

文章编号: 2095-2163(2023)03-0046-05

中图分类号: TP183

文献标志码: A

基于 BERT 和交互注意力的方面级情感分类模型

袁逸飞, 单剑锋

(南京邮电大学 电子与光学工程学院, 南京 210023)

摘要: 目前针对方面级的情感分类模型往往忽略了对方面词的单独建模, 传统的注意力机制模型可能会给句子中的情感词分配较低的注意力权重。鉴于此, 本文提出一种结合交互注意力网络 (IAN) 和双向编码器表征技术 (BERT) 的情感分类方法。首先将输入分为 3 部分: 文本、方面词、文本+方面词。通过 BERT 得到词向量, 双向门控单元 (BiGRU) 负责进行语义信息的提取, 隐藏向量分别输入特征提取层, 最后将 2 部分进行拼接, 输入 *softmax* 层得到情感极性。通过在公开数据集上的实验证明, 本文的模型效果能够比基线模型提高 5% 的精确度。

关键词: 方面级的情感分类; 交互注意力模型; 双向门控单元; BERT 预训练模型

Aspect level classification model based on BERT and interactive attention

YUAN Yifei, SHAN Jianfeng

(College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

[Abstract] The current emotion classification models for aspect level tend to ignore the independent modeling of aspect words. The traditional attention mechanism models may assign low attention weight to emotion words in sentences. In view of this, an emotion classification method combining interactive attention network (IAN) and bidirectional encoder representation technique (BERT) is proposed. First of all, the input is divided into three parts: text, aspect words, text + aspect words. Then, the word vector is obtained by BERT, the bidirectional gated unit (BiGRU) is responsible for semantic information extraction, and the hidden vectors are input into the feature extraction layer respectively. Finally, the two parts are spliced and input into *softmax* layer to get the emotional polarity. The experimental results on SemEval dataset show that compared with existing methods, the proposed method can effectively achieve the improvement of 5% in the sentiment classification.

[Key words] aspect level sentiment analysis; interactive attention model; BiGRU; BERT

0 引言

情感分析是当前自然语言处理领域的重要分支之一。而基于方面级的情感分类任务将情绪分类又推进了一步。通常情况下, 如果只知道一个句子或者文档的极性虽然有一定的作用, 但不能分辨出极性是针对谁, 于是方面级的情感分类应运而生^[1]。任务描述如图 1 所示。

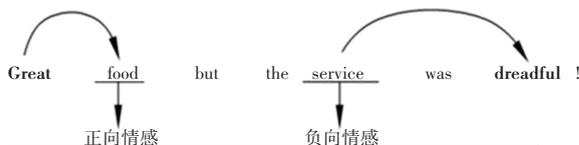


图 1 方面级情感分类任务实例

Fig. 1 Task instance of aspect level sentiment classification

当前, 神经网络模型已越来越多地用在方面级的情绪分类中^[2-4]。现有的研究大多采用监督学习的方法构建情感分类器, 比如基于特征的支持向量机^[5]、基于神经网络的方法^[6], 大多数基于注意力机制的模型没有考虑上下文与方面词之间的交互信息。Ma 等学者^[7]提出交互注意力网络, 但是研究中使用的双向长短期记忆网络 (BiLSTM) 可能存在于小数据集下表现不如双向门控单元 (BiGRU) 的结果, 同时传统的词向量工具 word2vec 不能很好地表征出每个词的向量, 不能产生一词多义的效果, 在实际的操作中, 则无法获取最终想要的向量。受文献^[8]的启发, 本文以 BiGRU 为基础, 提出面向方面级情感分类的模型。首先使用 BERT 预训练语言模型进行词嵌入, 并通过 BiGRU 获得输入向量间的隐藏

作者简介: 袁逸飞 (1997-), 男, 硕士研究生, 主要研究方向: 智能信息处理; 单剑锋 (1965-), 男, 博士, 副教授, 主要研究方向: 智能信息处理、目标识别。

