

基于文本信息和层次神经网络的产品评分方法

赵 贇 王中卿 李寿山

(苏州大学计算机科学与技术学院 江苏 苏州 215006)

摘 要 通常点评网站对商品的打分都是通过对商品评论的评分求均值而获得,但是这种方式严重依赖于评论的评分,而且对于评论数较少的商品,这种方式显得不够精确。不同于传统的产品打分机制,文中提出了一种根据产品评论的文本信息对产品进行整体打分的层次神经网络模型,该模型可以从有限的评论中分析出产品较为公正的得分。在产品评论中,存在着[词-句子-评论-商品]的层级结构,因此采用了三层 GRU 的结构分别来对句子、评论、商品进行表示,从而预测商品最终的打分。除此之外,还对评论层进行了额外地输出,进一步提高了商品得分预测的准确率。在回归和分类两种预测任务上的实验结果表明,模型的层次结构对于预测商品得分具有至关重要的作用,同时输出评论的得分可以进一步提高预测的准确率。

关键词 产品打分,评分预测,层次神经网络

中图法分类号 TP391 **文献标识码** A

Product Rating with Text Information and Hierarchical Neural Network

ZHAO Yun WANG Zhong-qing LI Shou-shan

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract Usually, the rating of the product on the website is obtained by averaging the rating of the product review, but this method relies heavily on the rating of reviews, which is not accurate enough for products with fewer reviews. Different from the traditional product scoring mechanism, this paper proposed a hierarchical neural network model for the overall scoring of products based on the text information of them, which can analyze the fair scores of products from limited reviews. In the product review, there is a hierarchical structure of [word-sentence-review-product], so the structure of three-layer GRU is used to get the representations of the sentences, reviews and products separately, so as to predict the final score of the product. In addition, this paper also makes additional output to the review layer to further improve the accuracy of the prediction. Experiments on the two prediction tasks of regression and classification show that the hierarchical structure of the model plays a crucial role in predicting the score of the product, and the score of outputting comment can further improve the prediction accuracy.

Keywords Product rating, Rating prediction, Hierarchical neural network

1 引言

随着在线点评网站逐步走进大众生活,人们乐意在这些网站上分享个人的观点和评价,同时又希望能够快速地了解某件产品或商户的综合评价。一般的网站对每个产品都提供了一个综合的评分,通常为 1 到 5 的分值,来代表评论过该商品的用户的综合评价,这些分值通常通过对所用评论的打分求均值获得。对于评论数目比较多的产品,这样的评分方式是比较简单有效的;然而对于评论数据比较少的商品,往往显得不够精确。其中一个原因可能是评论的打分很容易受到用户主观性的影响,其实际打分往往和用户的评分有所偏差。其次,评论的得分可能存在缺失或者与评论实际内容严重不符的情况,那么这样的得分是不可信的。一种很直观的想法是,直接通过浏览某个商品所有的评论来对该商品进行评分,这样的得分相对来说比较客观。

评论的评分预测可以视为文档级别的情感分析,而根据

文本信息对商品进行打分更侧重全局的情感。大多数现有的研究主要是预测评论的情感极性,Pang^[1]通过一些机器学习的方法把评论的评分预测视为一种多分类或是回归问题,通过特征工程的方式,利用文本特征或用户特征来进行预测。然而,用户对产品的评价很可能是多方面的,因此也有一些工作主要集中于属性级别的情感分析^[2]。近年来,神经网络结构在情感分析和评分预测邻域变得越来越流行,例如使用卷积神经网络(CNN)进行情感分析^[3],利用长短时记忆网络(LSTM)进行情感分类^[4]。

虽然目前也有一些利用用户评论对产品进行打分的研究,但他们通常忽略了评论的层次结构关系,然而层次结构对于预测全局情感往往是重要的,例如一句话的情感强度通常由该句子内部不同程度的情感词决定,这些词的情感强度通常只是与该句子有关,而不一定能决定全局情感;又比如对于某个评论,用户可能对于产品的不同方面表现出不同的情感倾向。本文的模型并不需要将所有评论的词级别的信息全部

