基于多粒度语义交互理解网络的幽默等级识别

张瑾晖[1], 张绍武[1], 樊小超[1,2], 杨亮[1], 林鸿飞[1]†
1.大连理工大学/辽宁省大连市
2.新疆师范大学/新疆自治区乌鲁木齐市
wszjh@mail.dlut.edu.cn, zhangsw@dlut.edu.cn
fxc1982@mail.dlut.edu.cn, liang@dlut.edu.cn
hflin@dlut.edu.cn

摘要

幽默在人们日常交流中发挥着重要作用。随着人工智能的快速发展，幽默等级识别成为自然语言处理领域的热点研究问题之一。已有的幽默等级识别研究往往将幽默文本看作一个整体，忽视了幽默文本内部的语义关系。本文将幽默等级识别视为自然语言推理任务，将幽默文本划分为"铺垫"和"笑点"两个部分，分别对其语义和语义关系进行建模，提出了一种多粒度语义交互理解网络，从单词和子句两个粒度捕获幽默文本中语义的关联和交互。本文在Reddit公开幽默数据集上进行了实验，相比之前最优结果，模型在语料上的准确率提升了1.3%。实验表明，引入幽默内部的语义关系信息可以提高模型幽默识别的性能，而本文提出的模型也可以很好地建模这种语义关系。

关键词： 幽默等级识别 ； 自然语言推理 ； 多粒度 ； 语义交互理解

A Multi-Granularity Semantic Interaction Understanding Network for Humor Level Recognition

Jinhui Zhang[1], Shaowu Zhang[1], Xiaochao Fan[1,2], Liang Yang[1], Hongfei Lin[1]†
1.Dalian University of Technology/Dalian, Liaoning
2.Xinjiang Normal University/Urumqi, Xinjiang Autonomous Region
wszjh@mail.dlut.edu.cn, zhangsw@dlut.edu.cn
fxc1982@mail.dlut.edu.cn, liang@dlut.edu.cn
hflin@dlut.edu.cn

Abstract

Humor plays an important role in daily communication, which makes it important problem for natural language processing. Existing works of humor level recognition tend to treat humor text as a whole, and ignore the study of the inner semantic relations of it. This paper regards humor level recognition as a kind of natural language inference task, divides humor text into two parts: "setup" and "punchline", and models the two and their relations respectively. This paper proposes a multi-granularity semantic interaction understanding network to capture semantic association and interaction in humor text from two granularity of word and clause. We conduct experiments on public humor data set Reddit, and the accuracy of the model on this corpus is improved by 1.3% compared with the previous optimal results. Our experiments show that the semantic relationship information inside humor can improve the performance of model on humor level recognition, and the model proposed by us can also represent the semantic relationship well.

Keywords: Humor Level Recognition , Natural Language Inference , Multi-Granularity , Semantic Interaction Understanding

† 通讯作者

第十九届中国计算语言学大会论文集，第645页-第655页，海南，中国，2020年10月30日至11月1日。

1 引言

幽默普遍存在于人们的日常交流中，是化解尴尬、活跃气氛、促进交流的重要手段，可以对人类身心健康产生积极的影响(Morse, 2007)。随着人工智能的快速发展，如何让计算机识别幽默，并进一步识别幽默的等级成为了目前自然语言处理领域的研究热点之一。幽默识别涉及认知语言学、人工智能、心理学等多个学科，其研究能够更好地促进计算机对人类语言的理解。同时，幽默识别能够赋予计算机从更深层次理解人类情感的能力，在机器翻译和人机交互等领域有着广泛的应用。因此，幽默识别及幽默等级识别具有重要的理论研究价值和广泛的应用价值。

传统的幽默识别通常是识别一个句子或段落是否具有幽默的含义(Mihalcea and Strapparava, 2005; Zhang and Liu, 2014; Blinov et al., 2019)。许多研究表明，幽默具有连续性(Blinov et al., 2019; Weller and Seppi, 2019; Hossain et al., 2019)。幽默等级识别，作为幽默识别任务的延伸，旨在根据幽默程度的不同将幽默文本划分为不同的等级。Paulos等(1980)的研究表明，幽默文本通常能够被划分为“铺垫”和“笑点”两个部分，其中“铺垫”一般先于“笑点”表述，是对背景和前提的交代，而“笑点”则是“铺垫”的延续和反转。Weller等(2019)指出，对“铺垫”和“笑点”两部分语义及其关系的深入理解有助于幽默等级识别。表1展示了一个幽默文本及其铺垫和笑点两部分：

幽默文本	As a typical example of failure, you are so successful.
子句1：铺垫部分	As a typical example of failure
子句2：笑点部分	you are so successful

