

文章编号: 2095-2163(2020)11-0083-05

中图分类号: TP183

文献标志码: A

基于 BiGRU 与方面注意力模块的情感分类方法

宋焕民, 张云华

(浙江理工大学 信息学院, 杭州 310018)

摘要: 方面级别的情感分析已广泛应用于文本信息挖掘,但目前广泛使用的 LSTM(Long Short-Term Memory)循环神经网络在处理方面级情感分类任务时,不能充分学习文本上下文信息,且存在模型计算复杂、训练时间长的问题。针对该问题,本文提出利用双向门控循环神经网络与方面注意力模块结合的情感分类方法。双向门控循环神经网络参数更少,模型训练更快,可以有效提取文本深层次的信息;将注意力操作与方面信息相结合,能充分提取特定方面的信息。该方法在 SemEval(Semantic Evaluation)数据集上的实验结果表明,相对现有的方面级情感分析方法,能有效提升处理速度和优化情感分类效果。

关键词: 基于方面的情感分析;自然语言处理;注意力机制;双向门控循环神经网络

Sentiment classification method based on BiGRU and aspect attention module

SONG Huanmin, ZHANG Yunhua

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

[Abstract] Aspect level sentiment analysis has been widely used in text information mining, but the widely used LSTM recurrent neural network can not fully learn the text context information when dealing with aspect level emotion classification tasks, and it has the problem of complex model calculation and long training time. This paper proposes an emotion classification method based on the combination of bidirectional gated cyclic neural network and aspect attention module. Firstly, the bidirectional gated recurrent neural network has fewer parameters and faster model training, which can effectively extract the deep-seated information of text; secondly, the combination of attention operation and aspect information can fully extract the hidden information of specific aspects. The experimental results on SemEval dataset show that compared with the existing aspect level sentiment analysis methods, the proposed method can effectively improve the processing speed and optimize the effect of sentiment classification.

[Key words] Aspect level Sentiment Analysis; text sentiment classification; attention mechanism; bidirectional gated recurrent unity(BiGRU)

0 引言

方面级别的情感分类是情感分析中的一项细粒度任务^[1],它能够提供更完整,更深入的结果。本文在方面级别的情感分类任务中发现,句子的情感极性与句子内容以及所属的方面高度相关。例如,句子“I liked the service and the staff, but not the food”中包含“service”和“food”两个方面。“service”方面它是积极的,而在“food”方面则是消极的。同一个句子中不同方面的情感极性可能不同,甚至是相反的。神经网络已在各种 NLP(Natural Language Processing)任务中达到了非常好的性能^[2]。如机器翻译、语义识别和智能问答等。但是,处理方面级别情感分类的神经网络模型还处在初期阶段。Zhou 等人^[3]提出基于特定方面的关系分类模型;Liu 等人^[4]利用区域卷积神经网络和分层长短期记忆网络结合的深度分层网络模型,处理方面情感分析任

务。现有方法可以提取特定方面的隐藏特征信息,但每次只能处理句子的一个特定方面。对于包含多个不同方面的句子需要重复输入,存在计算及训练时间过长的问题。

为了解决上述问题,本文提出基于 BiGRU 结合方面注意力模块^[5]的情感分类模型。首先采用双向 GRU 神经网络编码输入,然后使用多个方面注意力模块解码获得分类结果。解码时,多个方面注意力模块可以同时句子中不同方面进行情感分类,解决了传统方法重复输入的问题。将句子的输入向量通过 BiGRU 层编码,缩短了模型的训练时间;BiGRU 层的输出向量结合注意力机制,能充分提取句子的关键信息,取得更好的分类效果。

1 相关工作

机器学习的方法已较广泛地被用于方面级情感分析任务。如 Vo 等人^[6]采用多种词向量、情感词

作者简介: 宋焕民(1995-),男,硕士研究生,主要研究方向:智能信息处理;张云华(1965-),男,博士,教授,主要研究方向:软件工程、智能信息处理。

收稿日期: 2020-09-11

典和池化函数,以自动特征抽取方式,提取对应方面的情感极性。Kirit 等人^[7]采用情感词典结合 SVM 的方式,抽取方面级的情感极性。但是,机器学习方法基于复杂的特征工程,且依赖于大规模的文本预处理。

