目录

[ACL-2016 2](#_Toc472275392)

[1. CNN- and LSTM-based Claim Classification in Online User Comments【较相关---都是用CNN对自然语言文本进行分类】 2](#_Toc472275393)

[2. MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification【较相关---用CNN对句子进行分类】 3](#_Toc472275394)

[3. Incorporating Label Dependency for Answer Quality Tagging in Community Question Answering via CNN-LSTM-CRF【不太相关】 3](#_Toc472275395)

[4. UniMelb at SemEval-2016 Task 3: Identifying Similar Questions by combining a CNN with String Similarity Measures【较相关---问题相似度的比较】 4](#_Toc472275396)

[5. Scalable Semi-Supervised Query Classification Using Matrix Sketching 4](#_Toc472275397)

[Word Embeddings with Limited Memory【不太相关---没有用深度学习的方式】 4](#_Toc472275398)

[6. Hawkes Processes for Continuous Time Sequence Classification: an Application to Rumour Stance Classification in Twitter [不相关] 4](#_Toc472275399)

[7. Science Question Answering using Instructional Materials【不太相关】 5](#_Toc472275400)

[8. An Efficient Cross-lingual Model for Sentence Classification Using Convolutional Neural Network【较相关---使用CNN对句子进行分类】 5](#_Toc472275401)

[9. Word embeddings with limited memory【较相关---都使用了词向量】 5](#_Toc472275402)

[10. QA-It: Classifying Non-Referential It for Question Answer Pairs【不相关】 5](#_Toc472275403)

[AAAI-2016 6](#_Toc472275404)

[11. Convolutional Neural Networks over Tree Structures for Programming Language Processing【不相关】 6](#_Toc472275405)

[12. How Important Is Weight Symmetry in Backpropagation?【不太相关】 6](#_Toc472275406)

[13. All-in Text: Learning Document, Label, and Word Representations Jointly【不相关】 7](#_Toc472275407)

[14. Text Classification with Heterogeneous Information Network Kernels【不太相关】 7](#_Toc472275408)

[15. Instructable Intelligent Personal Agent【不太相关】 8](#_Toc472275409)

[16. Look, Listen and Learn — A Multimodal LSTM for Speaker Identification【不相关】 8](#_Toc472275410)

[17. SPAN: Understanding a Question with Its Support Answers[不相关] 8](#_Toc472275411)

[IJCAI-2016 9](#_Toc472275413)

[18. Recurrent Neural Network for Text Classification with Multi-Task Learning【不太相关】 9](#_Toc472275414)

[19. DrMAD: Distilling Reverse-Mode Automatic Differentiation for Optimizing Hyperparameters of Deep Neural Networks【较相关---涉及到超参数的优化问题】 9](#_Toc472275416)

[20. Towards Convolutional Neural Networks Compression via Global Error Reconstruction【较相关】 10](#_Toc472275418)

[21. Neural Enquirer: Learning to Query Tables in Natural Language【较相关】 11](#_Toc472275420)

[22. Bag-of-Embeddings for Text Classification[不太相关] 11](#_Toc472275422)

[KDD-2016 11](#_Toc472275424)

[23. On-line Active Reward Learning for Policy Optimisation in Spoken Dialogue Systems【不太相关】 11](#_Toc472275425)

[24. Compressing Neural Language Models by Sparse Word Representations【不太相关】 12](#_Toc472275427)

[25. Structural Deep Network Embedding【不太相关】 12](#_Toc472275429)

[26. Compressing Convolutional Neural Networks in the Frequency Domain【较相关】 13](#_Toc472275430)

[27. Large-Scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks【不太相关】 14](#_Toc472275431)

**ACL-2016**

1. [**CNN- and LSTM-based Claim Classification in Online User Comments**](http://aclanthology.info/papers/cnn-and-lstm-based-claim-classification-in-online-user-comments)**【较相关---都是用CNN对自然语言文本进行分类】**

When processing arguments in online user interactive discourse, it is often necessary to determine their bases of support. In this paper, we describe a supervised approach, based on deep neural networks, for classifying the claims made in online arguments. We conduct experiments using convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) on two claim data sets compiled from online user comments. Using different types of distributional word embeddings, but without incorporating any rich, expensive set of features, we achieve a significant improvement over the state of the art for one data set (which categorizes arguments as factual vs. emotional), and performance comparable to the state of the art on the other data set (which categorizes propositions according to their verifiability). Our approach has the advantages of using a generalized, simple, and effective methodology that works for claim categorization on different data sets and tasks.