Table 1: 幽默中的铺垫和笑点

在表1中，幽默文本被划分为两个子句，子句1为“铺垫”，子句2为“笑点”。“笑点”既对铺垫中的“failure”做了补充说明，是“铺垫”的延续，又使用“successful”与“铺垫”中的“failure”形成反转。“铺垫”和“笑点”之间对立统一的关系使句子包含了一定程度的幽默。

现有的幽默识别与幽默等级识别研究通常分两步进行：首先基于幽默理论，设计并实现一系列的幽默特征；然后采用传统的机器学习方法或结合神经网络方法对幽默文本或幽默等级进行识别。Weller等(2019)采用基于Transformer的预训练模型对幽默等级进行识别并取得了较好的性能。人工构造特征耗时耗力且难以对多样性的幽默表达进行全面表征，模型的泛化能力较弱。现有的神经网络模型和预训练模型将铺垫和笑点作为整体进行建模，忽略了其独立的语义信息和交互的关联信息。此外，由于语言的细微差别可能造成幽默的程度不同，仅从单一的粒度提取幽默特征，模型的性能可能受到限制。

综上所述，为了缓解幽默等级识别中的问题，本文提出了一种基于多粒度语义交互理解网络的幽默等级识别方法。针对幽默语言多样性的问题，采用了多种词嵌入表示融合的方法对幽默文本进行表征；针对幽默语义复杂性的问题，采用了局部语义交互理解模块和全局语义交互理解模块，分别从单词粒度和子句粒度提取幽默文本的高维潜在语义特征；针对幽默中“铺垫”和“笑点”的语义关联特点，采用“交互型”的神经网络模型对二者的关联信息进行建模；最后对多粒度的语义特征和交互关联特征进行融合并对幽默等级进行识别。本文的贡献如下：

1. 本文基于多种嵌入表示融合的幽默文本表示，提出了一种基于局部和全局语义理解的神经网络模型，分别从单词级别和子句级别提取幽默文本特征。

2. 本文提出了一种基于交互语义关联特征的神经网络模型，对幽默文本中“铺垫”和“笑点”的关联信息进行建模以抽取幽默语义关联特征。

3. 本文使用基于多粒度语义交互理解网络的幽默等级识别方法，在Reddit公开幽默数据集上进行对比实验，结果表明，本文提出的方法能够有效地提升幽默等级识别的性能。

2 相关工作

作为日常生活中常见的语言现象，幽默理论研究历史久远，基于幽默理论，幽默识别也有很多的研究成果，而幽默等级识别研究则刚刚起步。本节将从幽默理论，幽默识别和幽默等级识别三个方面总结前人的工作。

2.1 幽默理论

幽默理论对幽默等级识别研究具有重要的指导意义。在众多幽默理论中，乖讹论被广泛接受且具有深远的影响。乖讹论认为幽默是人类对不协调事物的感知，当事物的发展违背人们的常识和期望时，幽默就产生了(SULS, 1972)。基于乖讹论，Raskin等(1979)提出了第一个语言学意义上的幽默理论——语义脚本理论（Script Semantic Theory of Humor, SSTH），该理论认为语义对立是幽默产生的重要原因。基于以上幽默理论，Paulos等(1980)将幽默分为“铺垫”和“笑点”，认为两部分之间存在对立统一的关系。

2.2 幽默识别

传统的机器学习方法被广泛应用于幽默识别领域。Yang等(2015)从不一致性、歧义性、语音特性和人机交互特性四个方面提取幽默的语义特征，并采用了随机森林方法识别幽默。Barbieri等(2014)根据幽默问题的语音和歧义性特点，构造了多种幽默特征。Zhang等(2014)基于幽默的语言学理论，构建50多种幽默特征并将它们划分为五个类别。Liu等(2018b)提取了对话中的情感特征及情感关联特征识别对话中的幽默。此外，他们对幽默文本中句法结构特征进行了深入的分析并指出句法结构和幽默文本具有高度的相关性(Liu et al., 2018a)。

近年来，越来越多的深度学习方法被用于识别幽默。杨勇等(2020)从音形义三个维度对幽默特征进行建模，采用层次注意力机制对幽默进行识别。Bertero等(2016a; 2016b)由《生活大爆炸》中的文本和语音内容构建幽默数据集，采用长短期记忆网络和卷积神经网络自动抽取文本语义特征，从而预测对话中的幽默。Baziotis等(2017)利用注意力机制，更好的关注到句子中的特定单词，从而提高了幽默识别的性能。Zhao等(2019)提出了一种采用张量分解的方法提取幽默的语义特征。除了英文，研究者采用深度学习方法对西班牙文(Bahdanau et al., 2014)和俄文(Blinov et al., 2019)语料进行了幽默识别。