近年来,深度学习在自然语言处理任务中表现优异,相比机器学习方法,极大的减轻了模型对特征工程的依赖,是目前处理情感分类问题的主流方法。如 Tang 等人^[8]提出的以目标词为中心,将文本拆分为两部分,分别以正序和倒序输入到 LSTM 中的 TD-LSTM 模型。相比传统的 LSTM 模型不仅效果更佳,且提出的 TC-LSTM 模型在 TD-LSTM 的基础上,让目标词与上下文之间关联,能达到更好的效果。Bao 等人^[9]提出了基于双向 LSTM 网络和词语位置信息权重建模模型,相较于传统 LSTM 模型及利用手工特征的机器学习算法效果更好。Wang 等人^[10]提出基于注意力机制的 LSTM-ATT 模型,该模型能够根据不同方面关注句子的不同部分,在方面级情感分析任务中取得了较好的效果。Augenstein 等人^[11]提出的 BiLSTM 网络模型,能同时捕捉词语的前后时序关系,获得词语间的前后依赖关系。刘洋等人^[12]基于 GRU 网络对时间序列建模,有效提高了预测的准确性;李骁等人^[13]采用 BiGRU 网络处理互联网信息输入序列,能快速准确的提取信息;Rozenald 等人^[14]提出了 BiGRU 网络结合卷积神经网络的模型,经过多次池化操作再提取文本特征信息进行文本分类。

综上所述,相关工作大多集中在 BiLSTM 或 BiGRU 神经网络结合 attention 模型的研究上,且在情感分类任务上有不错的表现。本文方法是以前人研究成果为基础,采用结构简单、计算量少、所需存储空间小的 BiGRU 神经网络和方面注意力模块混合的网络模型,用于方面情感分析。该方法能使句子中的不同方面同时输入模型并同时处理,为此采用了多个方面注意力模块。每个方面注意力模块具有独立的方面信息及注意力参数,能对特定方面的特征信息进行提取,能够很好的识别不同方面的情感极性。

2 相关技术及 BiGRU-AAM 模型

2.1 相关技术

2.1.1 GRU

由于 LSTM 内部结构复杂计算量大,在模型训练上不仅需要的时间长,而且模型所需参数也较多。2014 年 GRU 模型被提出,其模型结构如图 1 所示。

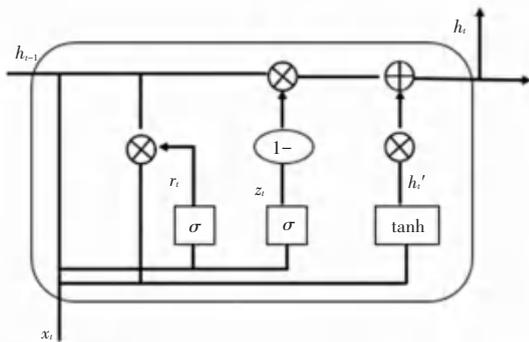


图 1 GRU 神经元结构

Fig. 1 GRU neuron structure

GRU 模型结构简单,收敛性好;合并了 LSTM 的隐层状态和单元状态,需要更少的参数,且效果不低于 LSTM 模型。该模型由两个门控单元组成,一个门控控制更新,另一个门控控制重置。更新控制被带入的前一时刻的状态信息程度;重置控制前一时刻的隐层信息被忽略的程度,忽略的信息越多值越小。GRU 模型更新方式如下:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), \quad (1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \quad (2)$$

$$h'_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]), \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t. \quad (4)$$

其中, z_t 为更新门控在 t 时刻的状态; r_t 为重置门控在 t 时刻的状态; h'_t 是 t 时刻对应的候选激活状态; h_t 是 t 时刻的激活状态; h_{t-1} 是 $(t - 1)$ 时刻的隐层状态。

2.1.2 BiGRU

文本情感分类任务中,要求模型能充分学习文本上下文的信息,将前一时刻的状态及后一时刻的状态关联,得出当前时刻的状态,以提取文本深层次的特征。BiGRU 是由两个单向的、方向相反的 GRU 构成。每一时刻,两个方向相反的 GRU 同时处理输入,共同决定当前时刻的输出。图 2 展示了一个沿着时间展开的 BiGRU 网络,其中每个时间步 t 的输入为 x_t 。