当在在线用户交互话语中处理论证时，常常需要确定它们的支持的基础。 在本文中，我们描述一种监督方法，基于深层神经网络，用于分类在在线论据的索赔。 我们使用卷积神经网络（CNN）和长期短期记忆网络（LSTM）对从在线用户评论编译的两个索赔数据集进行实验。 使用不同类型的分布式字嵌入，但没有并入任何丰富，昂贵的特征集合，我们实现了对于一个数据集（将参数分类为事实与情绪）的现有技术的显着改进， 其他数据集（根据其可验证性对命题进行分类）的技术水平。 我们的方法具有使用广泛，简单，有效的方法的优点，该方法适用于对不同数据集和任务的权利要求分类。

【解决的问题：对在线用户的谈话进行分类，这里的分类并没有特制是哪种分类。

使用的方法：分别进行了CNN、LSTM两种方式，进行有监督学习

达到的效果：（1）客观的与主观的分类提升显著、（2）可验证的与不可验证的分类达到目前最高水平，本文的模型泛化能力更强、且更简单和高效】

1. [**MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification**](http://aclanthology.info/papers/mgnc-cnn-a-simple-approach-to-exploiting-multiple-word-embeddings-for-sentence-classification)**【较相关---用CNN对句子进行分类】**

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。

【解决的问题：文本分类，可以同时处理多个不等长的嵌入层（句向量）

使用的方法：MGNC-CNN

达到的效果：对句子进行分类，使得训练时长小，且不需要输入嵌入层（句向量）等长】

1. [**Incorporating Label Dependency for Answer Quality Tagging in Community Question Answering via CNN-LSTM-CRF**](http://aclanthology.info/papers/incorporating-label-dependency-for-answer-quality-tagging-in-community-question-answering-via-cnn-lstm-crf)**【不太相关】**

In community question answering (cQA), the quality of answers are determined by the matching degree between question-answer pairs and the correlation among the answers. In this paper, we show that the dependency between the answer quality labels also plays a pivotal role. To validate the effectiveness of label dependency, we propose two neural network-based models, with different combination modes of Convolutional Neural Networks, Long Short Term Memory and Conditional Random Fields. Extensive experiments are taken on the dataset released by the SemEval-2015 cQA shared task. The first model is a stacked ensemble of the networks. It achieves 58.96% on macro averaged F1, which improves the state-of-the-art neural network based method by 2.82% and outperforms the Top-1 system in the shared task by 1.77%. The second is a simple attention-based model whose input is the connection of the question and its corresponding answers. It produces promising results with 58.29% on overall F1 and gains the best performance on the Good and Bad categories.

在社区问答（cQA）中，答案的质量由问答配对之间的匹配程度和答案之间的相关性决定。在本文中，我们表明，答案质量标签之间的依赖关系也发挥了关键作用。为了验证标签依赖的有效性，我们提出两个基于神经网络的模型，卷积神经网络，长短期内存和条件随机场的不同组合模式。对SemEval-2015 cQA共享任务发布的数据集进行大量实验。第一个模型是网络的堆叠集合。它在宏观平均F1上达到58.96％，这将基于神经网络的现有方法改进了2.82％，并且在共享任务中优于Top-1系统1.77％。第二个是一个简单的基于注意的模型，其输入是问题及其相应答案的连接。它产生有希望的结果与总体F1的58.29％，并获得在良好和坏类别的最佳性能。

