文章编号：2095-2163(2019)03-0300-04 中图分类号：TP391.41 文献标志码：A

基于时间序列网络的谣言检测研究

任文静，秦 兵，刘 挺
（哈尔滨工业大学 计算机科学与技术学院，哈尔滨 150001）

摘 要：本文主要研究了 GRU，LSTM 等深度学习模型在谣言检测上的应用，判断微博文本是否为谣言类信息。考虑到新浪微博平台的图结构，一条微博文本对应着多条评论信息，评论中可能包含对该条文本的态度，例如赞成、反对、怀疑等。因此，本文在判断微博文本是否为谣言时，融合了评论信息，将评论看作一条时间线上的各个时刻，按照时间节点展开，作为时间序列模型每个时刻的输入，并且利用注意力机制衡量每个时间节点对最终语义表示的重要程度。实验结果表明，在加入评论信息及 attention 机制后，实验结果具有明显提升，最后达到 92.66%的识别准确率。

关键词：谣言检测；深度学习；新浪微博；分类

Rumor detection based on time series model

REN Wenjing, QIN Bing, LIU Ting
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

【Abstract】 This paper mainly studies the application of GRU and LSTM model in rumor detection research, to judge whether a microblogging text is rumor information or not. Considering the graph structure of sina microblogging, that is to say, a microblogging text corresponds to a number of comments, and comments may contain attitude towards this text, such as favor, opposition, doubt. This paper integrates comments information to rumor detection, regards comments as time nodes on a time line. After unfold by time, every comment can be treated as the input of time series model at every time. The paper also proposes attention mechanism, which weighs importance degree of every time node to final semantic representation. The experimental results show that, after adding comment information and attention mechanism, result improves significantly, finally reaches 92.66%.

【Key words】 rumor detection; deep learning; sina microblogging; classification

0 引 言

目前，国内外研究者基于 twitter、新浪微博平台中的谣言信息已经展开了丰富的研究工作，从不同的角度着手构建谣言检测模型。大多数的研究都是将其看作分类任务，利用带标签的数据集进行有监督的学习。但构造特征工程费时费力，并且需要一定的专业背景知识。

相比于传统的机器学习方法，深度学习模型可以自动学习数据集中蕴含的特征，摒弃了繁琐的特征构造过程，无需掌握过多的领域背景知识，在一定程度上简化了设计开发步骤[1]。在本次研究中，主要利用 GRU、LSTM 模型对微博文本进行建模，考虑到评论信息对谣言检测具有重要的影响，评论文本中包含着否定、怀疑、肯定等态度。因此研究中将利用注意力模型对评论进行建模。评论数众多，采用分块的方式对评论进行划分，作为时间序列模型每个时刻的输入，并引入 attention 机制，衡量每个时间块对最终语义表示的影响程度。对此，将给出设计论述如下。

1 深度学习模型研究

1.1 LSTM 和 GRU 网络探析

长短期记忆网络（LSTM）[2] 是一种特殊的 RNN，通过内部的结构设计可以避免 RNN 的梯度消失问题，并且相比于 RNN 模型，能够记住更长远的信息。包含 3 个门结构，分别为输入门、输出门、遗忘门，可以去除或者增加长期信息，刻画长远信息对当前细胞状态的影响程度，衡量当前输入及长远信息对当前细胞状态的影响程度的差异性。每个门的计算公式可表示如下：

$$i_t = \sigma_i(W_i[h_{t-1}, x_t] + b_i);$$

$$\widetilde{C_t} = \sigma_C(W_C[h_{t-1}, x_t] + b_C);$$

$$f_t = \sigma_f(W_f[h_{t-1}, x_t] + b_f);$$

作者简介：任文静（1993-），女，硕士研究生，主要研究方向：自然语言处理；秦 兵（1968-），女，博士，教授，博士生导师，主要研究方向：自然语言处理；刘 挺（1972-），男，博士，教授，博士生导师，主要研究方向：自然语言处理、文本挖掘、文本检索等。
通讯作者：任文静 Email：wjren@ ir.hit.edu.cn

收稿日期：2017-06-12

$$o_t = \sigma_o(W_o[h_{t-1}, x_t] + b_o);$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t};$$

$$h_t = o_t * tanh(C_t). \quad (1)$$

其中，输入门 i_t 表示当前新的输入对细胞状态的影响程度，决定利用多少新信息对细胞状态进行修改，代替旧的信息；遗忘门 f_t 表示从先前的细胞状态中丢弃多少信息，0 表示舍弃，1 表示保留；输出门 o_t 确定更新后的细胞状态将有多少信息输出。