如图 2 所示,把 BiGRU 看作两个单向的 GRU,则前向隐层状态 \vec{h}_{t-1} 和反向隐层状态 \overleftarrow{h}_{t-1} 加权求和,得到 t 时刻的隐层状态。相关计算如式(5) - 式(7):

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}), \quad (5)$$

$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}), \quad (6)$$

$$h_t = w \vec{h}_t + v \overleftarrow{h}_t + b_t. \quad (7)$$

公式中的 $GRU()$ 函数表示对输入的词向量进

行非线性变换,对词向量编码,得到对应的隐层状态; w_t 是 t 时刻双向 GRU 的前向隐层状态的权重, v_t 是相应的反向隐层状态的权重; b_t 是 t 时刻隐层状态的偏置。

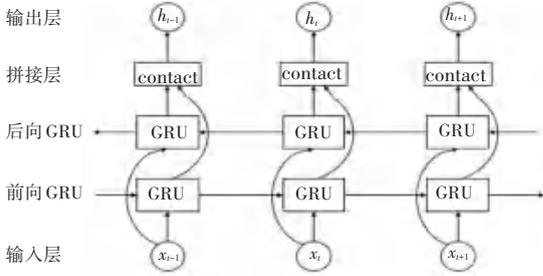


图 2 BiGRU 神经网络结构

Fig. 2 BiGRU neural network structure

2.2 BiGRU-AAM 模型

BiGRU-AAM 模型是一个编码解码器结构,如图 3 所示。输入层和 BiGRU 层是编码器部分,解码器部分是由多个方面注意力模块构成。

该方法由如下几步完成:

(1) 输入

对输入进行预处理,通过词嵌入操作,用词向量表示每个词。

(2) BiGRU 层

BiGRU 神经网络从 2 个方向对输入进行语义特征提取,得到每个时间步上的隐层状态和句子的向量表示。

(3) 方面注意力模块

BiGRU 层输出的隐层状态结合方面信息,通过注意力操作,可以得到句子特定方面的向量表示。

(4) 输出

特定方面的向量表示在全连接层利用 softmax 函数处理,得到句子的最终分类结果。

本文的情感分类任务是判断输入句子中的各个方面的情感极性。每个方面信息投影为一个多维连续的值向量,第 i 个方面的方面向量可以表示为 $v_{a,i} \in U^k$, k 是方面向量的维度。

2.2.1 编码器

编码器部分提取句子的语义特征,包括向前和向后的两个方向。通过 BiGRU 网络,获取第 t 个时间步上的隐藏层状态 h_t ,对于长度为 n 的句子,隐藏层输出矩阵为: $H = [h_1, h_2, \dots, h_n]$, 即 $H \in U^{d \times n}$ 。其中, $h_t \in U^d$, d 为 BiGRU 输出维度。利用式(8)计算句子的整体向量 v_s 。

$$v_s = \frac{1}{n} \sum_{t=1}^n h_t. \quad (8)$$

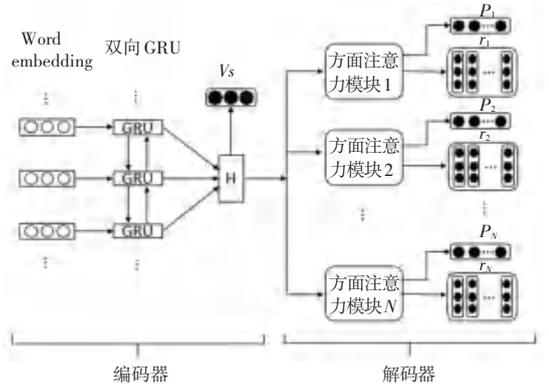


图 3 BiGRU-AAM 模型结构

Fig. 3 BiGRU-AAM model structure

2.2.2 解码器

解码器部分是多个方面注意力模块,分别对应数据集的不同方面。对于包含多个方面的句子,编码获得的输出 H 分别送入对应的方面注意力模块中。方面向量 $v_{a,i}$ 是每个方面注意力模块中的方面信息。各个方面的注意力模块独立执行操作,提取特定方面的隐藏特征。方面注意力模块由注意力部分、概率部分及重构部分组成,结构如图 4 所示。

