**NAACL会议-论文调研汇总**

**2016年accepted papers**

**有关文本分类（Classification）：**

1. **Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks**

摘要：Recent approaches based on artificial neural networks (ANNs) have shown promising results for short-text classification. However, many short texts occur in sequences (e.g., sentences in a document or utterances in a dialog), and most existing ANN-based systems do not leverage the preceding short texts when classifying a subsequent one. In this work, we present a model based on recurrent neural networks and convolutional neural networks that incorporates the preceding short texts. Our model achieves state-of-the-art results on three different datasets for dialog act prediction.

（最近基于人工神经网络（ANN）的研究已经显示出了用于短文本分类的有希望的结果。 然而，许多短文本以序列（例如，文档中的句子或对话中的话语）出现，并且当对后续的文本进行分类时，大多数现有的基于ANN的系统不利用前面的短文本。在这项工作中，我们提出一个基于循环神经网络和卷积神经网络，结合上述短文本的模型，我们的模型实现了三个不同的数据集的对话行为预测的最先进的结果。）

**总结：基于RNN与CNN，解决短文本分类的问题。**

1. **Improved Neural Network-based Multi-label Classification with Better Initialization Leveraging Label Co-occurrence**

摘要：In a multi-label text classification task, in which multiple labels can be assigned to one text, label co-occurrence itself is informative.We propose a novel neural network initialization method to treat some of the neurons in the final hidden layer as dedicated neurons for each pattern of label co-occurrence. These dedicated neurons are initialized to connect to the corresponding co-occurring labels with stronger weights than to others. In experiments with a natural language query classification task, which requires multi-label classification, our initialization method improved classification accuracy without any computational overhead in training and evaluation.

（在多标签文本分类任务中，其中多个标签可以被分配给一个文本，标签共现本身是信息性的。我们提出了一种新颖的神经网络初始化方法，以将最终隐藏层中的一些神经元作为专用神经元对于标签共生的每种模式。这些专用神经元被初始化以连接到具有比其他更强的权重的对应的共同出现的标签。在需要多标签分类的自然语言查询分类任务的实验中，我们的初始化方法提高了分类精度，在训练和评估中没有任何计算开销。）

**总结：神经网络解决多标签文本分类问题。**

1. **Combining Recurrent and Convolutional Neural Networks for Relation Classification**

摘要：This paper investigates two different neural architectures for the task of relation classification: convolutional neural networks and recurrent neural networks. For both models, we demonstrate the effect of different architectural choices. We present a new context representation for convolutional neural networks for relation classification (extended middle context). Furthermore, we propose connectionist bi-directional recurrent neural networks and introduce ranking loss for their optimization. Finally, we show that combining convolutional and recurrent neural networks using a simple voting scheme is accurate enough to improve results. Our neural models achieve state-of-the-art results on the SemEval 2010 relation classification task.

（本文研究两种不同的神经体系结构的关系分类任务：卷积神经网络和循环神经网络。对于这两个模型，我们演示不同的建筑选择的影响。我们为卷积神经网络的关系分类（扩展的中间上下文）提出一个新的上下文表示。此外，我们提出连接主义双向循环神经网络和引入排名损失为他们的优化。最后，我们表明，使用简单的投票方案组合卷积和复现神经网络是准确的，以改善结果。我们的神经模型在SemEval 2010关系分类任务中实现最先进的结果。）

Relation classification is the task of assigning sentences with two marked entities to a predefined set of relations.（关系分类是将具有两个标记的实体的句子分配到预定义的关系集合的任务。）

**总结：CNN、RNN解决文本关系分类问题。**

1. **Hierarchical Attention Networks for Document Classification**

摘要：We propose a hierarchical attention network for document classification. Our model has two distinctive characteristics: (i) it has a hierarchical structure that mirrors the hierarchical structure of documents; (ii) it has two levels of attention mechanisms applied at the wordand sentence-level, enabling it to attend differentially to more and less important content when constructing the document representation. Experiments conducted on six large scale text classification tasks demonstrate that the proposed architecture outperform previous methods by a substantial margin. Visualization of the attention layers illustrates that the model selects qualitatively informative words and sentences.