本文受国家自然科学基金青年项目(61806137)资助。

赵 贇(1995-),男,硕士生,主要研究方向为自然语言处理,E-mail:yzhao666@stu.suda.edu.cn;王中卿(1987-),男,博士,主要研究方向为自然语言处理,E-mail:wangzq.antony@gmail.com(通信作者)。

纳入最终产品评分的考察中,而是先获取局部表示,再获取全局表示,这样能够有效避免最终的情感预测引入其它不必要的噪声。鉴于此,本文提出了一种基于层次门控循环单元(GRU)结构的产品评分预测模型,希望通过三层 GRU 来获取对句子、评论以及商户(产品)的表示,从而最大程度地保留了词、句子、评论、商户之间的从属关系。实验表明,本文使用的模型对于产品评分预测的效果要明显好于抛弃层次结构而直接将产品评论的所有词作为特征的传统机器学习模型。

本文第 2 节简单介绍了相关工作;第 3 节介绍了模型每一层的结构并阐述了模型的基本原理;第 4 节介绍了实验的数据集和多种基准模型以及它们的参数设置,同时也对最终的实验效果进行了比较和分析;最后总结全文。

2 相关工作

产品评分预测方法经常涉及到个性化推荐领域^[5],但本文主要研究如何通过用户评论文本信息来进行产品打分,并不涉及用户的一些结构化属性特征。一些研究工作主要集中在评论的情感分类或是评分预测上,Ifrim 等^[6]提出了一种 bag-of-opinions 产品评论得分预测方法,克服了 n-gram 特征稀疏性的局限性,并提出了一种基于约束的岭回归算法来对每个意见进行打分。Fang 等^[7]对亚马逊产品评论数据进行了句子级别和文档级别的评分预测,并对朴素贝叶斯、支持向量机、随机森林这 3 种方法进行了深入的比较。Tang^[8]提出了一种基于神经网络的评分预测模型,不仅考虑了用户评论的文本特征,同时考虑了不同用户对于情感词的强度,结果表明结合用户属性要比单独用文本特征的预测效果要好。Li 等^[9]提出了一种新颖的评分预测模型,不光利用了评论的文本信息,同时结合了用户与产品的特征,并利用张量分解技巧避免了数据的稀疏性问题;实验表明,对于那些不流行的产品或不活跃的用户相关的评论,模型的预测效果最好。对于直接利用评论来对产品或商户进行评分,目前的研究还比较少,产品的打分一般都是通过产品评论的打分获得,然而评论的打分很容易受到用户主观性的影响,Fan 等^[10]在 Yelp 数据集上利用评论文本中的高频词以及词性作为特征,多种机器学习方法的对比实验表明,线性回归的预测效果比较好。

采用层次模型对文档段落进行表示的方法已经被广泛使用,例如,Xie 等^[11]提出了一种基于层次结构的多策略情感分析模型,使用支持向量机对微博数据集中每个句子的情感极性进行了分类,并获得了不错的效果。目前使用比较多的还是基于神经网络的层次模型,Li 等^[12]利用层次 LSTM 模型训练一个段落或文档自编码器,利用层次 LSTM 来对段落进行重构,结果表明层次 LSTM 模型能够有效地表示文档以及段落,结合注意力机制可以提升表示能力。Yang 等^[13]使用两层 GRU 在词级别和句子级别进行表示,同时结合了两层注意力机制来对文档进行分类。Ruder 等^[14]使用了双向的 LSTM 来对评论进行属性级别的情感分析,结果表明层次模型能够充分利用评论句子的结构和上下文信息提高预测的准确率。

3 层次 GRU 评分模型

3.1 任务描述

一般而言,点评网站对某个商家或某件产品的打分都是通过该商家或产品的所有相关评论的打分求均值获得。由于

某些商品的评论个数有可能比较少,而且评论内容与评论得分的也有可能不一致,甚至评论的评分是缺失的,由此得到的评分往往不够精确,其可靠性得不到保障。因此,我们希望浏览某个商品的所有评论的文本信息来对该商品进行打分,而不是依赖固有的评论的评分。商品一般包含多条评论,每条评论包含多个句子,每个句子包含多个词,因此我们希望模型能够保留这种层次结构。很自然地,我们采用了层次 GRU 模型。由于 RNN 的记忆性,LSTM 和 GRU 都适合处理顺序序列且能避免梯度消失问题,但 GRU 参数更少、更容易收敛,因此文中采用了三层 GRU 来对产品进行评分预测,依次从词级别、句子级别、评论级别的表示来探讨对生成评分的影响。图 1 是模型的示意图,下面从 3 个层面解释模型的结构。