2.3 幽默等级识别

幽默等级识别使计算机能够理解哪些语义和语义关系使句子更加有趣。Chris等(2018)对单词的幽默程度建模并对4997个单词的幽默程度进行了评分。Hossain等(2019)通过重新编辑新闻标题使其变得更加幽默，并对编辑前后文本语义的幽默程度进行了分析。Cattle等(2016)将文本划分为“铺垫”和“笑点”两个部分，并指出二者的语义相关性对文本的幽默等级具有显著影响。此外，一些国际著名评测也将幽默等级识别任务作为评测主题(Potash et al., 2017)。

综上所述，幽默理论为幽默等级识别的研究提供了理论研究基础。此外，从不同粒度提取幽默文本中“铺垫”和“笑点”的语义特征和语义关系特征有助于幽默等级识别性能的提升。

3 幽默等级识别方法

基于多粒度语义交互理解网络的幽默等级识别方法主要包括两个层次，语义的嵌入式表示层和交互特征提取层。交互语义特征提取层包括两个部分，局部语义交互理解模块和全局语义交互理解模块。

基于多粒度语义交互理解的神经网络模型如图1所示。语义的嵌入式表示层能够获取幽默文本中“铺垫”和“笑点”的高维潜在语义表示。首先为了更好地获取不同词嵌入表示的语义信息，融合多种词嵌入表示对“铺垫”和“笑点”中的单词进行表征；其次，为了获取高维潜在语义表示，采用双向长短期记忆网络（Bi-directional LSTM, Bi-LSTM）分别提取“铺垫”和“笑点”的语义信息并得到上下文表示。交互语义特征提取层将上下文表示作为输入，从局部和全局两个维度交互地提取“铺垫”和“笑点”中语义特征及两者之间的语义关联性特征。局部语义交互理解模块计算得到“铺垫”中单词语义表示和“笑点”中单词语义表示的关联信息，全局语义交互理解模块计算“铺垫”子句和“笑点”子句的关联信息。最后对局部和全局信息进行融合并对幽默等级进行识别。