（我们提出一个文档分类的层次关注网络。模型有两个独特的特点：（i）它有一个层次结构，反映文件的层次结构; （ii）它在字和句子级别上应用了两个级别的注意力机制，使得它在构造文档表示时能够不同地参与越来越重要的内容。在六个大规模文本分类任务上进行的实验表明，所提出的架构比以前的方法性能大得多。关注层的可视化示出了模型选择定性信息词和句子。）

**总结：神经网络解决文档分类问题。**

1. **Dependency Based Embeddings for Sentence Classification Tasks**

摘要：We compare different word embeddings from a standard window based skipgram model, a skipgram model trained using dependency context features and a novel skipgram variant that utilizes additional information from dependency graphs. We explore the effectiveness of the different types of word embeddings for word similarity and sentence classification tasks. We consider three common sentence classification tasks: question type classification on the TREC dataset, binary sentiment classification on Stanford’s Sentiment Treebank and semantic relation classification on the SemEval 2010 dataset. For each task we use three different classification methods: a Support Vector Machine, a Convolutional Neural Network and a Long Short Term Memory Network. Our experiments show that dependency based embeddings outperform standard window based embeddings in most of the settings, while using dependency context embeddings as additional features improves performance in all tasks regardless of the classification method.

（我们比较了来自基于标准窗口的skipgram模型，使用依赖上下文特征训练的skipgram模型和利用来自依赖图的附加信息的新颖skipgram变体的不同字嵌入。我们探讨了不同类型的词嵌入对词相似性和句子分类任务的有效性。我们考虑三个常见的句子分类任务：TREC数据集上的问题类型分类，斯坦福的情感树库的二元情感分类和SemEval 2010数据集上的语义关系分类。对于每个任务，我们使用三种不同的分类方法：支持向量机，卷积神经网络和长期短存储器网络。我们的实验表明，在大多数设置中，依赖性嵌入的性能优于标准的基于窗口的嵌入式，而使用依赖性上下文嵌入作为附加特性可提高所有任务的性能，而不管分类方法如何。）

**总结：SVM、CNN、LSTM解决句子问题类型分类。**

1. **Deep LSTM based Feature Mapping for Query Classification**

摘要：Traditional convolutional neural network(CNN) based query classification uses linear feature mapping in its convolution operation. The recurrent neural network (RNN), differs from a CNN in representing word sequence with their ordering information kept explicitly. We propose using a deep long-short-term-memory (DLSTM) based feature mapping to learn feature representation for CNN. The DLSTM, which is a stack of LSTM units, has different order of feature representations at different depth of LSTM unit. The bottom LSTM unit equipped with input and output gates, extracts the first order feature representation from current word. To extract higher order nonlinear feature representation, the LSTM unit at higher position gets input from two parts. First part is the lower LSTM unit’s memory cell from previous word. Second part is the lower LSTM unit’s hidden output from current word. In this way, the DLSTM captures the nonlinear nonconsecutive interaction within n-grams. Using an architecture that combines a stack of the DLSTM layers with a tradition CNN layer, we have observed new state-of-the-art query classification accuracy on benchmark data sets for query classification.

（传统的基于卷积神经网络（CNN）的查询分类在其卷积运算中使用线性特征映射。循环神经网络（RNN）与表示单词序列的CNN不同，它们的排序信息被明确保持。我们建议使用基于深长短期记忆（DLSTM）的特征映射来学习CNN的特征表示。DLSTM是LSTM单元的堆叠，在LSTM单元的不同深度处具有不同的特征表示次序。底部LSTM单元配备输入和输出门，从当前字提取一阶特征表示。为了提取高阶非线性特征表示，在较高位置的LSTM单元从两个部分获得输入。第一部分是下一个LSTM单元的存储单元。第二部分是较低的LSTM单元从当前字的隐藏输出。以这种方式，DLSTM捕获n-gram内的非线性非连续相互作用。使用将DLSTM层与传统CNN层组合的架构，我们观察到用于查询分类的基准数据集上的新的最先进的查询分类准确性。）