通讯作者: 单剑锋 Email: shanjf@njupt.edu.cn

收稿日期: 2022-04-26

向量,接着分别输入多头自注意力层和交互注意力层,最后两者拼接得到结果。

1 相关工作

在最初的基于机器学习方法中,方面级的情感分类被视作一个文本分类的任务。朴素贝叶斯和支持向量机的方法可以不再需要方面词就可预测情感极性。但是这类方法有个无法避免的缺陷,即要依赖情感词典和手工设计的特征。就使得训练结果强烈依赖手工设计的特征的质量,同时,由于特征都是手工标记,所以特征数量有上限的,预测精度也随之遇到瓶颈。

2 相关技术及 BiGRU-BERT 模型

模型由词嵌入层、语义学习层、注意力层和融合层组合而成。为了便于理解,文本上下文使用 $S = \{\omega_1, \omega_2, \dots, \omega_i, \dots, \omega_{i+l}, \omega_n\}$ 表示,其中, n 表示上下文中词的个数,方面词为 $\{\omega_i, \dots, \omega_{i+l}\}$ 。这里, t 和 l 分别表示词的位置和方面词的长度。

2.1 词嵌入层

传统使用的是 Word2Vec、FastText、Glove 等方法。但是本文引入 BERT 预训练模型。BERT 模型相比传统的其他方法,主要是引入了注意力机制,同时解决了传统方法不能做到的一词多义问题, BERT 预训练模型是基于语义理解的深度双向语言模型,由 Google 的 Devlin 等学者^[9]于 2018 年 10 月提出的预训练模型。模型用的是比 Word2Vec 更强大的双向 Transformer 编码器,是基于多层 Transformer 编码器演化而来的。每句话采用 [CLS] 开始, [SEP] 结束,并以遮蔽语言建模 (Masked language model) 和下一句话预测 (next sentence prediction) 为无监督目标,因此具有了更强的词向量抽象能力, BERT 在大量任务中都获得了相比 Glove 而言更好的性能。本文在将 [CLS] 文本 [SEP]、[CLS] 方面词 [SEP] 作为输入的基础上进行了拓展,加入了 [CLS]+文本+[SEP]+方面词+[SEP] 的第三个输入。

2.2 GRU

由于 LSTM 内部结构复杂、计算量大,在模型训练上不仅需要较长的时间,模型所需参数也较多。

2014 年,提出了 GRU 模型^[10]。与 LSTM 相比, GRU 模型解决了长短期记忆的问题。研究可知, GRU 只有 2 个门,参数减少了 1/3,因此更易收敛,收敛速度更快,可以大大加速迭代过程。GRU 模型

的单元结构如图 2 所示。对此模型中用到的数学方法,可由如下公式分别进行描述:

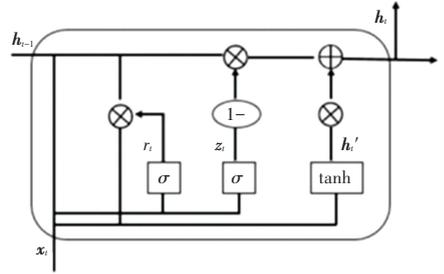


图 2 GRU 神经元结构

Fig. 2 GRU neuron structure

$$r_t = \text{sigmoid}(W_r [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \text{sigmoid}(W_z [h_{t-1}, x_t]) \quad (2)$$

$$\tilde{h}_t = \tanh(W[r_t \odot h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

其中, z_t 和 r_t 分别表示更新门和重置门; h_t 是 t 时刻的激活状态; h_{t-1} 是 $t - 1$ 时刻的隐层状态。重置门 r_t 控制着前一个状态的信息 h_{t-1} 传入状态。

2.3 BiGRU

在 GRU 基础上,若要当前时刻的输出能获取更多信息,故而使用双向 GRU 模型,将顺向和反向 GRU 相结合, BiGRU 是由 2 个 GRU 上下正反向合并在一起组成的。在每一个时刻 t , 输入会提供这 2 个方向相反的 GRU, 输出则是由 2 个单向 GRU 拼接而成。BiGRU 网络结构如图 3 所示。