【解决的问题：问答对中的答案质量评估

使用的方法：CNN、LSTM、CRF（conditional random fields）;利用答案标签之间的依赖关系进行验证

达到的效果：F1达到58.96%（提升2.82%）】

1. [**UniMelb at SemEval-2016 Task 3: Identifying Similar Questions by combining a CNN with String Similarity Measures**](http://aclanthology.info/papers/unimelb-at-semeval-2016-task-3-identifying-similar-questions-by-combining-a-cnn-with-string-similarity-measures)**【较相关---问题相似度的比较】**

This paper describes the results of the participation of The University of Melbourne in the community question-answering (CQA) task of SemEval 2016 (Task 3-B). We obtained a MAP score of 70.2% on the test set, by combining three classifiers: a NaiveBayes classifier and a support vector machine (SVM) each trained over lexical similarity features, and a convolutional neural network (CNN). The CNN uses word embeddings and machine translation evaluation scores as features.

本文介绍了墨尔本大学参加SemEval 2016（任务3-B）的社区问答（CQA）任务的结果。 我们通过组合三个分类器（NaiveBayes分类器和支持向量机（SVM），每个训练在词汇相似性特征上）和卷积神经网络（CNN），在测试集上获得70.2％的MAP分数。 CNN使用字嵌入和机器翻译评估得分作为特征。

【解决的问题：问题相似度比较（这个比赛的论文---University of Melbourne in the community question-answering (CQA) task of SemEval 2016 (Task 3-B)）

使用的方法：CNN（运用了词向量和机器翻译的特征）、贝叶斯分类器（运用词汇相似度作为特征）、SVM（运用词汇相似度为特征）三者的不同组合

达到的效果：MAP score of 70.2%】

1. [**Scalable Semi-Supervised Query Classification Using Matrix Sketching**](http://aclanthology.info/papers/scalable-semi-supervised-query-classification-using-matrix-sketching)

[**Word Embeddings with Limited Memory**](http://aclanthology.info/papers/word-embeddings-with-limited-memory)**【不太相关---没有用深度学习的方式】**

The enormous scale of unlabeled text available today necessitates scalable schemes for representation learning in natural language processing. For instance, in this paper we are interested in classifying the intent of a user query. While our labeled data is quite limited, we have access to virtually an unlimited amount of unlabeled queries, which could be used to induce useful representations: for instance by principal component analysis (PCA). However, it is prohibitive to even store the data in memory due to its sheer size, let alone apply conventional batch algorithms. In this work, we apply the recently proposed matrix sketching algorithm to entirely obviate the problem with scalability (Liberty, 2013). This algorithm approximates the data within a speciﬁed memory bound while preserving the covariance structure necessary for PCA. Using matrix sketching, we signiﬁcantly improve the user intent classiﬁcation accuracy by leveraging large amounts of unlabeled queries.

今天可用的大量未标记的文本需要在自然语言处理中的表示学习的可扩展方案。 例如，在本文中，我们有兴趣对用户查询的意图进行分类。 虽然我们的标签数据相当有限，我们可以访问几乎无限量的未标记的查询，可以用于诱导有用的表示：例如通过主成分分析（PCA）。 然而，由于其庞大的尺寸甚至将数据存储在存储器中是禁止的，更不用说应用常规的批处理算法。 在这项工作中，我们应用最近提出的矩阵草图算法，以完全消除可扩展性的问题（Liberty，2013）。 该算法近似在指定的存储器边界内的数据，同时保留PCA所需的协方差结构。 使用矩阵素描，我们通过利用大量未标记的查询来显着提高用户意图分类的准确性。

【解决的问题：用户查询的意图分类

使用的方法：矩阵草图算法 ，大量未标记的用户查询分类数据

达到的效果：提高了用户意图分类的准确性】

1. [**Hawkes Processes for Continuous Time Sequence Classification: an Application to Rumour Stance Classification in Twitter**](http://aclanthology.info/papers/hawkes-processes-for-continuous-time-sequence-classification-an-application-to-rumour-stance-classification-in-twitter)**[不相关]**

Classification of temporal textual data sequences is a common task in various domains such as social media and the Web. In this paper we propose to use Hawkes Processes for classifying sequences of temporal textual data, which exploit both temporal and textual information. Our experiments on rumour stance classification on four Twitter datasets show the importance of using the temporal information of tweets along with the textual content.