**总结：DLSTM与CNN结合解决查询分类问题。**

1. **Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents**

摘要：The goal of sentence and document modeling is to accurately represent the meaning of sentences and documents for various Natural Language Processing tasks. In this work, we present Dependency Sensitive Convolutional Neural Networks (DSCNN) as a generalpurpose classification system for both sentences and documents. DSCNN hierarchically builds textual representations by processing pretrained word embeddings via Long ShortTerm Memory networks and subsequently extracting features with convolution operators. Compared with existing recursive neural models with tree structures, DSCNN does not rely on parsers and expensive phrase labeling, and thus is not restricted to sentencelevel tasks. Moreover, unlike other CNNbased models that analyze sentences locally by sliding windows, our system captures both the dependency information within each sentence and relationships across sentences in the same document. Experiment results demonstrate that our approach is achieving state-ofthe-art performance on several tasks, including sentiment analysis, question type classification, and subjectivity classification.

（句子和文档建模的目标是准确表示各种自然语言处理任务的句子和文档的含义。在这项工作中，我们提出依赖敏感卷积神经网络（DSCNN）作为句子和文档的通用分类系统。 DSCNN通过经由Long ShortTerm存储器网络处理预训练的字嵌入并随后利用卷积运算符提取特征来分层地构建文本表示。与具有树结构的现有递归神经模型相比，DSCNN不依赖于解析器和昂贵的短语标记，因此不限于句子级任务。此外，与通过滑动窗口在本地分析句子的其他基于CNN的模型不同，我们的系统捕获每个句子内的依赖性信息和同一文档中的句子之间的关系。实验结果表明，我们的方法是在几个任务，包括情感分析，问题类型分类和主观性分类达到最先进的表现。）

**总结：DSCNN解决句子和文档分类问题。**

1. **MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification**

摘要：We introduce a novel, simple convolution neural network (CNN) architecture – multi-group norm constraint CNN (MGNC-CNN) – that capitalizes on multiple sets of word embeddings for sentence classification. MGNCCNN extracts features from input embedding sets independently and then joins these at the penultimate layer in the network to form a final feature vector. We then adopt a group regularization strategy that differentially penalizes weights associated with the subcomponents generated from the respective embedding sets. This model is much simpler than comparable alternative architectures and requires substantially less training time. Furthermore, it is flexible in that it does not require input word embeddings to be of the same dimensionality. We show that MGNC-CNN consistently outperforms baseline models.

（我们介绍一个新的简单的卷积神经网络（CNN）架构-多组规范约束CNN（MGNC-CNN）- 利用多组词嵌入为句子分类。MGNCCNN独立地从输入嵌入集中提取特征，然后在网络中的倒数第二层连接这些特征以形成最终特征向量。然后，我们采用群正则化策略，其对从相应嵌入集产生的子部分相关联的权重进行差分惩罚。该模型比可比较的替换架构简单得多，并且需要相当少的训练时间。此外，它是灵活的，因为它不需要输入字嵌入具有相同的维度。 我们显示MGNC-CNN一贯优于基准模型。）

**总结：MGNC-CNN解决句子分类问题。**

1. **Joint Learning with Global Inference for Comment Classification in Community Question Answering**

摘要：This paper addresses the problem of comment classification in community Question Answering. Following the state of the art, we approach the task with a global inference process to exploit the information of all comments in the answer-thread in the form of a fully connected graph. Our contribution comprises two novel joint learning models that are on-line and integrate inference within learning. The first one jointly learns two node- and edge-level MaxEnt classifiers with stochastic gradient descent and integrates the inference step with loopy belief propagation. The second model is an instance of fully connected pairwise CRFs (FCCRF). The FCCRF model significantly outperforms all other approaches and yields the best results on the task to date. Crucial elements for its success are the global normalization and an Ising-like edge potential.