GRU[3]（Gated Recurrent Unit）由 Cho 等人于 2014 年提出，是 LSTM 模型的一个变体。LSTM 包含 3 个门以及当前细胞状态的计算，参数较多，收敛较慢，训练时间较长。GRU 对 LSTM 进行了简化，将遗忘门和输入门合并为更新门，又引入了重置门，一定程度上加快了训练速度，减少了模型参数，并且不会降低模型效果。GRU 内部计算公式的数学表述如下：

$$z_t = \sigma_z(W_z[h_{t-1}, x_t] + b_z);$$

$$r_t = \sigma_r(W_r[h_{t-1}, x_t] + b_r);$$

$$\widetilde{h_t} = tanh(r_t \circ U h_{t-1} + W_h * x_t);$$

$$h_t = (1 - z_t) \circ \widetilde{h_t} + z_t \circ h_{t-1}. \quad (2)$$

其中，重置门 r_t 决定前一个时刻隐含层状态 h_{t-1} 对当前细胞状态 $\hat{h}_t$ 的影响程度，如果先前的状态对当前状态毫无影响，那么理论上，重置门 r_t 会完全屏蔽先前的信息。更新门 z_t 决定是否将先前的记忆进行后传，衡量先前记忆对未来信息的影响程度。如果 z_t 取值为 1 时，表示先前记忆完全不进行删减地后传；如果 z_t 等于 0，则表示只传递当前细胞状态，认为未来信息只与当前时刻相关，与先前的记忆都没有关系。

1.2　注意力模型探析

注意力模型起初用于编码解码模型中。编码解码模型的基本思想可描述如下：编码的过程是将输入序列 x 转化为固定长度的向量，解码的过程根据固定长度的向量以及之前预测出的词语生成输出序列，是一种端到端的学习过程。编码器、解码器选择自由，可以利用 RNN、LSTM、GRU、CNN 等深度学习模型的任意自由组合。

编码解码[4-5]模型虽然在多种任务上已取得了较为可观的研究效果，但依然存在一定的局限性。在模型编码的过程中，将输入信息压缩到固定长度的实数向量中，可能无法获得完整的文本表示语义。而且，在解码某个词时，只利用到了编码过程的最终表示，即固定长度的向量表示，而并未考虑到特定输入对当前解码的影响。这种局限性对于机器翻译、序列标注等任务来说，将显著降低模型的设计处理性能。

注意力模型[6]可以解决上述局限与不足。通过引入注意力，在解码时，不单单利用固定长度的向量表示，还将关注到每一个输入对当前预测值的影响。每一步预测时计算输入的影响程度，可以充分利用输入序列携带的信息，进而在解码过程中，输入序列的每个词都将对待预测词的选择产生影响。

注意力模型应用到分类任务中，对输入序列学习语义表示时，不再使用最后一个隐含层的输出作为特征表示，而是将每个词的重要程度融合进整个输入序列的语义表示中，更加直观清楚地解释了输入序列中的每个词对分类任务的影响程度，及对该任务的重要程度。

2　基于时间序列网络的谣言检测

2.1　基于 LSTM、GRU 的微博文本表示

GRU 模型在谣言检测中的应用如图 1 所示。在图 1 中，$w_1, \cdots, w_n$ 是微博文本中的每个词，通过查询词向量表，可以得到该词的分布式词向量表示。而经由神经单元的计算，则可以得到一系列隐含层的输出，并将前一时刻隐含层的输出和当前输入 w_i 作为当前时刻神经单元的输入。$h_1, \cdots, h_n$ 是 GRU、LSTM 模型隐含层的输出。接下来，研究将对所有隐含层的输出按维度进行均值计算或是取每一维的最大值（MAX pooling），作为句子的语义表示。最后，研究即将该语义表示作为最终分类的特征，送入 Softmax 分类器，判断是否为谣言。

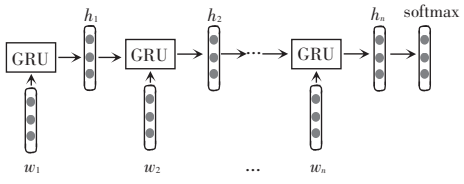

图 1　GRU 模型图

Fig. 1　GRU network

本节将 LSTM、GRU 模型应用到微博文本上，研究目的旨在判断当不引入其它资源的前提下，且仅是使用微博文本，深度学习模型能否利用深层语义分析识别谣言，能否学习到一些类似于主题、情感分布等特征，或者挖掘语言习惯判断是否为谣言，能否与基于手工抽取特征的传统机器学习方法进行抗衡。实验结果可见后文 3.2 节内容。