时间文本数据序列的分类是在诸如社交媒体和网络的各种领域中的常见任务。 在本文中，我们建议使用Hawkes过程来分类时间文本数据的序列，利用时间和文本信息。 我们对四个Twitter数据集的谣言姿态分类的实验表明使用tweet的时间信息连同文本内容的重要性。

【解决的问题：时间文本数据序列的分类

使用的方法：Hawkes过程，利用时间和文本信息

达到的效果：提高了时间文本数据序列的分类的准确性】

1. [**Science Question Answering using Instructional Materials**](http://aclanthology.info/papers/science-question-answering-using-instructional-materials)**【不太相关】**

We provide a solution for elementary science tests using instructional materials. We posit that there is a hidden structure that explains the correctness of an answer given the question and instructional materials and present a unified max-margin framework that learns to find these hidden structures(given a corpus of question answer pairs and instructional materials), and uses what it learns to answer novel elementary science questions. Our evaluation shows that our framework outperforms several strong baselines.

我们提供使用教学材料的基础科学测试的解决方案。 我们假设有一个隐藏的结构，解释了给定的问题和教学材料的答案的正确性，并提出一个统一的最大边缘框架，学习找到这些隐藏的结构（给定一个问题答案对和教学材料的语料库） 使用它学会回答新的小学科学问题。 我们的评估表明，我们的框架胜过几个强大的基线。

【解决的问题：假设问题和教材答案之间存在某种隐藏结构，通过学习找到这个隐藏结构，并用之回答新的基本科学问题

使用的方法：LSSVM

达到的效果：】

1. [**An Efficient Cross-lingual Model for Sentence Classification Using Convolutional Neural Network【较相关---使用CNN对句子进行分类】**](http://aclanthology.info/papers/an-efficient-cross-lingual-model-for-sentence-classification-using-convolutional-neural-network)

In this paper, we propose a cross-lingual convolutional neural network (CNN) model that is based on word and phrase embeddings learned from unlabeled data in two languages and dependency grammar. Compared to traditional machine translation (MT) based methods for cross lingual sentence modeling, our model is much simpler and does not need parallel corpora or language specific features. We only use a bilingual dictionary and dependency parser. This makes our model particularly appealing for resource poor languages. We evaluate our model using English and Chinese data on several sentence classification tasks. We show that our model achieves a comparable and even better performance than the traditional MT-based method.

在本文中，我们提出一个跨语言卷积神经网络（CNN）模型，基于从两种语言和依赖语法的未标记数据学习的词和短语嵌入。 与传统的基于机器翻译（MT）的跨语言句子建模方法相比，我们的模型更简单，不需要平行语料库或语言特定的特征。 我们只使用一个双语字典和依赖解析器。 这使我们的模型特别吸引资源贫乏的语言。 我们使用英语和中文数据对几个句子分类任务评估我们的模型。 我们表明，我们的模型实现了一个可比的，甚至比传统的基于MT的方法更好的性能。

【解决的问题：搭建跨语言的句子分类模型

使用的方法：CNN

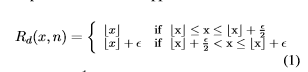
达到的效果：使用英语训练模型，用中文测试模型，结果显示同基于机器翻译的跨语言模型相比，本文的模型在问题分类、观点分析等分类任务上表现得更好】

1. **Word embeddings with limited memory【较相关---都使用了词向量】**

This paper studies the effect of limited precision data representation and computation on word embeddings. We present a systematic evaluation of word embeddings with limited memory and discuss methods that directly train the limited precision representation with limited memory. Our results show that it is possible to use and train an 8-bit fixed-point value for word embedding without loss of performance in word/phrase similarity and dependency parsing tasks.