（本文论述了社区中的评论分类问题。 根据现有技术，我们使用全局推理过程来处理任务，以完全连接的图形的形式利用回答线程中的所有评论的信息。我们的贡献包括两个联机学习模型，在线和集成推理学习。第一个联合学习具有随机梯度下降的两个节点和边缘级MaxEnt分类器，并将推理步骤与循环信任传播相结合。 第二个模型是完全连接的成对CRF（FCCRF）的实例。 FCCRF模型显着优于所有其他方法，并在迄今为止的任务上产生最好的结果。其成功的关键要素是全局正常化和类似于Ising的边缘势。）

1. **Learning to Recognize Ancillary Information for Automatic Paraphrase Identification**

摘要：Previous work on Automatic Paraphrase Identification(PI) is mainly based on modeling text similarity between two sentences. In contrast, we study methods for automatically detecting whether a text fragment only appearing in a sentence of the evaluated sentence pair is important or ancillary information with respect to the paraphrase identification task. Engineering features for this new task is rather difficult, thus, we approach the problem by representing text with syntactic structures and applying tree kernels on them. The results show that the accuracy of our automatic Ancillary Text Classifier (ATC) is promising, i.e., 68.6%, and its output can be used to improve the state of the art in PI.

（以前的自动释义识别（PI）的工作主要基于两个句子之间的文本相似性建模。相反，我们研究了用于自动检测仅出现在所评估的句子对的句子中的文本片段是否是关于复述识别任务的重要或辅助信息的方法。这个新任务的工程特性是相当困难的，因此我们通过用句法结构表示文本并在它们上应用树内核来处理问题。结果表明，我们的自动辅助文本分类器（ATC）的准确性是有希望的，即68.6％，并且其输出可以用于改进PI中的现有技术。）

**有关QA：**

1. **Question Answering over Knowledge Base using Factual Memory Networks**

摘要：In the task of question answering, Memory Networks have recently shown to be quite effective towards complex reasoning as well as scalability, in spite of limited range of topics covered in training data. In this paper, we introduce Factual Memory Network, which learns to answer questions by extracting and reasoning over relevant facts from a Knowledge Base. Our system generate distributed representation of questions and KB in same word vector space, extract a subset of initial candidate facts, then try to find a path to answer entity using multi-hop reasoning and refinement. Additionally, we also improve the run-time efficiency of our model using various computational heuristics.

（在问题回答的任务中，尽管在训练数据中涵盖的主题范围有限，但是存储器网络最近显示对于复杂的推理和可扩展性是相当有效的。在本文中，我们介绍了事实内存网络，它通过从知识库中提取和推理相关事实来学习回答问题。我们的系统在相同的单词向量空间中产生问题和KB的分布式表示，提取初始候选事实的子集，然后尝试使用多跳推理和细化找到回答实体的路径。此外，我们还使用各种计算启发式来提高我们的模型的运行时效率。）

1. **Learning to Compose Neural Networks for Question Answering**

摘要：We describe a question answering model that applies to both images and structured knowledge bases. The model uses natural language strings to automatically assemble neural networks from a collection of composable modules. Parameters for these modules are learned jointly with network-assembly parameters via reinforcement learning, with only (world, question, answer) triples as supervision. Our approach, which we term a dynamic neural module network, achieves state-of-theart results on benchmark datasets in both visual and structured domains.

（我们描述了一个适用于图像和结构化知识库的问题回答模型。该模型使用自然语言字符串从可组合模块的集合中自动组装神经网络。这些模块的参数通过强化学习与网络装配参数联合学习，只有（世界，问题，答案）三元组作为监督。我们的方法，我们称之为动态神经模块网络，在视觉和结构域的基准数据集上实现最先进的结果。）

**有关文本语义相似度：**

1. **Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement**

摘要：Textual similarity measurement is a challenging problem, as it requires understanding the semantics of input sentences. Most previous neural network models use coarse-grained sentence modeling, which has difficulty capturing fine-grained word-level information for semantic comparisons. As an alternative, we propose to explicitly model pairwise word interactions and present a novel similarity focus mechanism to identify important correspondences for better similarity measurement. Our ideas are implemented in a novel neural network architecture that demonstrates state-ofthe-art accuracy on three SemEval tasks and two answer selection tasks.

