TFL Traffic News Profile、TFL Traffic Syndicated Feeds、以及 google 交通信息[35]。

3. 系统描述

本节描述基于网络开源数据的交通信息获取及分析系统,系统结构如图 1 所示。

开源交通信息获取及分析系统结构主要包含四部分:(1)数据采集,(2)预处理,(3)信息建模与处理,(4)系统应用。

数据采集。数据采集模块用于采集网络上与交通相关的开源数据。通过网站提供的 API 接口或网络爬虫采集新浪微博、气象网站、bbs、论坛、在线地图、签到网站、网页中含有的指定区域和时间内具有指定关键词的交通内容原始数据,主要以文本形式存在。目前,该系统主要从新浪微博中提取交通事件,从气象网站和微博气象公共账号中提取天气信息,从在线地图中提取交通状态信息,从微博、bbs、论坛网页中获取分析交通舆情的数据。有关网络数据采集方法的介绍请见第 4 节。

预处理。在这里,主要是针对文本进行预处理。通过对文本预处理,得到标准化的文本,以便于信息抽取。预处理模块主要包括:(1)中文分词。分词一般是自然语言处理的第一步。在这里,我们采用中国科学院计算技术所开发的中文分词系统 ICTCLAS,其主要功能包括中文分词、词性标注、命名实体识别、新词识别等。(2)去重。去除重复的文本和转发的文本。(3)去除停用词、URL、表情符号、特殊字符、图片、过短微博。文本中的停用词、URL、表情符号、特殊字符、图片对于文本的理解几乎不起作用,为了提高处理效率和节约存储空间,故去除。同时,去除字数少于 10 的过短微博。(4)特征选择。从训练样本集中,结合经验和交通主题特点,通过手动筛选和扩展,提取样本特征词汇。(5)文本标准化入库。将经过预处理后的文本存入数据库。

建模与处理。对于预处理后的文本,提取文本中含有的交通主题、事件,对文本进行分类,分析文本交通话题、情感,多源数据交通信息融合分析。(1)交通主题建模。主题模型挖掘文字主题信息的一种方法,是文本处理中的重要环节,在文本分类、文本检索、语义分析等方面有广泛的应用。比较成功的模型主要有 LSA, PLSA, LDA 等。其中, LDA 模型及其变种因其良好的性能表现,已成为微博数据挖掘中应用较多的方法之一,本文采用 LDA 模型对主题进行建