本文研究了有限精度数据表示和计算对字嵌入的影响。 我们提出了一个有限内存的词嵌入的系统评价和讨论的方法，直接训练有限的精度表示与有限的记忆。 我们的研究结果表明，可以使用和训练一个8位的固定点值的词嵌入，而不会失去性能的词/短语相似性和依赖性解析任务。

【解决的问题：较低内存消耗的词向量表示

使用的方法：暂时不知道什么鬼

达到的效果：在句子相似度计算和语法依赖关系的任务上使用8位句向量即可】

1. [**QA-It: Classifying Non-Referential It for Question Answer Pairs**](http://aclanthology.info/papers/qa-it-classifying-non-referential-it-for-question-answer-pairs)**【不相关】**

This paper introduces a new corpus, QA-It, for the classification of non-referential it. Our dataset is unique in a sense that it is annotated on question answer pairs collected from multiple genres, useful for developing advanced QA systems. Our annotation scheme makes clear distinctions between 4 types of it, providing guidelines for many erroneous cases. Several statistical models are built for the classification of it, showing encouraging results. To the best of our knowledge, this is the first time that such a corpus is created for question answering.

本文介绍了一个新的语料库，QA-It，用于非参考的分类。 我们的数据集是独特的，它被注释在从多个类型收集的问题答案对，有用于开发先进的QA系统。 我们的注释方案明确区分其4种类型，为许多错误的情况提供指导。 为其分类建立了几个统计模型，显示令人鼓舞的结果。 就我们所知，这是第一次创建这样的语料库用于问题回答。

【解决的问题：对社区问答中的答案按照参考价值进行分类

使用的方法：。。。

达到的效果：论文作者认为他们史无前例，待考究】

**AAAI-2016**

1. [**Convolutional Neural Networks over Tree Structures for Programming Language Processing**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11775)**【不相关】**

Programming language processing (similar to natural language processing) is a hot research topic in the field of software engineering; it has also aroused growing interest in the artificial intelligence community. However, different from a natural language sentence, a program contains rich, explicit, and complicated structural information. Hence, traditional NLP models may be inappropriate for programs. In this paper, we propose a novel tree-based convolutional neural network (TBCNN) for programming language processing, in which a convolution kernel is designed over programs’ abstract syntax trees to capture structural information. TBCNN is a generic architecture for programming language processing; our experiments show its effectiveness in two different program analysis tasks: classifying programs according to functionality, and detecting code snippets of certain patterns. TBCNN outperforms base line methods, including several neural models for NLP.

编程语言处理（类似于自然语言处理）是软件工程领域的一个热点研究课题; 它也引起了人们对人工智能社区的越来越大的兴趣。 然而，与自然语言句子不同，程序包含丰富，明确和复杂的结构信息。 因此，传统的NLP模型可能不适合于程序。 在本文中，我们提出了一种新的基于树的卷积神经网络（TBCNN）用于编程语言处理，其中卷积核设计在程序的抽象语法树上以捕获结构信息。 TBCNN是用于编程语言处理的通用体系结构; 我们的实验表明它在两个不同的程序分析任务的有效性：根据功能分类程序，并检测某些模式的代码片段。 TBCNN优于基线方法，包括NLP的几个神经模型。

【解决的问题：类似于自然语言处理，本文研究的是编程语言的处理

使用的方法：TBCNN

达到的效果：】

1. [**How Important Is Weight Symmetry in Backpropagation?**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12325)**【不太相关】**

Gradient backpropagation (BP) requires symmetric feedforward and feedback connections—the same weights must be used for forward and backward passes. This “weight transport problem” (Gross berg 1987) is thought to be one of the main reasons to doubt BP’s biologically plausibility. Using 15 different classification datasets, we systematically investigate to what extent BP really depends on weight symmetry. In a study that turned out to be surprisingly similar inspirit to Lillicrap et al.’s demonstration (Lillicrap et al. 2014) but orthogonal in its results, our experiments indicate that: (1) the magnitudes of feedback weights do not matter to performance (2) the signs of feedback weights do matter—the more concordant signs between feedforward and their corresponding feedback connections, the better (3) with feedback weights having random magnitudes and 100% concordant signs, we were able to achieve the same or even better performance than SGD. (4) some normalizations/stabilizations are indispensable for such a symmetric BP to work, namely Batch Normalization (BN) (Ioffe and Szegedy 2015) and/or a “Batch Manhattan” (BM) update rule.