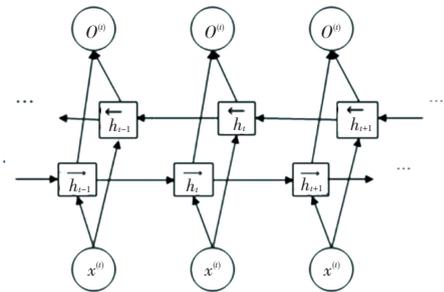


图 3 BiGRU 网络结构

Fig. 3 BiGRU neural network structure

2.4 交互注意力层

针对上下文单词表示,隐藏层的输出作为注意力层的输入 $[h_c^1, h_c^2, \dots, h_c^n]$, 注意力机制利用上下文表示生成注意力向量 α_i , 可由如下公式计算求值:

$$\alpha_i = \frac{\exp(\gamma(h_c^i, t_{avg}))}{\sum_{j=1}^n \exp(\gamma(h_c^j, t_{avg}))} \quad (5)$$

$$\gamma(h_c^i, t_{avg}) = \tanh(h_c^i \cdot W_a \cdot t_{avg}^T + b_a) \quad (6)$$

其中, t_{avg} 是目标表示; γ 是计算 h_c^i 在上下文中重要性的函数; W_a 是权重; b_a 是偏置项; \tanh 是非线性函数; t_{avg}^T 是 t_{avg} 的转置。

同理, 针对方面词的表示, 见如下:

$$\beta_i = \frac{\exp(\gamma(h_r^i, c_{avg}))}{\sum_{j=1}^m \exp(\gamma(h_r^j, c_{avg}))} \quad (7)$$

其中, 各个参数的含义与式(6)相同。通过计算单词的注意力权重后, 可以利用注意力向量得到上下文 c_r 和方面词 t_r 的表示, 即:

$$c_r = \sum_{i=1}^n \alpha_i h_c^i \quad (8)$$

$$t_r = \sum_{i=1}^m \beta_i h_r^i \quad (9)$$

2.5 多头自注意力层

多头注意力是指能够同时执行多个注意力函数。一个注意力函数映射键向量 $K = \{k_1, k_2, \dots, k_n\}$ 和查询向量 $Q = \{q_1, q_2, \dots, q_m\}$ 作为输出向量。同时, 计算一组查询 Q 上的注意力函数, 用到的公式为:

$$Attention(k, q) = softmax(f_s(k, q))k \quad (10)$$

$$f_s(k_i, q_j) = \tanh([k_i; q_j] \cdot W_{att}) \quad (11)$$

其中, f_s 表示对齐函数, 能够学习 q_j 和 k_i 之间的语义关系; $W_{att} \in U^{2d_{hid}}$ 是可学习的权重。

多头注意力机制可同时关注来自并行子空间中学习 n 个不同的分数, 并且在对齐方面的功能尤为强大。 n 头的输出是串联起来投射到特定隐藏维度 d_{hid} , 推得的数学公式具体见如下:

$$MHA(k, q) = [o^1; o^2; \dots; o^{n_{head}}] \cdot W_{mh} \quad (12)$$

$$o^m = Attention^m(k, q) \quad (13)$$

其中, $W_{mh} \in U^{d_{hid} \times d_{hid}}$ 表示向量的拼接; o^m 是第 m 个注意力头的输出; $m \in [1, n_{head}]$ 。

多头自注意力是一种 $q = k$ 的特殊情况。给定一个上下文的词嵌入向量 e^c , 可以得到多头自注意力上下文表示:

$$e^{self} = MHA(e^c, e^c) \quad (14)$$

其中, 学习到的 $e^{self} = \{c_1^{self}, c_2^{self}, \dots, c_n^{self}\}$ 是长期依赖。

3 模型实现

本文给出的模型框架如图4所示。由图4可看到, 主要由词嵌入层、特征提取层、词级交互注意力层、输出层组成:

(1) 词嵌入层。使用预训练 BERT 模型进行词嵌入, 将训练数据输入。

(2) 特征提取层。BiGRU 对词向量进行编码, 用于提取语义特征。前一个隐层的输出是后一个隐层的输入, 最后得到隐藏层。

(3) 词级交互注意力层。将方面词和上下文的隐藏层分别进行池化, 提取出其中的特征。

(4) 多头自注意力层。计算每个词向量和其它词向量之间的关系。

(5) 输出层。将得到的局部特征和全局特征进行融合, 再经过 $softmax$ 归一化操作, 得到最终的情感预测结果。

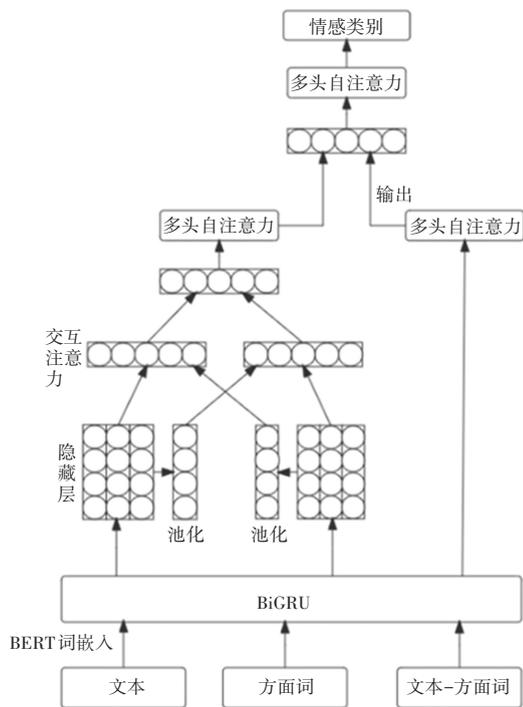


图4 BiGRU-BERT模型的整体框架

Fig. 4 BiGRU-BERT model structure

3.1 实验平台与实验数据

本文选取 Twitter 数据集^[11], SemEval2014 Task4 中 Restaurant 和 Laptop 的数据集, 其中极性分成消极、中性、积极三种, 具体见表1。

表1 数据集信息

Tab. 1 Data set

数据集	积极	中性	消极
Restaurant-Train	2 164	637	807
Restaurant-Test	728	196	196
Laptop-Train	994	464	870
Laptop-Test	314	169	128
Twitter-Train	1 561	3 127	1 560
Twitter-Test	173	346	173

3.2 实验参数设置

BERT词向量维度设置为768,用正态分布对权重矩阵进行初始化,偏差初始化为0。在训练中,用Adam优化器更新参数。设置Batch为16,学习率为 $5e-5$ 。实验软硬件环境见表2。

表2 实验环境

Tab. 2 Experimental environment

实验环境	具体信息
操作系统	Windows11
CPU	R7 5800X
显卡	RTX3070
内存	16 GB
开发语言	Python3.6
开发框架	Pytorch
开发工具	Pycharm

3.3 实验结果分析

为了全面地评价网络的性能,列出了一些基线模型方法进行比较,对比的基线模型介绍如下。

(1)SVM^[12]:支持向量机,通过人工构造特征,引入外部情感词典完成分类。

(2)LSTM:LSTM模型,只使用LSTM对上下文进行处理。

(3)Bi-LSTM^[13]:直接将文本输入LSTM,只是从前后有双向的LSTM。

(4)TD-LSTM^[14]:用了2个LSTM模型,分别输入左半部分的上文带上方面词,右半部分的下文带上方面词。

(5)ATAE-LSTM^[15]:提出了融合方面信息的LSTM和注意力机制的模型。能够通过注意力机制只关注特定方面。

(6)IAN:用LSTM对方面词和句子分别建模,并通过注意力机制,分别对其中的隐层进行提取,再进行拼接得到情感分类。

(7)RAM^[16]:循环注意力网络,提出循环注意力机制,多跳数目设置为3。

(8)BiIAGRU-BERT:基于BERT的交互注意力模型,即本模型。

仿真实验对比结果见表3。实验结果表明,加入方面信息和交互注意力能够显著改善分类效果。因为LSTM神经网络相对复杂且在小数据集下的效果不如GRU,因此本文采用GRU作为特征提取层。BiIAGRU-Glove模型证明,效果比仅有交互注意力IAN模型和注意力ATAE-LSTM模型要好。BiIAGRU-BERT和BiIAGRU-Glove具有相同的模