（文本相似度测量是一个具有挑战性的问题，因为它需要理解输入句子的语义。大多数以前的神经网络模型使用粗粒度的句子建模，其难以捕获细粒度的词级信息用于语义比较。作为替代，我们建议明确建模配对词的相互作用，并提出一个新的相似性聚焦机制，以识别重要的对应更好的相似性测量。我们的想法是在一个新颖的神经网络架构中实现的，展示了三个SemEval任务和两个答案选择任务的最先进的精度。）

**总结：DNN解决文本相似度测量问题。**

1. **Bayesian Supervised Domain Adaptation for Short Text Similarity**

摘要：Identification of short text similarity (STS) is a high-utility NLP task with applications in a variety of domains. We explore adaptation of STS algorithms to different target domains and applications. A two-level hierarchical Bayesian model is employed for domain adaptation (DA) of a linear STS model to text from different sources (e.g., news, tweets). This model is then further extended for multitask learning (MTL) of three related tasks: STS, short answer scoring (SAS) and answer sentence ranking (ASR). In our experiments, the adaptive model demonstrates better overall cross-domain and crosstask performance over two non-adaptive baselines.

（短文本相似性（STS）的识别是具有在各种领域中的应用的高实用性NLP任务。 我们探索STS算法适应不同的目标域和应用程序。 采用两级分级贝叶斯模型用于线性STS模型到来自不同源（例如，新闻，tweet）的文本的域适应（DA）。然后将该模型进一步扩展为三个相关任务的多任务学习（MTL）：STS，短回答计分（SAS）和答案句数排序（ASR）。在我们的实验中，自适应模型展示了比两个非自适应基线更好的整体跨域和交叉表。）

1. **Fast and Easy Short Answer Grading with High Accuracy**

摘要：We present a fast, simple, and high-accuracy short answer grading system. Given a shortanswer question and its correct answer, key measures of the correctness of a student response can be derived from its semantic similarity with the correct answer. Our supervised model (1) utilizes recent advances in the identification of short-text similarity, and (2) augments text similarity features with key grading-specific constructs. We present experimental results where our model demonstrates top performance on multiple benchmarks.

（我们提供一个快速、简单、高精度的短回答分级系统。给定一个简短回答问题及其正确答案，学生反应的正确性的关键措施可以从其与正确答案的语义相似性导出。我们的监督模型（1）利用在识别短文本相似性的最新进展，和（2）增加文本相似性特征与关键的分级特定结构。我们提出实验结果，其中我们的模型演示在多个基准的顶级性能。）

**总结：回答分级，涵盖有关短文本相似、分级结构。**

1. **Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks**

摘要：A key challenge in entity linking is making effective use of contextual information to disambiguate mentions that might refer to different entities in different contexts. We present a model that uses convolutional neural networks to capture semantic correspondence between a mention’s context and a proposed target entity. These convolutional networks operate at multiple granularities to exploit various kinds of topic information, and their rich parameterization gives them the capacity to learn which n-grams characterize different topics. We combine these networks with a sparse linear model to achieve state-of-the-art performance on multiple entity linking datasets, outperforming the prior systems of Durrett and Klein and Nguyen et al.