梯度反向传播（BP）需要对称前馈和反馈连接 - 前向和反向传递必须使用相同的权重。这种“重量转移问题”（Gross berg 1987）被认为是怀疑BP生物学可信性的主要原因之一。使用15个不同的分类数据集，我们系统地研究BP真正取决于重量对称性的程度。在一项研究中，Lillicrap等人的证明（Lillicrap et al。2014）显示出令人惊讶的类似灵感，但在其结果中是正交的，我们的实验表明：（1）反馈权重的大小与性能无关（2）反馈权重的符号有关 - 前馈和它们相应的反馈连接之间的更一致的符号，具有随机幅度和100％一致符号的反馈权重的更好（3），我们能够实现相同或甚至更好性能比SGD。 （4）一些归一化/稳定对于这种对称BP工作是必不可少的，即批量标准化（BN）（Ioffe和Szegedy 2015）和/或“批量曼哈顿”（BM）更新规则。

【解决的问题：反向传播中反馈权重对称有多重要，只是介绍重要性

使用的方法：

达到的效果：（1）反馈权重的大小与性能无关（2）反馈权重的符号有关 - 前馈和它们相应的反馈连接之间的更一致的符号，具有随机幅度和100％一致符号的反馈权重的更好（3），我们能够实现相同或甚至更好性能的SGD。 （4）一些归一化/稳定对于这种对称BP工作是必不可少的，即批量标准化（BN）（Ioffe和Szegedy 2015）和/或“批量曼哈顿”（BM）更新规则。】

1. [**All-in Text: Learning Document, Label, and Word Representations Jointly**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12058)**【不相关】**

Conventional multi-label classification algorithms treat the target labels of the classification task as mere symbols that are void of an inherent semantics. However, in many cases textual descriptions of the se labels are available or can be easily constructed from public document sources such as Wikipedia. In this paper, we investigate an approach for embedding documents and labels into a joint space while sharing word representations between documents and labels. For finding such embeddings, we rely on the text of documents as well as descriptions for the labels. The use of such label descriptions not only lets us expect an increased performance on conventional multi-label text classification tasks, but can also be used to make predictions for labels that have not been seen during the training phase. The potential of our method is demonstrated on the multi-label classification task of assigning keywords from the Medical Subject Headings (MeSH) to publications in biomedical research, both in a conventional and in a zero-shot learning setting.

常规的多标签分类算法将分类任务的目标标签作为仅没有固有语义的符号。然而，在许多情况下，对标签的文本描述是可用的，或者可以容易地从诸如维基百科的公共文档源构建。在本文中，我们调查了一种将文档和标签嵌入到联合空间中，同时在文档和标签之间共享单词表示的方法。为了找到这样的嵌入，我们依赖文档的文本以及标签的描述。使用这样的标签描述不仅使我们期望在常规多标签文本分类任务上的增加的性能，而且还可以用于对在训练阶段期间没有看到的标签进行预测。我们的方法的潜力展示在多标签分类任务中，将关键词从医学主题标题（MeSH）分配到生物医学研究中的出版物，无论是在传统的还是在零的学习环境。

【解决的问题：常见的多标签分类仅仅将标签看作一个没有固定语义的词，而没有考虑到标签的文档描述（比如维基百科对这个词的解释），本文文档和标签共同嵌入词向量，提高多标签分类的性能

使用的方法：

达到的效果：】

1. [**Text Classification with Heterogeneous Information Network Kernels**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12392)**【不太相关】**

Text classification is an important problem with many applications. Traditional approaches represent text as a bag of-words and build classifiers based on this representation. Rather than words, entity phrases, the relations between the entities, as well as the types of the entities and relations carry much more information to represent the texts. This paper presents a novel text as network classification framework, which introduces 1) a structured and typed heterogeneous information networks (HINs) representation of texts, and 2) a meta-path based approach to link texts. We show that with the new representation and links of texts, the structured and typed information of entities and relations can be incorporated into kernels. Particularly, we develop both simple linear kernel and indefinite kernel based on meta paths in the HIN representation of texts, where we call them HIN-kernels. Using Freebase, a well-known world knowledge base, to construct HIN for texts, our experiments on two benchmark datasets show that the indefinite HIN-kernel based on weighted meta-paths outperforms the state-of-the art methods and other HIN-kernels.