第十九届中国计算语言学大会论文集，第645页-第655页，海南，中国，2020年10月30日至11月1日。

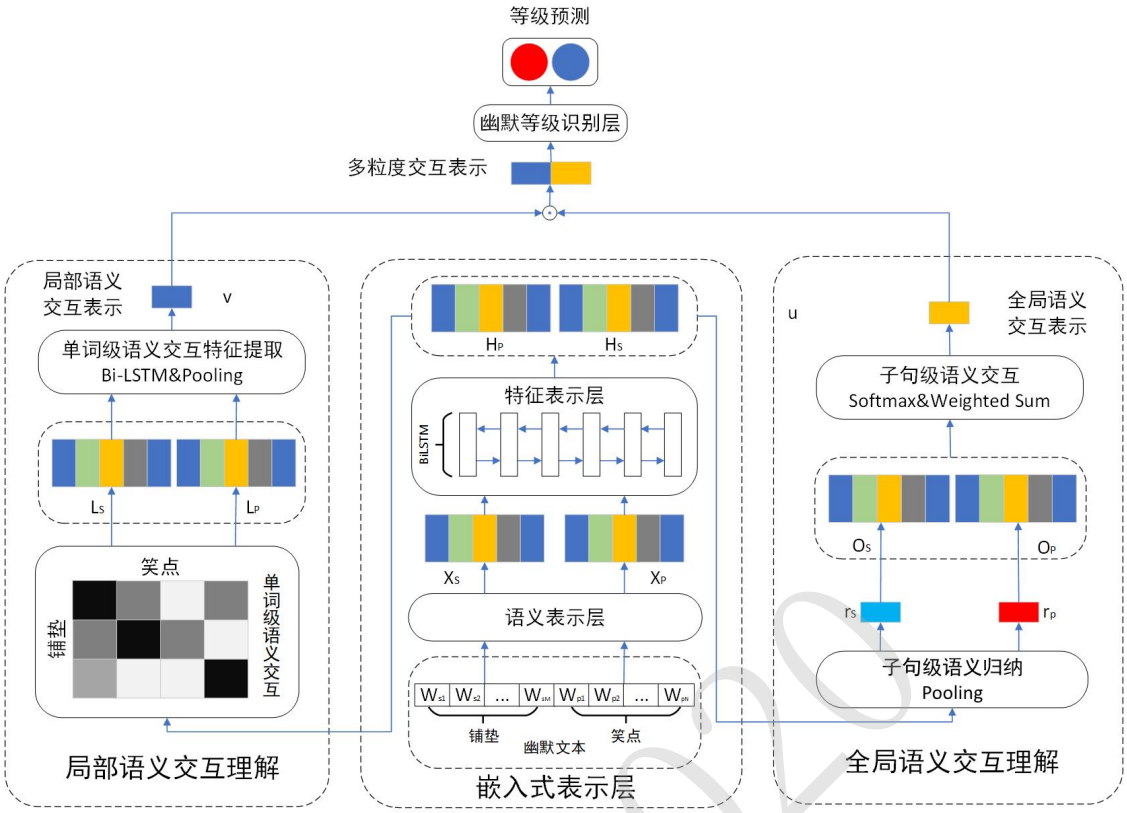

Figure 1: MSIN模型框架图

3.1 语义的嵌入式表示层

幽默是一种复杂的语言现象，一词多义等特征使得幽默特征表示和提取变得更加困难(Yang et al., 2015)。Xu等(2018)指出领域内和领域外的词嵌入表示的融合有助于文本分类模型性能的提升。目前还没有由幽默语料训练得到的词嵌入表示，而大规模的的词嵌入表示，如GloVe(Pennington et al., 2014)、BERT(Devlin et al., 2018)等，一般是利用通用语料或者新闻语料训练得到的。直接采用单一的词嵌入表示往往使得幽默等级识别的性能欠佳。此外，“铺垫”和“笑点”在幽默等级识别中起着不同的作用，将二者统一建模，不利于文本的幽默等级识别。因此，本文将幽默文本的“铺垫”和“笑点”分别建模，采用多个领域词嵌入表示进行融合，并采用Bi-LSTM提取两个部分的高维语义特征。

3.1.1 语义表示层

该层将“铺垫”和“笑点”中的每个单词映射到多个高维特征空间，并对其进行融合以获取有意义的语义表示。设幽默语句为$W = \{w_{s1}, w_{s2}, \ldots, w_{sM}, w_{p1}, \ldots, w_{pN}\}$，其“铺垫”为$W_S = \{w_{s1}, w_{s2}, \ldots, w_{sM}\}$，“笑点”为$W_P = \{w_{p1}, w_{p2}, \ldots, w_{pN}\}$，其中$w_i$为语句中的任一单词，$M + N$为句子总长度，$M$和$N$分别为“铺垫”和“笑点”的长度。将幽默语句中每个单词表示为K种低维稠密向量，并对同一单词多种向量进行拼接，得到单词的向量表示。则“铺垫”的向量表示为$X_S = \{x_{s1}, x_{s2}, \ldots, x_{sM}\} \in R^{D \times M}$，“笑点”的向量表示为$X_P = \{x_{p1}, x_{p2}, \ldots, x_{pN}\} \in R^{D \times N}$，$D$是词向量的维度，$D = D_1 + D_2 + \ldots + D_K$。

3.1.2 特征表示层

在该层中，模型利用Bi-LSTM分别提取“铺垫”和“笑点”子句的语义特征，作为幽默文本的“铺垫”和“笑点”的特征表示。LSTM(Hochreiter and Schmidhuber, 1997)能够对文本语义上的长距离依赖关系进行建模，而Bi-LSTM能够从正反两个方向提取潜在语义特征，并融合两部分的语义信息。在每个时间步t，正向和反向LSTM对输入词向量x_t的处理过程可以分别形式化的表示为：

$$\overrightarrow{h}_t = LSTM\left(h_{t-1}, x_t\right) \tag{1}$$

$$\overleftarrow{h}_t = LSTM\left(h_{t+1}, x_t\right) \tag{2}$$

其中，h_t表示t时刻的隐态向量，x_t为t时刻输入的词向量。将每个时间步正反两个方向的隐态向量拼接就得到Bi-LSTM单个时间步的输出，记作$h_t = \left[\overrightarrow{h}_t, \overleftarrow{h}_t\right] \in R^{2h}$，$h$表示隐态向量的维度。

特征表示层能够得到幽默文本的潜在语义特征，记为$H = [H_S, H_P] = [h_{s1}, h_{s2}, \ldots, h_{sM}, h_{p1}, \ldots, h_{pN}] \in R^{(M+N)\times 2h}$，其中$H_S$和$H_P$分别是铺垫和笑点的潜在语义表示。

3.2 交互语义特征提取层

Chen等(2016)指出不同粒度的语义单元及其交互信息能够有效地提高模型对文本语义的理解。铺垫和笑点作为两个语义单元，两者在不同粒度上相互作用，铺垫中单个词语及铺垫整体都会影响到笑点的语义表达，反之亦然。此外，Engelthaler等(2017)指出不同单词在句子中表现出不同的幽默程度。单词的语义信息与语句的幽默等级具有一定的相关性。

为使神经网络模型能够学习到单词和句子的语义信息，并且能够获取“铺垫”和“笑点”之间的关联信息，本文采用局部语义交互理解模块和全局语义交互理解模块对来自上层的潜在语义表示做处理。

3.2.1 局部语义交互理解模块

Yang等(2015)研究发现，在幽默文本中，不同词语的重要程度不同，当删除幽默文本中的某些词语后，文本的幽默程度下降甚至完全消失。本文采用局部语义交互理解模块从单词级别提取幽默文本的语义信息和语义关联信息。局部语义交互理解模块包括单词级语义交互层和单词级语义特征提取层。