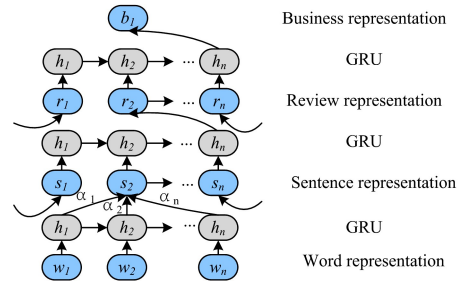


图 1 模型的结构

3.2 词级别表示

本文采用 GRU 来对句子进行表示,当然也可以利用 CNN 在短语级别对句子进行表示。将句子中的词输入到词嵌入层,每个词可以表示成一个低维的向量。通常词汇表中的词以 one-hot 形式存储,其词向量可以通过词嵌入矩阵 $W^{(w)} \in \mathbb{R}^{d \times V}$ 的每一列获得,其中 d 表示词向量的维度, V 表示词汇表的大小。假设句子 $S = (v_1, v_2, \dots, v_n)$,其中 $v_i \in \mathbb{R}^V$,则第 i 个词的词向量表示如下:

$$x_i = W^{(w)} v_i \quad (1)$$

则 $S = (x_1, x_2, \dots, x_n) \in \mathbb{R}^{d \times n}$ 。得到词的表示后,我们将整个句子序列输入到第一个 GRU,从而得到句子级别的表示。

3.3 句子级别表示

输入序列中的每个元素,即句子中的每个词向量,都会输入到 GRU 单元进行循环计算,公式如下:

$$z_t = \sigma(W^{(z)} x_t + U^{(z)} h_{t-1}) \quad (2)$$

$$r_t = \sigma(W^{(r)} x_t + U^{(r)} h_{t-1}) \quad (3)$$

$$\tilde{h}_t = \tanh(W x_t + r_t \circ U h_{t-1}) \quad (4)$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \quad (5)$$

其中, x_t 为第 t 个时间步输入的词向量, h_{t-1} 为上一次计算后的隐层输出,它们分别与权重矩阵 $W^{(z)}$ 、 $U^{(z)}$ 相乘,即线性变换后相加,再输入到 sigmoid 激活函数进行线性化,即公式中的 σ 。我们将 z_t 称为更新门,更新门决定过去的记忆信息有多少需要被传递到将来。类似地,我们将当前的输入向量和上个时间步的输出 h_{t-1} 经过另一个非线性变换,我们称 r_t 为重置门,其主要决定哪些过去残留的信息需要被遗忘。然后,上一步的输出经过线性变换后与重置门 r_t 进行哈达马积,公式中用 \circ 表示,与当前输入经过线性变换后的结果相加,再经过双曲正切激活函数来决定保留与遗忘的信息。最后,更新门 z_t 在上一步的输出 h_{t-1} 与当前记忆 \tilde{h}_t 之前做了一个权衡,计算的结果为当前 GRU 单元的输出 h_t 。

一般来说,一句话中每个词的重要性都是不一样的,因此

我们考虑在词级别添加注意力机制,利用注意力机制可以学到不同情感词在某个句子中的权重分布。一句话中的每个词依次输入到 GRU 后,GRU 在每个时间步的输出可以视为该词的表示,我们把所有时间步的输出按照注意力的权重进行加权求和后的表示视为该句子的最终的表示:

$$s_i = \sum_{j=1}^{l_i} \alpha_j^i h_j^i \quad (6)$$

其中, l_i 表示第 i 个句子的长度; α_j^i 表示第 j 个词的重要性,其为一个 0 到 1 之间的数值表示该词的重要程度; h_j^i 表示 GRU 第 j 个时间步的输出。 α_j^i 可由如下公式计算得到:

$$\alpha_j^i = \frac{\exp(e(h_j^i))}{\sum_{k=1}^{l_i} \exp(e(h_k^i))} \quad (7)$$