文本分类是许多应用程序的重要问题。传统方法将文本表示为词袋并基于该表示来构建分类器。而不是词，实体短语，实体之间的关系，以及实体和关系的类型携带更多的信息来表示文本。本文提出了一种新颖的文本作为网络分类框架，其中介绍了1）结构化和类型的异质信息网络（HIN）表示文本，2）基于元路径的方法链接文本。我们表明，使用新的表示和文本的链接，实体和关系的结构化和类型化的信息可以被并入内核。特别地，我们在文本的HIN表示中基于元路径开发简单线性内核和不确定内核，其中我们称之为HIN-内核。使用Freebase，一个知名的世界知识库，为文本构建HIN，我们在两个基准数据集上的实验表明，基于加权元路径的不确定的HIN内核超过了最先进的方法和其他HIN内核。

【解决的问题：考虑了文档之间关系、句子以及实体类别等因素的文档分类问题

使用的方法：

达到的效果：】

1. [**Instructable Intelligent Personal Agent**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12383)**【不太相关】**

Unlike traditional machine learning methods, humans often learn from natural language instruction. As users become increasingly accustomed to interacting with mobile devices using speech, their interest in instructing these devices in natural language is likely to grow. We introduce our Learning by Instruction Agent(LIA), an intelligent personal agent that users can teach to perform new action sequences to achieve new commands, using solely natural language interaction. LIA uses a CCG semantic parser to ground the semantics of each command in terms of primitive executable procedures defining sensors and effectors of the agent. Given a natural language command that LIA does not understand, it prompts the user to explain how to achieve the command through a sequence of steps, also specified in natural language. A novel lexicon induction algorithm enables LIA to generalize across taught commands, e.g., having been taught how to “forward an email to Alice,” LIA can correctly interpret the command “forward this email to Bob.” A user study involving email tasks demonstrates that users voluntarily teach LIA new commands, and that these taught commands significantly reduce task completion time. These results demonstrate the potential of natural language instruction as a significant, under-explored paradigm for machine learning.

与传统的机器学习方法不同，人类经常学习自然语言教学。随着用户越来越习惯于使用语音与移动设备交互，他们对以自然语言指示这些设备的兴趣可能增长。我们介绍了我们的学习指导代理（LIA），一个智能的个人代理，用户可以教它执行新的动作序列来实现新的命令，仅使用自然语言交互。 LIA使用CCG语义解析器，根据定义代理的传感器和效应器的原始可执行过程，使每个命令的语义接地。给定LIA不理解的自然语言命令，它提示用户解释如何通过也以自然语言指定的步骤序列来实现命令。新颖的词典诱导算法使得LIA能够在教导的命令之间概括，例如，已经教导了如何“将电子邮件转发到Alice”，LIA可以正确地解释命令“将该电子邮件转发给Bob”。涉及电子邮件任务的用户研究表明，用户自愿地教LIA新命令，并且这些教导的命令显着减少任务完成时间。这些结果表明自然语言教学作为机器学习的重要的，未充分研究的范例的潜力。

【解决的问题：用户使用自然语言同系统交互，教系统执行新的动作序列来实现新的命令

使用的方法：CGG语义解析器、传感器等

达到的效果：】

1. [**Look, Listen and Learn — A Multimodal LSTM for Speaker Identification**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12386)**【不相关】**

Speaker identification refers to the task of localizing the face of a person who has the same identity as the ongoing voice in a video. This task not only requires collective perception over both visual and auditory signals, the robustness to handle severe quality degradations and unconstrained content variations are also indispensable. In this paper, we describe a novel multimodal Long Short-Term Memory (LSTM) architecture which seamlessly unifies both visual and auditory modalities from the beginning of each sequence input. The key idea is to extend the conventional LSTM by not only sharing weights across time steps, but also sharing weights across modalities. We show that modeling the temporal dependency across face and voice can significantly improve the robustness to content quality degradations and variations. We also found that our multimodal LSTM is robustness to distractors, namely the non-speaking identities. We applied our multimodal LSTM to The Big Bang Theory dataset and showed that our system outperforms the state-of-the-art systems in speaker identification with lower false alarm rate and higher recognition accuracy.