单词级语义交互层 单词级语义交互层使用软对齐的方式获取“铺垫”和“笑点”的单词粒度语义交互表示。具体地讲，对来自特征表示层的铺垫和笑点的潜在语义表示H_S和H_P，该层首先将两者中每个单词对应向量两两之间做点乘，计算公式为：

$$e_{ij} = h_{si} * h_{pj} \tag{3}$$

可以得到铺垫和笑点的相似度矩阵$E = \{e_{ij} | i \in [1, M], j \in [1, N]\} \in R^{M\times N}$，其中$e_{ij}$表示铺垫中第$i$个单词和笑点中第$j$个单词的相似度。

然后，该层以加权求和的形式求出铺垫和笑点中每个单词对应的交互表示：

$$\widetilde{h}_{si} = \sum_{j=1}^{N} \frac{exp\left(e_{ij}\right)}{\sum_{k=1}^{N} exp\left(e_{ik}\right)} h_{pj}, \forall i \in [1, M] \tag{4}$$

$$\widetilde{h}_{pj} = \sum_{i=1}^{M} \frac{exp\left(e_{ij}\right)}{\sum_{k=1}^{M} exp\left(e_{kj}\right)} h_{si}, \forall j \in [1, N] \tag{5}$$

由上面两个式子可知，交互表示包括由铺垫表示的笑点和由笑点表示的铺垫，模型使用铺垫或者笑点中所有向量的加权和来得到对方每个单词的表示，以这种方式实现铺垫和笑点的交互。

最后，模型融合每部分文本各自的潜在语义表示及交互表示，得到两部分在该层的输出：

$$L_S = [l_{s1}, l_{s2}, \ldots, l_{sM}] \tag{6}$$

$$L_P = [l_{p1}, l_{p2}, \ldots, l_{pN}] \tag{7}$$

其中$L_S \in R^{M*8h}$、$L_P \in R^{N*8h}$分别是铺垫和笑点的单词级语义交互表示。本文使用如下方法对每个单词的潜在语义表示和交互表示进行融合：

$$l_{si} = [h_{si}, \widetilde{h}_{si}, h_{si} - \widetilde{h}_{si}, h_{si} * \widetilde{h}_{si}], \forall i \in [1, \ldots, M] \tag{8}$$

$$l_{pj} = [h_{pj}, \widetilde{h}_{pj}, h_{pj} - \widetilde{h}_{pj}, h_{pj} * \widetilde{h}_{pj}], \forall j \in [1, \ldots, N] \tag{9}$$

其中l_{si}、$l_{pj} \in R^{8h}$，

单词级语义交互特征提取层 该层对单词级交互信息进一步抽象，获取单词级语义交互特征。首先，该部分分别将L_s和L_p经过Bi-LSTM来提取单词级交互表示的高层特征，计算过程为：

$$G_S = [g_{s1}, g_{s2}, \ldots, g_{sM}] = Bi - LSTM(L_S) \tag{10}$$

$$G_P = [g_{p1}, g_{p2}, \ldots, g_{pM}] = Bi - LSTM(L_P) \tag{11}$$

其中$G_S \in R^{M \times 2h}$，$G_P \in R^{N \times 2h}$。分别对G_S和G_P做平均池化和最大池化，并将池化获得的四个向量拼接，最终得到幽默文本的局部语义交互特征向量v，计算过程如下：

$$v_{s,ave} = ave\_pool(G_S), v_{s,max} = max\_pool(G_S) \tag{12}$$

$$v_{p,ave} = ave\_pool(G_P), v_{p,max} = max\_pool(G_P) \tag{13}$$

$$v = [v_{s,ave}, v_{s,max}, v_{p,ave}, v_{p,max}] \tag{14}$$

3.2.2 全局语义交互理解模块

Ma等(2017)研究表明，对于文本中的不同语义单元，其单词的含义会受到其他语义单元的有影响。铺垫和笑点作为幽默文本的两个子句级语义单元，二者互相作用，对幽默等级识别产生重要影响。本文采用全局语义交互理解模块从子句级别提取幽默文本的语义信息和语义关联信息。全局语义交互理解模块包括子句级语义归纳层和子句级语义特征提取层。

子句级语义归纳层 该层分别对“铺垫”和“笑点”子句的上下文表示（H_S或者H_P）做平均池化和最大池化，将两部分拼接得到二者的归纳表示。公式如下：

$$r_{s,ave} = ave\_pool(H_S), r_{s,max} = max\_pool(H_S), r_s = [r_{s,ave}, r_{s,max}] \tag{15}$$

$$r_{p,ave} = ave\_pool(H_P), r_{p,max} = max\_pool(H_P), r_p = [r_{p,ave}, r_{p,max}] \tag{16}$$