（实体链接中的一个关键挑战是有效利用上下文信息来消除可能涉及不同上下文中不同实体的提及。我们提出一个模型，使用卷积神经网络捕获提及的上下文和建议的目标实体之间的语义对应。这些卷积网络以多个粒度操作以利用各种类型的主题信息，并且它们丰富的参数化给予他们学习哪些n-gram表征不同主题的能力。我们将这些网络与稀疏线性模型组合，以在多个实体链接数据集上实现最先进的性能，优于先前的Durrett和Klein以及Nguyen等人的系统。）

**总结：CNN解决实体语义对应问题。**

**2015年accepted papers**

1. **Effective Use of Word Order for Text Categorization with Convolutional Neural Networks**

摘要：Convolutional neural network (CNN) is a neural network that can make use of the internal structure of data such as the 2D structure of image data. This paper studies CNN on text categorization to exploit the 1D structure (namely, word order) of text data for accurate prediction. Instead of using low-dimensional word vectors as input as is often done, we directly apply CNN to high-dimensional text data, which leads to directly learning embedding of small text regions for use in classification. In addition to a straightforward adaptation of CNN from image to text, a simple but new variation which employs bag-ofword conversion in the convolution layer is proposed. An extension to combine multiple convolution layers is also explored for higher accuracy. The experiments demonstrate the effectiveness of our approach in comparison with state-of-the-art methods.

（卷积神经网络（CNN）是一种可以利用诸如图像数据的2D结构的数据的内部结构的神经网络。本文研究CNN对文本分类，利用文本数据的1D结构（即词序）进行准确预测。代替使用低维字向量作为输入，如我们经常所做的那样，我们直接将CNN应用于高维文本数据，这导致直接学习嵌入小文本区域以用于分类。 除了CNN从图像到文本的直接适应之外，提出了一种在卷积层中采用袋字转换的简单但新的变体。还研究了组合多个卷积层的扩展以获得更高的精度。 实验证明了我们的方法与现有技术方法相比的有效性。）

**总结：CNN解决文本的分类。**

1. **Convolutional Neural Network for Paraphrase Identification**

摘要：We present a new deep learning architecture Bi-CNN-MI for paraphrase identification (PI). Based on the insight that PI requires comparing two sentences on multiple levels of granularity, we learn multigranular sentence representations using convolutional neural network (CNN) and model interaction features at each level. These features are then the input to a logistic classifier for PI. All parameters of the model (for embeddings, convolution and classification) are directly optimized for PI. To address the lack of training data, we pretrain the network in a novel way using a language modeling task. Results on the MSRP corpus surpass that of previous NN competitors.

（我们提出一个新的深度学习架构Bi-CNN-MI翻译识别（PI）。基于PI需要在多个粒度级别上比较两个句子的洞察，我们使用卷积神经网络（CNN）和每个级别的模型交互特征来学习多粒度句子表示。这些特征是PI的逻辑分类器的输入。模型的所有参数（用于嵌入，卷积和分类）直接针对PI进行优化。为了解决缺乏训练数据，我们使用语言建模任务以一种新颖的方式预训练网络。MSRP语料库的结果超过以前的NN竞争对手。）

1. **Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval**

摘要：Methods of deep neural networks (DNNs) have recently demonstrated superior performance on a number of natural language processing tasks. However, in most previous work, the models are learned based on either unsupervised objectives, which does not directly optimize the desired task, or singletask supervised objectives, which often suffer from insufficient training data. We develop a multi-task DNN for learning representations across multiple tasks, not only leveraging large amounts of cross-task data, but also benefiting from a regularization effect that leads to more general representations to help tasks in new domains. Our multi-task DNN approach combines tasks of multiple-domain classification (for query classification) and information retrieval (ranking for web search), and demonstrates significant gains over strong baselines in a comprehensive set of domain adaptation.

（深度神经网络（DNN）的方法最近已经证明在许多自然语言处理任务上的优越性能。 然而，在大多数以前的工作中，基于不直接优化所期望的任务的无监督目标或者经常遭受不足的训练数据的单目标监督目标来学习模型。 我们开发了一个多任务DNN，用于跨多个任务学习表示，不仅利用大量的跨任务数据，而且还受益于正则化效应，这导致更一般的表示以帮助新域中的任务。 我们的多任务DNN方法结合多域分类（用于查询分类）和信息检索（用于网络搜索的排名）的任务，并且在一组全面的域适配中证明了对强基线的显着增益。）

**该会议其它论文的类型：跨语言翻译、聚类、LSTM等**