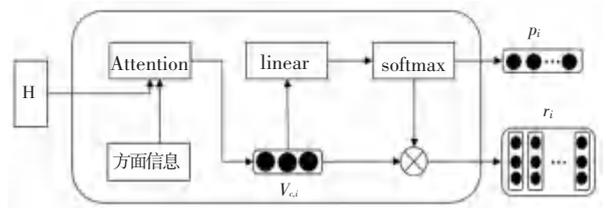


图 4 方面注意力模块

Fig. 4 Aspect attention module

注意力部分将方面向量 $v_{a,i}$ 与输入 H 中的各个隐层状态拼接,然后执行注意力操作。加权平均获得的特定方面文本的向量用 $v_{c,i}$ 表示,计算公式如下:

$$t_{t,i} = \tanh(w_{a,i} [h_t, v_{a,i}] + b_{a,i}), \quad (9)$$

$$a_{t,i} = \frac{\exp(u_{t,i})}{\sum_{j=1}^n \exp(u_{j,i})}, \quad (10)$$

$$v_{c,i} = \sum_{t=1}^n a_{t,i} h_t. \quad (11)$$

其中, i 表示第 i 个方面注意力模块; $W_{a,i} \in U^{d+k}$ 表示注意力的权重矩阵; $b_{a,i} \in U$ 表示注意力的偏置。

概率部分得到各个情感极性的分类概率。特定方面文本的向量 $v_{c,i}$ 经过全连接层,由 softmax 函数计算输入样本在每个分类上的概率并输出。计算公式如下:

$$p_i = \text{softmax}(w_{p,i} v_{c,i} + b_{p,i}), \quad (12)$$

其中, $p_i = [p_{i,1}, p_{i,2}, \dots, p_{i,c}]$; $W_{p,i} \in U^{c \times d}$

是全连接层的权重矩阵; $b_{p,i} \in U^c$ 是全连接层的偏置; c 是类别数。

重构部分对有关特定方面文本的向量 $v_{c,i}$ 进行重构,得到句子整体向量 v_s 在特定方面的情感分类的新向量。通过分类概率对 $v_{c,i}$ 重构,可以获得多个重构文本,重构文本的向量 $r_i = [r_{i,1}, r_{i,2}, \dots, r_{i,c}]$, 计算公式如下:

$$r_{i,j} = p_{i,j} v_{c,i} \quad (13)$$

其中, $r_{i,j} \in U^d, j = 1, 2, \dots, c$ 是类别数。

2.3 模型训练

BiGRU-AAM 网络模型的训练采用 Pytorch 深度学习框架,编程语言为 Python3.6; 运行环境为 Pycharm Ubuntu16 系统,内存大小为 16GB。实验模型的训练使用反向传播算法,利用式(12)中的损失函数优化模型。该损失函数由 2 个折页损失函数组成。 $L(\theta)$ 函数让每个方面注意力模块输出中,真实情感极性概率尽可能大,且让其它类别的概率尽可能小; $J(\theta)$ 函数使每个方面注意力模块输出的重构文本的向量表示中,真实情感极性的向量表示与句子整体向量表示 v_s 趋近。计算过程如下:

$$R(\theta) = L(\theta) + J(\theta), \quad (14)$$

$$L(\theta) = \sum_m \sum_i \sum_{j \neq y_{m,i}} \max(0, 1 - p_{i,y_{m,i}} + p_{i,j}), \quad (15)$$

$$J(\theta) = \sum_m \sum_i \sum_{j \neq y_{m,i}} \max(0, 1 - v_s r_{i,y_{m,i}} + v_s r_{i,j}). \quad (16)$$