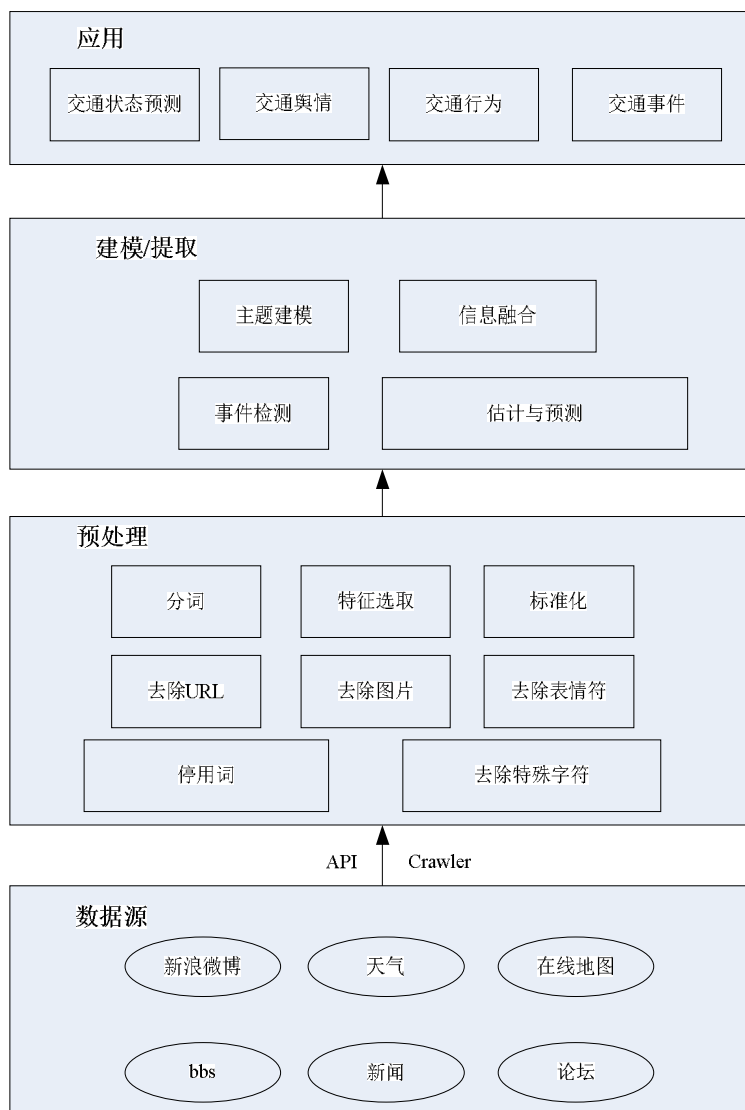


图 1 系统结构

模,有关 LDA 的详细介绍请见第 4 节。(2) 交通事件抽取。交通事件抽取中,需确定事件发生时间、事件发生地点、事件本身内容这三个属性。从新浪微博上抓取的数据都具有时间属性,但由于微博发布时间与其描述的交通信息的发生时间并不总是一致,因此需要对微博文本进行有效解析以获取交通信息的发生时间。相对而言,事件的空间属性更难一些,需精确定位事件发生的道路名称、具体位置或范围、行驶方向;对于无法准确获取上述信息的交通事件,只能给出事件大致发生范围。(3) 信息融合。多个微博博主发的微博中可能描述的是同一个交通

事件 E,其中含有的信息量可能有多有少,将对同一个交通事件的若干个描述融合起来,达到尽可能准确描述交通事件的目的。

(4) 有效性验证,指的是对微博信息的真实性验证。结合其他微博信息,或实际交通物理系统采集到的交通数据进行对比分析,确定信息的有效性。

应用。系统主要用于网络开源数据挖掘交通事故、交通拥堵、交通管制、道路施工、道路损坏等交通事件信息,并对交通运行状态进行溯源分析;结合活动、天气等信息,确定交通瓶颈,分析交通瓶颈产生原因,以及估计和预测未来交通状态;获得人们对交

通政策、交通系统的观点、喜好、评价、吐槽等内容,全面了解社会公众关切的交通运输便利和安全问题,诊断交通系统运营、管理、规划中存在的问题,快速准确监测网上各类交通运输负面舆情,从而及时采集准确的应对措施,以预防、减少和消除突发交通运输舆情造成的负面影响。

4. 主要技术

本系统所涉及的主要技术包括如何采集网络开源交通信息、如何从网络文本中挖掘交通信息,以及如何预测未来交通状态。

4.1 网络开源交通信息采集:

(1) 微博中交通信息抓取

微博的普及,使得它成为一个无处不在的社会传感器,其蕴含大量实时信息,从微博中挖掘事件信息,对于分析、预测、处置事件具有极大的益处。要从微博中采集交通信息,首先要实时抓取微博内容。新浪微博内容抓取方法主要有两类,一类是基于爬虫的方法,一类是基于新浪微博 API 的方法。

利用新浪微博开放平台提供的 API 接口可以获取公共微博、用户发布的微博、搜索某一话题下的微博、获取某个用户的位置等。利用新浪微博 API 抓取微博内容的过程主要有: 1) 获得新浪微博的授权; 2) 通过认证之后,获得 Access Token; 3) 利用 Access Token 调用 API 接口,返回特定的 XML 或 JSON 文件; 4) 解析文件,得到需要的数据,将其保存在数据库中。

采用网络爬虫方式也可以获取新浪微博内容。基于爬虫的方法需要解决模拟登录和页面数据存储问题。常用的抓取策略有深度优先遍历策略、宽度优先遍历策略、Partial PageRank 策略、大站优先策略等。给定微博 URL,可利用网络爬虫抓取页面内容,对其进行拆解分析,获取已发布的微博数量、微博内容、评论数、转发数等信息。

(2) 官网和专业论坛发布的交通相关信息抓取

为了便民出行,交通部门官网也公布交通信息,其公布的常见交通信息包含: 交通实时路况、交通事故、交通拥堵路段、交通事件、道路施工、交通管制等。

不利天气条件和大型活动对交通运行会造成严重影响。天气预报网站会发布每天分时段天气信息。在线售票网站、专业论坛等会公布演唱会、比赛、运动会、大型商场促销、中小学上学等活动信息。