得到的向量r_s、r_p通过全连接层把它们的维度投影到$2h$。

子句级语义交互层 该层对铺垫和笑点子句做交互，然后对交互信息进一步抽象，以获取全局的语义交互特征。首先，计算子句与各词之间的交互权重：

$$O_S = [o_{s1}, o_{s2}, ..., o_{sM}] = [h_{s1} * r_p, h_{s2} * r_p, \ldots, h_{sM} * r_p] \tag{17}$$

$$O_P = [o_{p1}, o_{p2}, ..., o_{pN}] = [h_{p1} * r_s, h_{p2} * r_s, \ldots, h_{pN} * r_s] \tag{18}$$

其中$O_S \in R^{M \times 2h}$，$O_P \in R^{N \times 2h}$。

然后，通过加权求和的方式获得两个子句的交互特征，并最终得到全局语义交互特征向量u：

$$u_s = \sum_{j=1}^{N} \frac{exp(o_{pj})}{\sum_{k=1}^{N} exp(o_{pk})} h_{pj} \tag{19}$$

$$u_p = \sum_{i=1}^{M} \frac{exp(o_{si})}{\sum_{k=1}^{M} exp(o_{sk})} h_{si} \tag{20}$$

$$u = [u_s, u_p] \tag{21}$$

其中u_s、$u_p \in R^{2h}$分别是铺垫和笑点的子句级别交互特征，将两者拼接得到u。

3.3 幽默等级识别层

该层由全连接层及softmax层组成。首先将局部和全局语义信息进行融合，然后通过全连接层和softmax层，得到幽默等级的概率分布，计算公式如下：

$$T = [v_{s,ave}, v_{s,max}, u_s, v_{p,ave}, v_{p,max}, u_p] \tag{22}$$

$$humor_{cls} = softmax\left(T\right) = \frac{e^{ti}}{\sum_{i=1}^{12h} e^{ti}} \tag{23}$$

其中$T \in R^{10h}$，$humor_{cls} \in R^C$是概率分布，C是幽默等级数量。本文采用交叉熵作为损失函数，其形式化表示如下：

$$loss = -\sum_{i=1}^{Num}\sum_{j=1}^{C} y_i^j \, log\widehat{y}_i^j + \lambda \parallel \theta \parallel^2 \tag{24}$$

其中，Num是训练集样本数，i是样本序号，j是标签序号，y_i^j是样本的真实标签类别，$\widehat{y}_i^j$是样本的预测标签类别，λ是L_2正则化项的超参数，θ是模型参数的集合。

4 实验结果

本节首先介绍了实验数据、评价指标、实验设置和基线方法，然后对比了基线方法和本文提出的MSIN方法的幽默等级识别性能，最后通过实验分析了本文提出方法的有效性。

4.1 实验数据与评价指标

Reddit数据集：该数据集由Weller等(2019)构建。幽默语句来自Reddit中带有“humor”标签的文本，采用众包方式对幽默语句的“铺垫”和“笑点”进行了标注，且对幽默语句的强弱进行了人工标注。数据集规模详见下表。

	弱幽默	强幽默	总计
训练集	9719	9719	19438
验证集	304	304	608
测试集	304	304	608

Table 2: Reddit幽默数据集统计信息

评价指标：为了便于和基线方法进行比较，本文采用了被广泛接受并应用于文本分类任务中的精确率（Acc）、准确率（P）、查全率（R）和F1 Score （F1）作为评价指标。

4.2 实验设置

词嵌入：在训练过程中，词嵌入表示分别采用了Glove以及Word2Vce(Mikolov et al., 2013)，维度均为300，词嵌入在训练的过程中固定。对未登录词使用$(-0.01, 0.01)$上的平均分布随机初始化。

超参数：在实验中，设置L_2正则化项的超参数$\lambda = 10^{-5}$，Bi-LSTM的神经元个数为128，CNN三个卷积核的尺寸分别为2、3和5，优化方法为Adam(Kingma and Ba, 2014)，Batch大小为64，dropout为0.5。为了防止过度拟合，在训练过程中使用了学习率衰减和早停机制。为了便于和基线模型对比，采用了Weller等(2019)对数据的划分。

4.3 基线方法

本文使用下述基线方法进行对比实验：

- Human(Weller and Seppi, 2019)*：人工预测结果。
- CNN(Weller and Seppi, 2019)*：采用CNN自动提取幽默语句的潜在语义特征并进行幽默等级识别。

https://nlp.stanford.edu/projects/glove/
https://code.google.com/archive/p/word2vec/