其中, m 表示第 m 个输入样本, $y_{m,i}$ 表示第 m 个输入样本在第 i 个方面的真实情感极性。

3 实验与分析

3.1 实验数据

由于方面情感分类任务需要判断其不同的方面的情感极性,实验采用 SemEval2014 任务 4 中的 restaurants 数据集进行。在 restaurant 数据集中,包括 5 个方面和 3 种情感极性,数据中每个句子涉及一个或多个不同方面。实验数据详情见表 1。

表 1 restaurants 数据集

Tab. 1 Data set

方面	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Food	867	302	209	69	90	31
Service	324	101	218	63	20	3
Price	179	51	115	28	10	1
Ambience	263	76	98	21	23	8
Anecdotes	546	127	199	41	357	51
总计	2179	657	839	222	500	94

3.2 实验设置

实验中,词向量采用 word2vec 技术^[15]进行初始化,单词向量的维数、方面嵌入及隐藏层大小均为 300。采用均匀分布 $U(-0.01, 0.01)$ 进行采样,来初始化其他参数。训练的更新规则采用 Kingam 等人^[16]提出的 Adam 规则。训练过程中使用 dropout 机制防止数据的过拟合。批处理大小为 25,初始学习率为 0.01。

3.3 实验结果分析

为验证本文方法的有效性,将本文提出的 BiGRU-AAM 模型与 Bi-LSTM、ATAE-LSTM 和 BLSTM-AAM 等经典情感分析模型,在 restaurant 数据集上进行对比实验。实验结果见表 2。

表 2 不同模型的实验结果

Tab. 2 Experimental results of different models

模型	指标	情感极性			time/s
		Positive	Negative	Neutral	
BiLSTM	精确率/%	90.02	80.24	36.52	1262
	查全率/%	84.15	77.6	52.75	
	F1 值	86.99	78.90	43.16	
ATAE-LSTM	精确率/%	91.21	81.58	42.13	1316
	查全率/%	86.61	77.14	59.62	
	F1 值	88.85	79.30	49.37	
BLSTM-AAM	精确率/%	89.81	81.52	56.32	1582
	查全率/%	90.1	81.72	51.95	
	F1 值	89.95	81.62	54.05	
BiGRU-AAM	精确率/%	90.15	83.72	58.21	1115
	查全率/%	89.96	81.11	52.64	
	F1 值	90.05	82.39	55.29	

实验结果对比可以看出,加入方面信息或使用注意力机制,能明显改善分类效果。体现在评价指标性能方面,就是提高准确率、查全率、F1 值。但是,由于 attention 层突出信息的功能是通过不断加权计算进行的,因此增加了时间代价。BLSTM-AAM 模型的时间代价相对较高。因 BLSTM 神经网络计算相对复杂,增加了计算时间,并且 attention 层在突出重点信息的同时也增加了加权计算时间。BiGRU-AAM 模型在实验的评价指标方面优于 BLSTM-AAM,这两个实验模型均结合了 attention 层,且具有相同的模型结构;不同的是前一个模型第一层采用 BiGRU 神经网络,另一个采用 BiLSTM 神经网络。由实验结果可以看出,由于 BiGRU 神经网络比 BiLSTM 神经网络结构简单,减少了计算及参数,使之达到更快的收敛速度。在时间代价指标

上,从 F1 指标上可以发现,BiGRU 与 Attention 的结合更有效,在方面级情感分类任务中效果更优。

4 结束语

本文提出一种基于 BiGRU 与方面注意力模块的情感分类方法,与广泛使用的 BiLSTM 网络结合 attention 的混合模型相比,本文方法能够提高方面级情感分类的准确率并降低损失率,取得很好的 F1 评价指标,达到了较好的分类效果,有效地降低了模型的训练时间。实验结果表明 BiGRU 结合 attention 的有效性以及本文方法在方面级情感分类的有效性。