其中, e 为相似度计算函数,它将词表示映射成一个数值,然后对其进行了 softmax。对于相似度计算函数,我们简单地采用了一层非线性变换:

$$e(h_j^i) = v^T \tanh(W h_j^i + b) \quad (8)$$

其中, W 、 b 分别为权重矩阵和偏置, v 表示权重向量, V^T 表示其转置。

3.4 篇章级别表示

上节我们已经得到了句子的表示,一条评论一般有多个句子,我们将某条评论中的所有句子表示输入另一个 GRU。从图 1 可以看出,我们只是将 GRU 的最后一个时间步的输出作为该条评论的表示 R ,因为 GRU 最后时间步的输出保留了其之前的记忆。实际上,有些论文中采用了不同的操作,例如将所有时间步的输出的每一维求均值或求最大值,用最后的结果表示整个评论。一个商家或产品通常包含多条评论,我们有了评论的表示后,再将该商户的所有评论输入到第三个 GRU,类似地,用最后时间步的输出来表示该商家或产品,最后将该商家的表示 B 输入一层全连接层,如果是回归预测,则全连接层直接输出预测值;如果是分类,则全连接层输出多个标签,再经过 softmax 计算每个标签的概率。需要说明的是,图 1 中在句子级别以及评论级别并没有应用注意力机制,在下文的实验中我们也列举了这两种情况的实验效果。

除此之外,实际的评论往往自带打分,为了利用这些已存在的得分,我们可以在评论层同时输出评论的打分,从而模型变成了一个多输出模型,分别对评论和商品进行了得分预测,模型的结构图如图 2 所示。

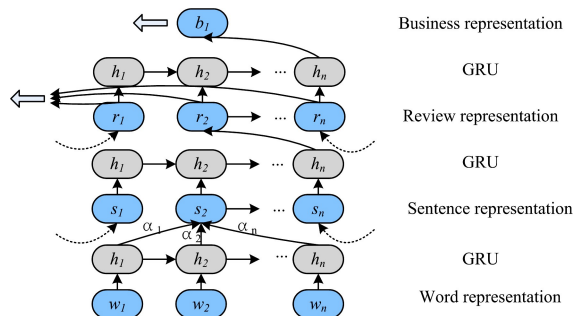


图 2 多输出模型

4 数据集及实验参数设置

4.1 数据集及预处理

本文使用了 Yelp¹⁾ 点评网站的评论数据集,我们重新计

算了每个商户的评论个数,并在这些商户中选取了评论数为 10 条的商户,共 5 027 个,其中划分为训练集 4 021 个和测试集 1 006 个。数据集中评论所包含的句子的长度最小为 1,最大为 2 604,平均长度为 80。我们限制了每条评论的句子数最多为 20,每个句子的词的个数最多为 30,下文实验中用于比较的其他机器学习模型则不加这个限制。我们选择了训练集中所有的词作为词特征,共 35 865 个。为了充分比较模型的预测效果,我们采用了两种度量指标(即回归预测评分(小数)或分类)进行预测。数据集评分统计如图 3 所示,可以看出数据集中各个评分的个数是不平衡的,集中在 4 分,一共有 9 种评分,因此可以看成是一个 9 分类问题。

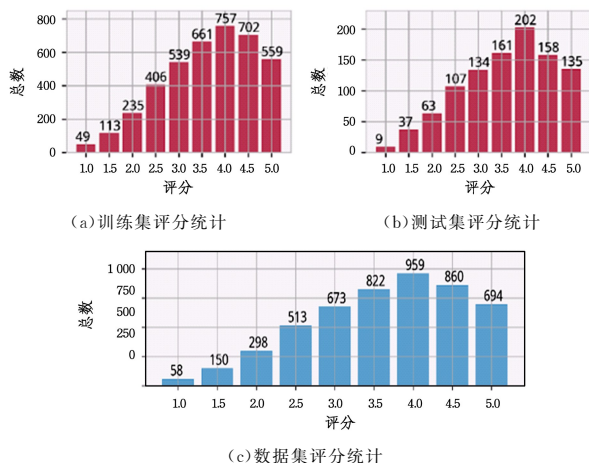


图 3 数据集柱状图

4.2 对比方法及参数设置

对于回归问题,我们同时提供了多个基准模型进行比较,这些模型没有考虑词、句子、评论以及商家之间的层次结构,只是简单地用某个商家的所有评论中的所有词来表示这个商家。除了本文提出的模型之外,其他比较模型都采用了所有词 TF-IDF 来构建特征表示商家,文档 j 中第 i 个词的权重可通过式(9)计算得到。