Method			Precision	Recall	F1_Score	Accuracy
Human(Weller et al.)*			-	-	-	66.30
分类任务	分类模型	CNN(Weller et al.)*	-	-	-	68.80
		CNN(Kim et al.)	68.22	68.85	68.18	68.16
		LSTM(Hochreiter et al.)	69.46	68.09	68.77	69.08
		Bi-LSTM-Attention	68.70	73.62	71.02	69.98
		Transformer(Weller et al.)*	-	-	-	72.40
		BERT(Devlin et al.)	72.06	74.67	73.34	72.86
推理任务	表示模型	CNN(Kim et al.)	69.78	69.87	69.41	69.42
		LSTM(Hochreiter et al.)	70.66	71.61	70.89	70.71
		BiLSTM-Attention	69.93	73.88	71.70	70.97
		BERT(Devlin et al.)	73.27	73.03	73.15	73.19
	交互模型	ESIM(Chen et al.)	73.38	70.72	72.03	72.53
		MSIN	**74.10**	**74.34**	**74.22**	**74.18**

Table 3: Reddit数据集实验结果

- CNN(Kim, 2014)：本文复现的基于CNN的方法，使用3种不同尺寸卷积核的CNN提取幽默文本特征进行幽默等级识别。
- LSTM(Hochreiter and Schmidhuber, 1997)：使用LSTM提取幽默特征并进行幽默等级识别。
- Bi-LSTM-Attention：使用双向LSTM和注意力机制提取幽默文本特征，并对幽默等级进行识别。
- Transformer(Weller and Seppi, 2019)*：使用基于transformer结构(Vaswani et al., 2017)的预训练模型对幽默文本整体做特征提取，以进行幽默等级识别。
- BERT(Devlin et al., 2018)：本文复现的基于BERT方法的结果，在任务语料上做微调然后进行幽默等级识别。
- ESIM(Chen et al., 2016)：只基于局部语义交互信息进行幽默等级识别。
- MSIN：本文提出的多粒度语义交互理解网络，综合使用语义嵌入、局部语义交互和全局语义交互进行幽默等级识别。

4.4 实验结果分析

本文在Reddit数据集上的实验结果见表3。表格整体分为三部分，第一部分为人工进行幽默等级识别的结果；第二部分采用之前研究的通用方法，将幽默等级识别视作文本分类任务，把幽默文本整体编码后进行分类；第三部分基于本文观点，即可将幽默等级识别任务视作自然语言推理任务，把幽默文本划分为铺垫和笑点两个语义部分，以这两部分作为模型的输入，使用表示型模型或交互型模型预识别文本蕴含的幽默等级。

在第二部分，本文的使用的CNN与Weller等(2019)的CNN结果相近，且两者均取得了明显好于人工预测的结果，证明了神经网络在幽默等级识别上的有效性。然而CNN由于卷积核尺寸固定，难以捕获长距离的语义关系，这对需要充分理解上下文的幽默等级识别任务是不利的。相比CNN，LSTM使用隐态向量捕获句子在长距离上的语义关系，可对时间序列进行有效建模，在数据集上取得了好于CNN的结果。然而LSTM是有偏倚的模型，后送入模型的信息会比先送入模型的信息拥有更大的权重，因此文本又使用Bi-LSTM+Attention进行改进。一方面，BiLSTM可以编码句子从前到后和从后到前两个方向上的信息，获取的特征更丰富，另一方面，Attention将所有时间步上的隐态向量赋予权重，让模型关注在文本分类过程中起关键作用的部分，缓解了由于LSTM的偏倚性造成的信息损失，因此模型相比LSTM取得了更好的结果。最后，本文使用BERT识别文本的幽默等级，其结果与Weller等(2019)使用Transformer的结果相近，并且两者均明显优于之前的模型。

Method	Precision	Recall	F1_Score	Accuracy
MSIN+Glove	73.20	70.07	71.60	72.20
MSIN+Word2Vec	73.33	72.37	72.85	73.03
MSIN+Both	**74.10**	**74.34**	**74.22**	**74.18**

Table 4: 不同词向量使用方式结果比较

Method	Precision	Recall	F1_Score	Accuracy
Word Level	72.64	73.36	73.00	72.86
Sub-sentence Level	70.68	73.22	71.91	71.41
MSIN	**74.10**	**74.34**	**74.22**	**74.18**

Table 5: 不同粒度实验结果比较

在第三部分，本文分别使用表示型和交互型两类模型进行幽默等级识别。

表示模型分别将铺垫和笑点编码为向量，然后将两向量与他们之间作差及点乘的结果拼接以捕获两部分的关系，最后基于拼接后的向量进行分类。为方便与第一部分的结果作比较，本文仍采用CNN、LSTM、Bi-LSTM-Attention和BERT四个模型。首先做内部比较，可以发现四个模型的结果依次递增，与第一部分的趋势保持一致；其次将表示模型与第一部分比较，发现四个模型的结果均高于第一部分中对应的模型，证明将幽默文本拆分为铺垫和笑点两部分，并让模型学习两部分之间的关系信息有助于幽默等级的识别。