官网和专业论坛发布的交通相关信息采用基于爬虫的方法抓取。

(3) 在线互联网地图服务商提供的交通信息采集

在过去十几年内,在线互联网地图服务发展迅猛。百度地图、高德地图、腾讯地图、mapbar、google 地图、bing 地图、MapQuest,以及其他地图服务均在互联网上可容易被人们访问、使用。特别是一些地图服务商,如高德地图、百度地图、bing 地图,对外提供 API 接口,允许外部人员获取地图信息,创建自己的应用服务。

4.2 Latent Dirichlet Allocation

以文本形式存在的交通信息文档 d 具有不同的类别 $c = \{c_1, c_2, \dots, c_n\}$, 需要

对其进行类别自动标注,这一过程称之为文本分类。常用的分类方法有 Naïve Bayes、SVM、决策树等。在分类过程中,文本通常以词袋 (bag of words) 形式表示,利用训练文本对模型进行训练,模型在有效地学习到训练文本特征后可对测试文本进行标记。

传统的文档相似性判断方法是根据两个文档共同出现的单词的多少,如 TF-IDF 等,但这种方法没有考虑文字的语义关联,可能两个文档很少或几乎没有共同出现的单词,但这两个文档是相似的。Latent Dirichlet Allocation (LDA) 在处理这方面是其中一种比较有效的模型。LDA 是一种文档主题生成模型,可用来识别文档或语料库中的主题信息。LDA 认为一篇文章的每个词都是通过“以一定概率选择了某个主题,并从这个主题中以一定概率选择某个词语”。在 LDA 模型中一篇文档生成的方式如下: 从 Dirichlet 分布 α 中取样生成文档 i 的主题分布 θ_i ; 从主题分布 θ_i 中取样生成文档 i 第 j 个词的主题 z_{ij} ; 从 Dirichlet