$$\omega_{i,j} = tf_{i,j} \times \log \frac{N}{df_i} \quad (9)$$

其中, $tf_{i,j}$ 表示文档 j 中第 i 个词出现的频数, df_i 表示包含第 i 个词的文档的个数, N 表示所有文档数目。下面列举了本文中用于比较的几种基准模型:

- (1)Lasso:使用 Lasso 回归进行评分预测。
- (2)KNN:使用 K 近邻回归进行评分预测,K 近邻数设置为 7,我们对近邻点按距离进行了加权,距离预测目标越近的近邻的权重越大。
- (3)Decision Tree:决策树回归,最大深度设置为 5,叶子节点最少的样本数设置为 16。
- (4)Random Forest:使用随机森林进行回归,决策树数目为 40,叶子节点最少样本数为 5。
- (5)Linear:线性回归。
- (6)Ridge:岭回归。
- (7)SVR:支持向量回归,采用线性核。
- (8)GRU:直接将某个产品的所有评论全部输入到一个 GRU 中预测最终的打分。

¹⁾ <https://www.yelp.com/dataset>

(9)H-GRU:3层GRU,每层采用最后时间步的输入进行表示。

(10)H-GRU(1-att):3层GRU,并在词级别应用注意力机制,即图1描述的结构。

(11)H-GRU(2-att):3层GRU,并在词级别和句子级别应用注意力机制。

(12)H-GRU(3-att):3层GRU,并分别在词级别、句子级别和评论级别应用注意力机制。

(13)H-GRU(multi):3层GRU,同时增加了评论部分的输出。

同样地,我们也采用了多个分类模型尝试对商家进行分类,用商家的所有评论的所有词构建特征表示该商家。下面列举了本文中用于比较的分类模型:

(1)KNN:K近邻分类,K近邻数设置为20,对近邻点距离进行了加权。

(2)Decision Tree:采用决策树进行分类,最大深度设置为6,最少叶子节点样本数为1。

(3)Random Forest:采用随机森林进行分类,决策树数目为45,叶子节点最少样本数为2。

(4)Logistic:逻辑回归,采用liblinear优化损失函数,采用一对多分类并按照训练样本数目计算每个类的权重。

(5)Naive Bayes:朴素贝叶斯分类器,采用多元伯努利模型,平滑系数 α 设置为0.007。

(6)SVM:支持向量机,采用线性核,惩罚系数C设置为1.5,同样计算了每个类的权重。

(7)Ensemble:对以上6种分类模型按照投票策略进行了集成,同时对以上模型预测标签的次数设置了权重,分别为0.05,0.1,0.2,0.3,0.3,0.05。

普通的GRU模型与层次GRU模型与回归部分描述的类型,只是将最后一层全连接层换成了预测多分类问题。

4.3 目标函数及参数优化

对于回归问题,层次GRU模型的输出是1~5的值,对于分类问题,层次GRU的输出是多个类的概率值,两种模型的超参数保持一致。词嵌入的维度设置为100,我们使用了Glove¹⁾预训练的词向量进行初始化并随模型训练,隐层维度统一设置为100,批量更新数目为64,学习率设置为0.001,梯度裁剪设置为5,并使用了Adam^[15]优化器。回归损失函数采用均方误差损失函数:

$$Loss_b(\theta) = \frac{1}{|B|} \sum_{b \in B} (label(b) - predict(b|\theta))^2 \quad (10)$$

其中,B表示所有商家的集合,label(b)表示某个商户的真实评分,为模型需要训练的参数。分类模型采用交叉熵损失函数:

$$Loss_b(\theta) = - \sum_{b \in B} label(b) \cdot \log \frac{e^{x_c}}{\sum_{j=1}^N e^{x_j}} \quad (11)$$