2.2　基于注意力模型的评论表示

在时下研究中，探讨可知只靠微博文本是远远

不够的,语义信息并不充足,同时仅凭一条文本,人类也很难判别真假。通常,微博的评论者会对信息的真实度进行肯定或否定,对于谣言检测任务来说,评论内容也具有至关重要的作用。在微博文本的传播周期中,不同的评论者会发表不同的意见或看法,这些信息可以帮助研究者甄别内容的真假。针对一条微博文本的所有评论,研究也可以将其看作是与时间序列相关的。在该条微博的传播周期中,每一时刻对应不同的用户,每个用户产生了不同的评论。当研究将所有的评论平铺到一条线上时,每条评论就对应于每个时间节点的输入信息,所有评论构成了整个输入空间。因此,研究中就可以利用时间序列模型对所有评论进行建模,将每条评论视作 LSTM 每一时刻的输入,学习评论间的相互影响及整个评论的语义表示,此时,模型待处理的时间序列长度是评论数量大小。

通过观察语料还将发现,针对一条微博文本,最大评论数可达 5 万条。虽然 LSTM 具有长依赖关系,但也不能学习到如此范围规模的知识,而且在后期的学习过程中会逐渐忘记先前的东西。因此,研究中将考虑对评论进行划分。划分后得到的块作为 LSTM、GRU 模型每一时刻的输入,减少时间序列的长度,降低模型复杂度。块与块之间依然存在时间上的顺序关系,前后互相影响,而且也依然可以利用时间序列模型对其实现建模。

与此同时,研究后又得知评论也具有爆发期,即在不同的时间段,评论的增加或衰减程度是不同的,故而在对评论进行划分时若能捕捉到这种评论的爆发期及衰减期,将使得划分更趋精准,如此划分后每个块内的内容表意可能更加相近,持有相同的观点,对爆发期的评论也能进行更细致的划分。Ma 等人[7]提出一种动态划分方式。与均等划分不同的是,在动态划分过程中,时间间隔将随着样本密度不断变化的,样本划分后的块数并不固定。

在本次研究中,则将微博文本作为第一个块的内容,即时间序列模型初始时刻的输入。考虑到每个块对谣言检测的影响程度都各不相同,设计时在模型中引入注意力机制,获取那些对文本分类有重要影响的块,并且增大这些块的权重,从而改善样本的表示。基于 GRU 的注意力模型如图 2 所示。

输入是划分成块的评论样本,每个时间节点对应一个评论块,利用 GRU 模型,结合当前输入及前一个时刻隐含层的输出学习当前时刻隐含层的输出 h_{it}。输入序列的样本表示不再是最后一个隐含层的输出表示,而是利用数学公式计算得到每个隐含层的输出权重 ∂_{it},整个评论样本的语义表示为所有隐含层输出值的加权和。研究推得各数学运算公式具体如下:

$$u_{it} = tanh(W_w h_{it} + b_w);$$

$$\partial_{it} = \frac{\exp(u_{it}^{\mathrm{T}} u_w)}{\sum_t \exp(u_{it}^{\mathrm{T}} u_w)};$$

$$s = \sum_i \partial_{it} h_{it}. \tag{3}$$

其中,u_w 是网络中的一个参数,可以被视为问句“输入序列中哪部分是最重要的”的语义表示,随机初始化,然后在不断的训练过程中学习得到。

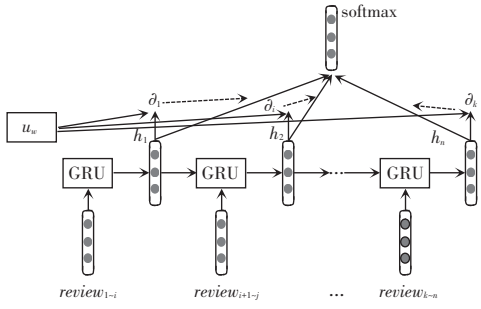

图 2 基于时间序列的评论表示学习

Fig. 2 Comments representation learning based on attention model

这种计算形式,充分利用了每个输入的信息,衡量了每个时间节点的贡献,最终整个评论的语义表示将更倾向于评论中蕴含的大多数的重要信息。

在本次研究中,在学习评论的样本表示时,按照评论的先后时间,结合注意力机制,构建时间序列模型。联合微博文本语义表示,送入 Softmax 分类器进行分类,判断文本是否是谣言。实验结果可参见后文 3.2 节内容。

3 实验与分析

3.1 语料库建设

新浪微博社区管理中心是新浪微博官方成立的,用来协助管理微博的委员会。如若发布淫秽、违法、谣言、辱骂、骚扰等违反社区规定的言论,并经他人举报时,信息就会出现在社区管理中心,等待中心人员的手工审核。