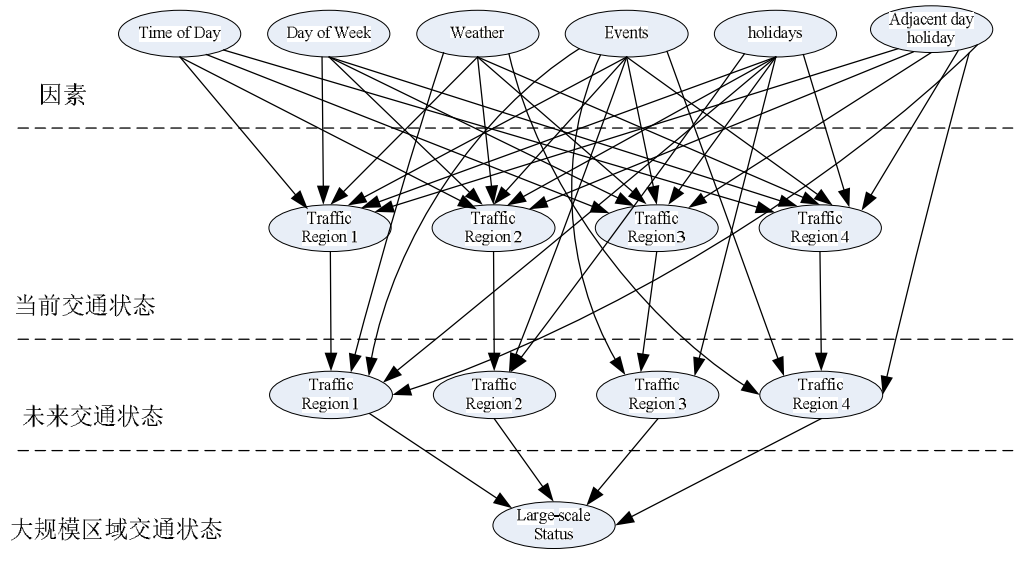


图 2 分层贝叶斯预测模型示例

分布 β 中取样生成主题 z_{ij} 的词语分布

$\phi_{z_{ij}}$; 从词语分布 $\phi_{z_{ij}}$ 中采样最终生成词

语 w_{ij} 。LDA 模型训练好后, 可以用于文档分类、信息提取等。

4.3 分层贝叶斯预测模型

交通系统是复杂的巨系统。影响交通预测的因素众多, 贝叶斯网络可以用于描述交通预测中的因果关系。在利用贝叶斯模型进行交通预测时, 将指定时间窗内每条道路的交通预测对象(如流量、速度、交通状态等)和相关影响因素分别描述为贝叶斯网络模型中的节点, 影响关系描述为贝叶斯网络模型的边。其中: 利用物理检测器可以获得交通流量、速度、排队、交通状态等交通信息, 利用网络开源信息可以收集到天气、事件、节假日等影响交通的因素。图 2 给出了基于分层贝叶斯模型的交通状态预测示例。随着交通数据量的大幅增长, 交通预测对数据存储和计算资源的需求越来越大, 为此, 基于 MapReduce 架构来实现分布式大规模交通预测。

模型参数采用期望最大化 (Expectation Maximization, EM) 算法估计。EM 算法的每次迭代由 E 步 (求期望) 和 M 步 (求极

大) 构成, 其中 E 步用模型当前参数求得响应度, M 步则用响应度来更新模型参数。

5. 结论

海量交通信息的获取是实现智能交通的基础和关键技术。随着移动互联网的高速发展, 作为基于传统物理传感器方法的补充, 基于社交媒体等网络开源内容的方法将成为获取实时、准确交通信息的重要方法。本文对网络开源交通信息获取及分析系统进行了讨论和研究。该系统通过网络爬虫或网站提供的 API 接口采集专题网站、论坛、微博、微信中含有的指定区域内交通内容; 对于以文本形式存在的交通内容, 主要利用中文自然语言处理技术进行分析; 通过机器学习方法挖掘、分析网络中获取的开源交通信息, 主要包括交通事件检测、交通舆情分析、交通行为推断以及交通预测。

目前, 在交通文本分析方面, 笔者已利用词嵌入 (word embedding)、基于深度学习的语言模型等技术提取文本中的交通信息, 效果显著。下一步, 将把这些先进的算法集成到开源交通信息获取及分析系统中。可以说, 基于社交媒体、在线地图等网络开源信息的交通预测、交通事件提取、交通舆情分析等方面的研究刚刚开展, 但仍有大量的工作需要进一步去做。

参考文献

- [1] N. Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, "DynaCAS: Computational Experiments and Decision Support for ITS," *IEEE Intelligent Systems*, vol. 23, no. 6, pp. 19-23, 2008.
- [2] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1-9, 2014.
- [3] F.-Y. Wang, "Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 630-638, 2010.
- [4] K. Zeng, W. Liu, X. Wang, and S. Chen, "Traffic Congestion and Social Media in China," *IEEE Intelligent Systems*, vol. 28, no. 1, pp. 72-77, 2013.
- [5] J. Cao *et al.*, "Web-Based Traffic Sentiment Analysis: Methods and Applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 844-853, 2014.
- [6] 王飞跃, "社会信号处理与分析的基本框架:从社会传感网络到计算辩证解析方法," *中国科学:信息科学*, no. 12, pp. 1598-1611, 2013.
- [7] X. Wang, X. Zheng, X. Zhang, K. Zeng, and F. Y. Wang, "Analysis of Cyber Interactive Behaviors Using Artificial Community and Computational Experiments," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 6, pp. 995-1006, 2017.
- [8] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 4, pp. 919-931, 2013.
- [9] L. Zhao, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Spatiotemporal Event Forecasting in Social Media," presented at the SIAM international conference on Data Mining, 2015.
- [10] 王飞跃, "软件定义的系统与知识自动化:从牛顿到默顿的平行升华," *自动化学报*, no. 01, pp. 1-8, 2015.
- [11] F. Y. Wang, X. Wang, L. Li, and L. Li, "Steps toward Parallel Intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 4, pp. 345-348, 2016.
- [12] Y. Lv, Y. Chen, X. Zhang, Y. Duan, and N. L. Li, "Social media based transportation research: the state of the work and the networking," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 1, pp. 19-26, 2017.
- [13] X. Wang, X. Zheng, Q. Zhang, T. Wang, and D. Shen, "Crowdsourcing in ITS: The State of the Work and the Networking," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1-10, 2016.
- [14] X. Zheng *et al.*, "Big Data for Social Transportation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 3, pp. 620-630, 2016.
- [15] E. Morgul *et al.*, "Virtual Sensors: Web-Based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2442, pp. 106-116, 2014.
- [16] A. Kurkcu, M. Eng, E. F. Morgul, and K. Ozbay, "Extended Implementation Methodology for Virtual Sensors: Web-based Real Time Transportation Data Collection and Analysis for Incident Management," in *Transportation Research Board 94th Annual Meeting*, 2015, no. 15-3374.
- [17] J. He, W. Shen, P. Divakaruni, L. Wynter, and R. Lawrence, "Improving traffic prediction with tweet semantics," presented at the Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, Beijing, China, 2013.
- [18] M. Ni, Q. He, and J. Gao, "Using social media to predict traffic flow under special event conditions," in *The 93rd Annual Meeting of Transportation Research Board*, 2014.
- [19] S. Grosenick, "Real-Time Traffic Prediction Improvement through Semantic Mining of Social Networks," Master, University of Washington,