参考文献

- [1] WANG Yequan, HUANG Minlie, ZHAO Li, et al. Attention-based LSTM for aspect-level sentiment classification [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2016; 606-615.
- [2] YU Kai, JIA Lei, CHEN Yuqiang, et al. Deep learning: yesterday, today, and tomorrow [J]. Journal of Computer Research and Development, 2013, 50 (9): 1799-1804.
- [3] ZHOU Peng, SHI Wei, TIAN Jun, et al. Attention-based bidirectional longshort-term memory networks for relation classification [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2016; 207-212.
- [4] LIU Quan, LIANG Bin, XU Jin, et al. A deep hierarchical neural network model for aspect-based sentiment analysis [J]. Chinese Journal of Computers, 2018, 41 (12): 2637-2652.
- [5] PENG Zhuliang, LIU Bowen, FAN Cheng, et al. Sentiment classification method based on BLSTM and aspect attention module [J]. Computer Engineering, 2020, 46(3): 60-65, 72.
- [6] VO D T, ZHANG Y. Target-dependent twitter sentiment

- classification with rich automatic features [C]//Twenty-Fourth International Joint Conference on Artificial Intelligence. 2015.
- [7] KIRITCHENKO S, ZHU X, CHERRY C, et al. NRC-Canada-2014; Detecting aspects and sentiment in customer reviews [C]// Proceedings of the 8th International workshop on Semantic Evaluation. 2014; 437-442.
- [8] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification [J]. arXiv preprint arXiv:1512.01100, 2015.
- [9] BAO H, LI S T. Object-level sentiment analysis combining BiLSTM and positional relationship [J]. Information System Engineering, 2018(3):149-151.
- [10] WANGY, HUANGM, ZHAOL. Attention-based LSTM for aspect-level sentiment classification [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016;606-615.
- [11] AUGENSTEIN I, ROCKTÄSCHEL T, VLACHOS A, et al. Stance detection with bidirectional conditional encoding [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA: ACL, 2016;876-885.
- [12] 刘洋. 基于 GRU 神经网络的时间序列预测研究 [D]. 成都: 成都理工大学, 2017.
- [13] 李骁, 黄征. 基于 GRU 网络的互联网信息挖掘 [J]. 信息技术, 2018(3):1-5, 9.
- [14] ROZENTAL A, FLEISCHER D. Amobee at semEval-2018 task 1; GRU neural network with a CNN attention mechanism for sentiment classification [C] // Proc of the 12th International on Semantic Evaluation. [S. l.]: Association for Computational Linguistics, 2018;218-225.
- [15] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality [J]. Advances in neural information processing systems, 2013, 26.
- [16] KINGMA D, BA J. Adam: A Method for Stochastic Optimization [J]. Computer Science, 2014.

(上接第 82 页)

4 结束语

针对自然采样法需要求解复杂的超越方程,规则采样法采样精度低等特性,本文采用了不规则采样的方法,并且通过使用 TMS320F28335 控制器的浮点计算可以快速运算出结果,并在示波器上显示 SPWM 波形。通过实验和仿真,证明变压变频调速的可行性,具有一定的实际意义和理论价值。

参考文献

- [1] 袁雷,胡冰新等. 现代永磁同步电机控制原理及 MATLAB 仿真 [M]. 北京:北京航空航天大学出版社,2016.3.
- [2] 庞丽坤,吴钦木. 三相交流电机 SPWM 变频调速系统设计与仿真 [J]. 智能计算机与应用,2019,9(3):201-205.
- [3] 陈伯时,陈敏逊. 交流调速系统 [M]. 3 版.北京:机械工业出版社,2013.

- [4] 符晓,朱洪顺. TMS320F28335DSP 原理、开发及应用 [M]. 北京:清华大学出版社,2017.
- [5] 黄立培,张学. 变频器应用技术及电动机调速 [M]. 北京:人民邮电出版社,1998.
- [6] 张卿杰,徐友等. 手把手教你学 DSP [M]. 北京:北京航空航天大学出版社,2018.
- [7] 陈坚,康勇. 电力电子学 [M]. 北京:高等教育出版社,2011.
- [8] 王兆安,刘进军等. 电力电子技术 [M]. 北京:机械工业出版社,2009.
- [9] 李祝知非,王林芝,邵璇,等. SPWM 自然采样法和规则采样法的仿真实现 [J]. 电子技术,2018,47(9): 66-68.
- [10] GAVASKAR R B, MAHESWARI L, GANESWARI K A. Performance improvement of single phase inverter using SPWM [J]. IOP Conference Series: Materials Science and Engineering, 2017, 225(1): 012071.