利用新浪微博官方发布的 API,研究时可以获取新浪微博社区管理中心的不实消息版块中的谣言信息。同样地,也可以随机选取一些用户,爬取由其发布的微博,过滤后作为真实信息集。迄至当下,已有 Ma 等人[7]整理公布了一系列的基于微博平台上的数据,且采样方法相同,所以在本次研究中就选择采用了公开的微博数据集。

研究将基于任务相关的语料库按照层次采样的

方式对数据集进行划分，10%为开发集，用作模型调参，剩余的数据就按照 3∶1 的比例，分别用作模型的训练集及测试集，从而得到数据集的分布统计参数详见表 1。

表 1　谣言检测数据集分布

Tab. 1　The distribution of rumor detection dataset

数据集	非谣言	谣言	总数
开发集	235	231	466
测试集	529	520	1 049
训练集	1 587	1 562	3 149

3.2　实验结果与分析

在谣言检测任务中，文中使用了精确率、准确率、召回率以及 F_1 值作为每个类别的评价指标。研究收集了 2 千万的大规模微博数据集，并利用 word2vec[8] 模型，训练得出了针对特定任务的 50 维分布式词向量。

针对微博文本，研究利用 LSTM 模型、GRU 模型学习语义表示，构建分类器，实验结果详见表 2。

表 2　基于 GRU 和 LSTM 的微博文本表示

Tab. 2　Weibo text representation based on GRU and LSTM

方法	类别	精确率/%	准确率%	召回率/%	F_1
LSTM	R	84.74	83.64	86.06	84.83
	N		85.89	83.45	84.65
GRU	R	85.20	83.18	87.98	85.51
	N		87.47	82.51	84.92

实验结果表明，简单的深度学习模型在谣言检测任务上已经可以取得 84%左右的准确率，同时也说明，深度学习模型的优越性与普适性。如果只利用微博文本，GRU 结果略微优于 LSTM 模型结果。

GRU_R 模型对评论内容进行建模，将评论划分成块，各块将作为 GRU 每个时刻的输入，时间序列的长度为块的个数。GRU_Att 模型将注意力机制与 GRU 模型结合起来，并将其应用到微博文本及评论内容的表示学习上，衡量每块评论对微博文本的影响。实验结果详见表 3。

表 3　基于注意力模型的评论表示

Tab. 3　Comment representation based on attention model

方法	类别	精确率/%	准确率%	召回率/%	F_1
GRU_R	R	90.95	89.35	92.79	91.04
	N		92.63	89.13	90.85
GRU_Att	R	92.66	91.71	93.65	92.67
	N		93.63	91.68	92.64

考虑到在谣言检测任务中，GRU 模型在文本语义表示方面略优于 LSTM，且具有速度快，参数少的优点，因此在注意力模型中，研究只在 GRU 模型上进行了尝试。分析了评论对于谣言检测的重要性，通过对评论划分成块，利用时间序列模型学习语义表示，并引入 attention 机制，衡量不同时间节点影响程度，最终可达到 92.66%，相比只利用微博文本的 GRU 模型，提升 7 个百分点。实验结果证明，Attention 机制及评论内容的引入，可以大幅度提升模型的准确率。注意力模型在建模的过程中，着重考虑每个块内评论对谣言检测的影响程度，利用这种不同的影响度，刻画整体评论的语义表示，使得语义表示更趋丰富，更加贴合评论中重要信息，例如怀疑、肯定等态度。

4　结束语

本文主要利用深度学习模型进行谣言的自动识别。实验中，首先尝试了利用 GRU、LSTM 序列模型对微博文本进行建模，并获得了 85.2%的准确率。接着，引入了评论信息，由于评论数过多，对评论按照时间密度划分成块，每块作为时间序列模型每个时刻的输入，同时，引入 attention 机制，重点关注有影响力的评论块。最终，本文提出的模型可以获得 92.66%的准确率，相比只用微博文本，提升了近 8 个百分点。

参考文献

[1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

[2] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997,9(8):1735-1780.

[3] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated Recurrent Neural Networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[4] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.

[5] VINYALS O, KAISER Ł, KOO T, et al. Grammar as a foreign language[J]. arXiv preprint arXiv:1412.7449,2014.

[6] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.

[7] MA Jing, GAO Wei, MITRA P, et al. Detecting rumors from Microblogs with Recurrent Neural Networks [C]// The 25th International Joint Conference on Artificial Intelligence (IJCAI 2016). New York, USA:IJCAI/AAAI Press,2016:3818-3824.

[8] MIKOLOV T, SUTSKEVER I, CHEN Kai, et al. Distributed representations of words and phrases and their compositionality[J]. arXiv preprint arXiv:1310.4546,2013.