- 2012.
- [20] T. Zhou, L. Gao, and D. Ni, "Road traffic prediction by incorporating online information," presented at the Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 2014.
- [21] D. Pathania and K. Karlapalem, "Social Network Driven Traffic Decongestion Using Near Time Forecasting," presented at the Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey, 2015.
- [22] Y.-y. Chen, Y. Lv, Z. Li, and F. Y. Wang, "Long short-term memory model for traffic congestion prediction with online open data," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 132-137.
- [23] 熊佳茜, "基于 CRF 的中文微博交通信息事件抽取," 硕士, 上海交通大学, 2014.
- [24] 崔健, 冯璇, 张佐, "基于微博的交通事件提取与文本分析系统," *交通信息与安全*, no. 06, pp. 132-135, 2013.
- [25] C. Gutierrez, P. Figuerias, P. Oliveira, R. Costa, and R. Jardim-Goncalves, "Twitter mining for traffic events detection," in *Science and Information Conference (SAI), 2015*, 2015, pp. 371-378.
- [26] E. D'Andrea, P. Ducange, B. Lazzarini, and F. Marcelloni, "Real-Time Detection of Traffic From Twitter Stream Analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2269-2283, 2015.
- [27] M. Maghrebi, A. Abbasi, and S. T. Waller, "Transportation application of social media: Travel mode extraction," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 1648-1653.
- [28] Zhenhua Zhang, Qing He, and S. Zhu, "Exploring Travel Behavior with Social Media: An Empirical Study of Abnormal Movements Using High-Resolution Tweet Trajectory Data," in *Proceedings of 96th Transportation Research Board Annual Meeting* Washington DC, 2017.
- [29] S. Hasan, X. Zhan, and S. V. Ukkusuri, "Understanding urban human activity and mobility patterns using large-scale location-based data from online social media," presented at the Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, Illinois, 2013.
- [30] D. Semwal, S. Patil, S. Galhotra, A. Arora, and N. Unny, "STAR: Real-time Spatio-Temporal Analysis and Prediction of Traffic Insights using Social Media," presented at the Proceedings of the 2nd IKDD Conference on Data Sciences, Bangalore, India, 2015.
- [31] F. Lécué *et al.*, "STAR-CITY: semantic traffic analytics and reasoning for CITY," presented at the Proceedings of the 19th international conference on Intelligent User Interfaces, Haifa, Israel, 2014.
- [32] F. Lécué *et al.*, "Semantic Traffic Diagnosis with STAR-CITY: Architecture and Lessons Learned from Deployment in Dublin, Bologna, Miami and Rio," in *The Semantic Web – ISWC 2014*, vol. 8797, P. Mika *et al.*, Eds. (Lecture Notes in Computer Science: Springer International Publishing, 2014, pp. 292-307.
- [33] F. Lécué, R. Tucker, V. Bicer, P. Tommasi, S. Tallevi-Diotallevi, and M. Sbodio, "Predicting Severity of Road Traffic Congestion Using Semantic Web Technologies," in *The Semantic Web: Trends and Challenges*, vol. 8465, V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, and A. Tordai, Eds. (Lecture Notes in Computer Science: Springer International Publishing, 2014, pp. 611-627.
- [34] F. Lécué *et al.*, "Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system and lessons learned in Dublin City," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 27–28, pp. 26-33, 2014.
- [35] B. Singh, "Real Time Prediction of Road Traffic Condition in London via Twitter and Related Sources," Master, Middlesex University, Middlesex University, 2012.