型结构,不同的是前一个模型采用的是BERT预训练模型,结果表明与BERT模型相结合能在方面级情感分类任务中取得更好的效果。

表3 实验结果对比

Tab. 3 Experimental results of different models

模型	指标	Restaurant	Laptop	Twitter
SVM	Acc/ %	80.2	70.5	63.4
	F_1 值	NA	NA	63.3
LSTM	Acc/ %	74.5	66.5	64.4
	F_1 值	56.1	60.6	62.7
TD-LSTM	Acc/ %	75.5	67.7	65.2
	F_1 值	62.3	61.3	64.9
ATAE-LSTM	Acc/ %	77.3	68.7	69.4
	F_1 值	63.5	63.0	67.7
IAN	Acc/ %	78.3	72.1	71.6
	F_1 值	66.7	67.4	70.2
RAM	Acc/ %	80.2	74.2	70.1
	F_1 值	70.3	69.9	68.5
BiIAGRU-Glove	Acc/ %	79.3	74.9	71.6
	F_1 值	68.1	66.9	70.3
BiIAGRU-BERT	Acc/ %	83.2	78.3	74.3
	F_1 值	73.8	73.0	73.4

4 结束语

本文提出一个基于注意力编码器与交互注意力网络模型,采用基于注意力的编码器对上下文和目标之间建模,还将预训练模型BERT应用到模型中,结果表明能够提高方面级情感分类的准确率和 F_1 值,达到了较好的分类的效果。实验结果表明注意力机制结合BERT预训练语言模型在基于方面级的情感分类任务中的有效性。

参考文献

- [1] ZHANG Lei, WANG Shuai, LIU Bing. Deep learning for sentiment analysis: A survey [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.
- [2] JIANG Long, YU Mo, ZHOU Ming, et al. Target-dependent twitter sentiment classification [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, Oregon: ACL, 2011: 151-160.
- [3] DONG Li, WEI Furu, TAN Chuanqi, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C]//Proceedings of The 52nd Annual Meeting of The Association for Computational Linguistics (Volume 2: Short Papers). Baltimore: ACL, 2014: 49-54.
- [4] DUY-TIN V, ZHANG Yue. Target-dependent twitter sentiment classification with rich automatic features [C]// Proceedings of

- The 24th International Conference on Artificial Intelligence (Ijcai'15). Buenos Aires Argentina; AAAI, 2015; 1347-1353.
- [5] CHERKASSKY V, MA Yunqian. Practical selection of SVM parameters and noise estimation for SVM regression [J]. Neural Networks, 2004, 17(1): 113-126.
- [6] WANG S C. Artificial neural network [M]//Interdisciplinary Computing in Java Programming. Boston, MA: Springer, 2003: 81-100.
- [7] MA Dehong, LI Sujian, ZHANG Xiaodong, et al. Interactive attention networks for aspect-level sentiment classification [J]. arXiv preprint arXiv:1709.00893, 2017.
- [8] SONG Youwei, WANG Jiahai, JIANG Tao, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv preprint arXiv:1902.09314, 2019.
- [9] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4171-4186.
- [10] FU Rui, ZHANG Zuo, LI Li. Using LSTM and GRU neural network methods for traffic flow prediction [C]//2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). Wuhan: IEEE, 2016: 324-328.
- [11] KIRITCHENKO S, ZHU Xiaodan, CHERRY C, et al. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews [C]//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin: ACL, 2014: 437-442.
- [12] DONG Li, WEI Furu, TAN Chuanqi, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). Baltimore, Maryland: ACL, 2014: 49-54.
- [13] YU Yong, SI Xiaosheng, HU Changhua, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural Computation, 2019, 31(7): 1235-1270.
- [14] HUANG Zhiheng, XU Wei, YU Kai. Bidirectional LSTM-CRF models for sequence tagging [J]. arXiv preprint arXiv: 1508.01991, 2015.
- [15] WANG Yequan, HUANG Minlie, ZHU X, et al. Attention-based LSTM for aspect-level sentiment classification [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: ACL, 2016: 606-615.
- [16] CHEN Peng, SUN Zhongqian, BING Lidong, et al. Recurrent attention network on memory for aspect sentiment analysis [C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: ACL, 2017: 452-461.