其中, $x \in \mathbb{R}^N$, x 表示模型的输出,其没有经过softmax,N表示输出的维度, c 是一个0到C-1之间的标量,代表所属的类别。对于多输出模型,我们简单地将所有评论得分的损失和该产品得分的损失进行了相加来代表总的损失。

4.4 回归结果分析

表1列出了回归模型预测的结果,可以看出在传统的回归方法中,线性回归、岭回归以及支持向量回归均能取得相对不错的效果。层次GRU模型的效果相对于其他方法具有十分明显的提高,最主要的原因还要归功于模型的结构与表示以及深度学习的计算能力,例如词向量能够保留词之间的语义相似性,循环神经网络适合语言模型。最主要的是,本文模型保留了词、句子、篇章之间的从属结构,这非常符合实际情况以及人的认知,传统的模型则完全抛弃了这种结构,由于本文模型的结构达到三层,所以这种结构差异带来的效果差异非常明显。从表中可以看出,直接将所有评论输入到一个GRU的效果仅仅比SVR好一点,完全比不上层次模型。其次,我们比较了每层GRU的输出添加注意力机制后的效果,发现在词级别添加注意力机制可以一定程度上改善准确率,然而在句子级别和评论级别反而不如不加的效果。对评论进行多输出可以明显地提高预测的准确率,取得了目前最好的结果。

表1 回归预测结果比较

模型	均方误差(MSE)
Lasso	0.985
KNN	0.752
Decision Tree	0.493
Random Forest	0.336
Linear	0.242
Ridge	0.222
SVR	0.214
GRU	0.193
H-GRU	0.130
H-GRU(1-att)	0.129
H-GRU(2-att)	0.133
H-GRU(3-att)	0.137
H-GRU(multi)	0.097

4.5 分类结果分析

表2列举了分类模型的准确率比较结果,由于数据集评分划分得比较细,因此这是一个9分类问题,而不是通常的5星分类。而且由于评分是人为的,存在一定程度的随意性,因此本文实验的准确率并不是非常的高,但足以说明各个模型之间的差异性。可以看出,传统分类模型的差异性并不大,逻辑回归、朴素贝叶斯以及支持向量机的分类效果要好一些,集成的效果比它们要好一些,但是和层次GRU模型相比差异还是比较大的。同样地,我们发现只在词级别添加注意力机制的效果最好,在句子级别和评论级别添加注意力的准确率反而下降了。其原因可能是注意力层可能会更偏向于记住固定的信息,对于词级别来说,输入第一层GRU的是一些比较固定的词向量,不同句子对于同一个GRU的输出也可能具有类似的部分,比如不同的句子可能包含相似的情感词和短语等等,所以这层的注意力机制能够识别出这些关键的部分。然而,对于句子级别甚至评论级别来说,句子的表示或评论的表示变得更加抽象了,不同的句子表示之间或不同的评论表示之间可能不具备相似性,因此注意力层难以识别更重要的部分。对于评论级别的表示来说,评论之间的顺序或重要性似乎并没有什么侧重点,所以准确率并不会得到改善。对于

¹⁾ <https://nlp.stanford.edu/projects/glove/>

句子级别来说,或许句子之间的顺序以及重要性是存在的,但在本文的实验中,效果并没有显著提升。另一方面,循环神经网络需要综合考虑所有时间步输入的综合情感,所以对于句子级别和评论级别来说,最后一个时间步的输出可能更具有代表性。从表 2 中依然可以看出,层次 GRU 明显要比普通的 GRU 效果要好。为了比较不同层次注意力机制的表现,图 4 显示了层次 GRU 模型在训练过程中的损失以及测试集上的准确率,其中训练集上的损失基本变化不大,层次 GRU 在词级别加入注意力机制的损失要稍微小一些。可以明显看出,模型在第 5 个 epoch 后准确率开始趋于稳定,在词级别加入注意力机制的效果最好,加入 3 层注意力的效果比其他要差。由此可见,注意力机制的效果并不明显,仅仅反映在词级别上,而多输出模型的效果比较明显,同样取得了目前最好的效果,这通常是由于多个任务之间可以在公共层共享某些特征,提高了模型的泛化能力,减少了过拟合的风险。