在交互模型部分，本文使用ESIM与本文提出的MSIN进行比较。ESIM通过计算两部分文本之间单词的相似度矩阵来构建局部语义交互表示，并以此来推断前后文本的关系，在没有大量预训练知识的情况下，取得了略低于BERT的结果。本文提出的MSIN综合考虑交互过程中局部和全局语义信息的影响，取得了好于ESIM的最优结果。因此可以证明，相比表示模型，交互模型可以更好地捕捉到铺垫和笑点之间的关系；本文提出的多粒度语义交互理解模型融合单词和子句两个级别的交互信息，在幽默等级识别任务上取得了提升。

同时，本文进行消融实验，证明了词向量融合及多粒度交互两个结构的有效性，实验结果分别见表4和表5。表4前两行分别为只使用Glove和只使用Word2Vec的结果，第三行是使用融合词向量的结果，可以发现，融合之后效果更佳。表5前两行分别为只使用单词和子句交互的结果，第三行为融合两个粒度进行交互的结果，可以发现，多粒度交互网络取得了最优结果。

5 结论

本文将幽默文本划分为铺垫和笑点两部分，提出对两者之间的关系进行建模可以显著提升模型识别幽默等级的性能。基于这个观点，首先，本文在融合多种嵌入表示的基础上，从局部和全局两个粒度来对幽默中的语义关系进行理解和建模。其次，本文对幽默中“铺垫”和“笑点”两部分的关联信息做交互建模，从而实现充分挖掘铺垫和笑点之间的关系。最后，本文在Reddit幽默数据集上进行实验，取得了最优结果，同时结合消融实验证实了模型设计的有效性。在以后的工作中，我们将在幽默文本自动切分及基于铺垫的笑点文本生成方面做更多的探索。

参考文献

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv: Computation and Language*.

Francesco Barbieri and Horacio Saggion. 2014. Automatic detection of irony and humour in twitter. *Process Biochemistry*, 40(8):2637–2642.

Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis. 2017. Datastories at semeval-2017 task 6: Siamese lstm with attention for humorous text comparison. pages 390–395.

Dario Bertero and Pascale Fung. 2016a. Deep learning of audio and language features for humor prediction. page 496.

Dario Bertero and Pascale Fung. 2016b. A long short-term memory framework for predicting humor in dialogues. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Vladislav Blinov, Valeria Bolotovabaranova, and Pavel Braslavski. 2019. Large dataset and language model fun-tuning for humor recognition. pages 4027–4032.

Andrew Cattle and Xiaojuan Ma. 2016. Effects of semantic relatedness between setups and punchlines in twitter hashtag games. pages 70–79.

Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2016. Enhanced lstm for natural language inference.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Tomas Engelthaler and Thomas T. Hills. 2017. Humor norms for 4,997 english words. *Behavior Research Methods*, 50(1):1–9.

Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Nabil Hossain, John Krumm, and Michael Gamon. 2019. "president vows to cut hair": Dataset and analysis of creative text editing for humorous headlines. *arXiv: Computation and Language*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. pages 1746–1751.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *Computer Science*.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018a. Exploiting syntactic structures for humor recognition. pages 1875–1883.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018b. Modeling sentiment association in discourse for humor recognition. 2:586–591.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Twenty-Sixth International Joint Conference on Artificial Intelligence*.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. pages 531–538.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Computer Science*.

Donald R. Morse. 2007. Use of humor to reduce stress and pain and enhance healing in the dental setting. *J N J Dent Assoc*, 78(4):32–36.

John Allen Paulos. 1980. *Mathematics and Humor*. University of Chicago Press.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Peter Potash, Alexey Romanov, and Anna Rumshisky. 2017. Semeval-2017 task 6: hashtagwars: Learning a sense of humor. In *International Workshop on Semantic Evaluation*.

Victor Raskin. 1979. Semantic mechanisms of humor. *Synthese Language Library*, 5(4):409–415.

J. M. SULS. 1972. A two-stage model for the appreciation of jokes and cartoons : An information-processing analysis. *Psychology of Humor*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. pages 5998–6008.

Orion Weller and Kevin D Seppi. 2019. Humor detection: A transformer gets the last laugh. pages 3619–3623.

Chris Westbury and Geoff Hollis. 2018. Wriggly, squiffy, lummox, and boobs: What makes some words funny? *Journal of Experimental Psychology General*, 148(1).

Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. pages 2367–2376.

Renxian Zhang and Naishi Liu. 2014. Recognizing humor on twitter. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 889–898.

Zhenjie Zhao, Andrew Cattle, Evangelos E Papalexakis, and Xiaojuan Ma. 2019. Embedding lexical features via tensor decomposition for small sample humor recognition. pages 6375–6380.

杨勇，杨亮，邹艳波，任鸽，樊小超. 2020. 基于音形义特征和层次注意力机制的幽默识别. 计算机工程, pages 1–12.