(上接第45页)

- [6] 张俊涛,李志勇,张浩,等. 利用出租车轨迹数据估计城市道路拥堵状况 [J]. 测绘工程, 2016, 25(09): 68-72, 76.
- [7] YUANJing, ZHENG Yu, XIE Xing, et al. T-Drive: Enhancing driving directions with taxi drivers' intelligence [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(1): 220-232.
- [8] 邬群勇,胡振华,张红. 基于多源轨迹数据的城市交通状态精细划分与识别 [J]. 交通运输系统工程与信息, 2020, 20(01): 83-90.
- [9] 魏薇,丁浪. 基于使用者运动轨迹大数据的城市绿道系统效用评估——以杭州市为例 [J]. 建筑与文化, 2018, 15(07): 155-156.
- [10] 邬群勇,张良盼,吴祖飞. 利用出租车轨迹数据识别城市功能区 [J]. 测绘科学技术学报, 2018, 35(04): 413-417, 424.
- [11] 朱延冰. 基于轨迹大数据的出租车司机就餐点选址问题研究 [D]. 上海: 华东师范大学, 2019.
- [12] ZHENG Yu, XIE Xing, MA Weiyang. GeoLife: A collaborative social networking service among user, location and trajectory [J]. Bulletin of the Technical Committee on Data Engineering, 2010, 33(2): 32-39.
- [13] LIU Huiping, JIN Cheqing, ZHOU Aoying. Popular route planning with travel cost estimation from trajectories [J]. Frontiers of Computer Science, 2020, 14(1): 191-207.
- [14] ZHENG Yu, ZHANG Lizhu, XIE Xing, et al. Mining interesting locations and travel sequences from GPS trajectories [C]//Proceedings of the 18th International Conference on World Wide Web. Madrid, Spain. ACM, 2009: 791-800.
- [15] 郎月华,李仁杰,傅学庆. 基于GPS轨迹栅格化的旅游行为空间模式分析 [J]. 旅游学刊, 2019, 34(06): 48-57.
- [16] 钮心毅,康宁. 上海郊野公园游客活动时空特征及其影响因素——基于手机信令数据的研究 [J]. 中国园林, 2021, 37(08): 39-43.
- [17] 林麟. 全球及我国软件产业发展研究 [J]. 信息通信技术与政策, 2019(03): 61-62.
- [18] 周盖,文勇军,樊志良,等. 基于Android平台的校车管理系统设计与实现 [J]. 智能计算机与应用, 2018, 8(02): 88-91.
- [19] 兰志辉,陈莉,段治州. 基于时间序列聚类的轨迹停留点检测算法 [J]. 计算机应用研究, 2020, 37(12): 3557-3560.
- [20] 兰文涛. 基于Android智能手机监控与轨迹分析系统的设计与实现 [D]. 西安: 西安科技大学, 2018.
- [21] 陈丽璐. 基于出租车轨迹数据的载客热点与打车热点的研究与实现 [D]. 镇江: 江苏大学, 2019.
- [22] 裴玉龙,李浩然. 基于手机GPS轨迹的出行方式判别分析研究 [J]. 重庆理工大学学报, 2018, 32(06): 145-152.
- [23] 赵瑜. 不同交通状态下基于手机GPS轨迹的出行信息采集效果评估研究 [D]. 成都: 西南交通大学, 2016.
- [24] 肖艳丽,张振宇,杨文忠. 基于GPS轨迹的用户移动行为挖掘算法 [J]. 计算机应用与软件, 2015, 32(11): 83-87.
- [25] 胡开喜. 移动轨迹的异常检测及其聚类研究 [D]. 重庆: 重庆大学, 2018.
- [26] 沈程,谭海波,许金林. 基于Android和GPS轨迹记录和能耗测算研究 [J]. 计算机技术与发展, 2013, 23(11): 173-176.
- [27] 王欢. 步行能量消耗特征的研究与应用 [D]. 上海: 上海体育学院, 2013.
- [28] 张开斌,阮廷勇. 中国六城市汽车行驶工况的测试统计分析 [J]. 汽车研究与开发, 2005(12): 33-36.
- [29] 梁春岩. 自行车交通流特性及其应用研究 [D]. 长春: 吉林大学, 2007.