表 2 分类的准确率比较

模型	准确率
KNN	0.233
Decision Tree	0.296
Random Forest	0.328
Logistic	0.336
Naive Bayes	0.338
SVM	0.357
Ensemble	0.378
GRU	0.442
H-GRU	0.544
H-GRU(1-att)	0.556
H-GRU(2-att)	0.553
H-GRU(3-att)	0.535
H-GRU(multi)	0.594

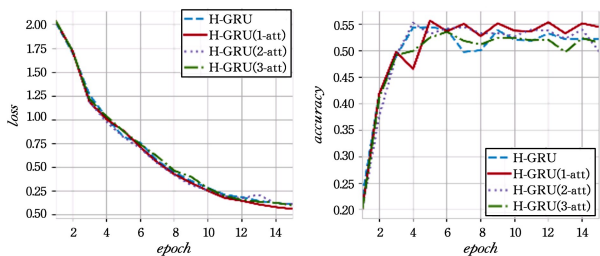


图 4 模型训练损失及准确率

结束语 本文首次应用一种 3 层 GRU 模型来对商品或商家进行表示,通过利用某个商家或产品的所有评论的文本信息来对该商家或产品进行打分。实验表明,层次 GRU 模型能够保留原始的词、句子、评论之间的层次关系,其评分预测的准确率明显超过了抛弃层次结构的传统机器学习模型,说明层次结构对于预测产品打分十分重要;同时,如果数据集中评论存在打分,利用多任务学习模型可以大幅度提高准确率。在下一步的工作中将研究如何将用户信息融入注意力机制或采用更有效的语义表示模型对句子建模。

参考文献

- [1] PANG B, LEE L. Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales[J/OL]. <http://arXiv.org/abs/cs/0506075>.
- [2] THET T T, NA J C, KHOO C S G. Aspect-based sentiment analysis of movie reviews on discussion boards[J]. *Journal of Information Science*, 2010, 36(6): 823-848.
- [3] SEVERYN A, MOSCHITTI A. Twitter Sentiment Analysis with Deep Convolutional Neural Networks[C]// *International ACM SIGIR Conference*. ACM, 2015: 959-962.
- [4] TANG D, QIN B, FENG X, et al. Target-Dependent Sentiment Classification with Long Short Term Memory[J/OL]. <http://arXiv.org/abs/1512.01100v1>.
- [5] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// *International Conference on World Wide Web*. ACM, 2001: 285-295.
- [6] IFRIM G, WEIKUM G. The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns[C]// *计算语言学国际学术会议*. 2010: 913-921.
- [7] FANG X, ZHAN J. Sentiment analysis using product review data[J]. *Journal of Big Data*, 2015, 2(1): 1-14.
- [8] TANG D, QIN B, YANG Y, et al. User modeling with neural network for review rating prediction[C]// *International Conference on Artificial Intelligence*. AAAI Press, 2015: 1340-1346.
- [9] LI F, LIU N, JIN H, et al. Incorporating reviewer and product information for review rating prediction[C]// *International Joint Conference on Artificial Intelligence*. AAAI Press, 2011: 1820-1825.
- [10] FAN M, KHADEMI M. Predicting a Business Star in Yelp from Its Reviews Text Alone[J/OL]. <http://arXiv.org/abs/1401.0864>.
- [11] XIE L X, ZHOU M, SUN M S. Hierarchical Structure Based on Hybrid Approach to Sentiment Analysis of Chinese Micro Blog and Its Feature Extraction[J]. *Journal of Chinese Information Processing*, 2012, 26(1): 73-84.
- [12] LI J, LUONG T, JURAFSKY D, et al. A Hierarchical Neural Autoencoder for Paragraphs and Documents[J]. *International Joint Conference on Natural Language Processing*, 2015: 1106-1115.
- [13] YANG Z, YANG D, DYER C, et al. Hierarchical Attention Networks for Document Classification[C]// *North American Chapter of the Association for Computational Linguistics*, 2016: 1480-1489.
- [14] RUDER S, GHAFARI P, JOHN G. A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis[J/OL]. <http://arXiv.org/abs.1609.02745>.
- [15] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[J/OL]. <http://arXiv.org/abs/1412.6980>.