讲话者识别是指将与正在进行的语音具有相同身份的人的面部定位在视频中的任务。这个任务不仅需要对视觉和听觉信号的集体感知，处理严重的质量劣化和不受约束的内容变化的鲁棒性也是不可缺少的。在本文中，我们描述一个新的多模态长短期记忆（LSTM）架构，从每个序列输入的开始无缝地统一视觉和听觉模态。关键思想是通过不仅跨越时间步长共享权重，而且跨模态共享权重来扩展常规LSTM。我们显示建模的时间依赖性跨面和语音可以显着提高内容质量退化和变化的鲁棒性。我们还发现，我们的多模态LSTM是对干扰者的鲁棒性，即不说话的身份。我们将我们的多模式LSTM应用于大爆炸理论数据集，并且表明我们的系统优于说话人识别中的最先进的系统，具有较低的误报率和较高的识别精度。

【解决的问题：通过图像和音频定位正在说话的人

使用的方法：LSTM

达到的效果：说话人识别最优系统】

1. [**SPAN: Understanding a Question with Its Support Answers**](http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11888)**[不相关]**

Matching a question to its best answer is a common task in community question answering. In this paper, we focus on the non-factoid questions and aim to pick out the best answer from its candidate answers. Most of the existing deep models directly measure the similarity between question and answer by their individual sentence embeddings. In order to tackle the problem of the information lack in question’s descriptions and the lexical gap between questions and answers, we propose a novel deep architecture namely SPAN in this paper. Specifically we introduce support answers to help understand the question, which are defined as the best answers of those similar questions to the original one. Then we can obtain two kinds of similarities, one is between question and the candidate answer, and the other one is between support answers and the candidate answer. The matching score is finally generated by combining them. Experiments on Yahoo! Answers demonstrate that SPAN can outperform the baseline models.

将问题与其最佳答案相匹配是社区问题回答中的常见任务。在本文中，我们专注于非事实型问题，旨在从候选答案中挑选最佳答案。大多数现有的深度模型直接测量问题和答案之间的相似性通过他们各自的句子嵌入。为了解决信息缺乏问题的描述和问题和答案之间的词汇差距的问题，我们提出一个新的深层架构，即本文中的SPAN。具体来说，我们介绍支持答案以帮助理解问题，这被定义为那些类似问题的最佳答案与原始答案。然后我们可以获得两种相似性，一种是在问题和候选答案之间，另一种是在支持答案和候选答案之间。通过组合它们最终生成匹配分数。在Yahoo! Answers上的实验证明SPAN可以胜过基准模型。

【解决的问题：给问题匹配最佳答案，比如论文上一个问题有很多答案，现有的技术只是通过词向量计算答案和问题的相似度，并没有考虑到更多信息。

使用的方法：SPAN

达到的效果：在yahoo!answers上证明性能优于基准模型】

**IJCAI-2016**

1. [**Recurrent Neural Network for Text Classification with Multi-Task Learning**](http://xueshu.baidu.com/s?wd=paperuri%3A%2855c87a8566543c206403233d8e99905f%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Farxiv.org%2Fabs%2F1605.05101&ie=utf-8&sc_us=16522293048132723077)**【不太相关】**

Abstract Neural network based methods have obtained great progress on a variety of natural language processing tasks. However, in most previous works, the models are learned based on single-task supervised objectives, which often suffer from insufficient training data. In this paper, we use the multi-task learning framework to jointly learn across multiple related tasks. Based on recurrent neural network, we propose three different mechanisms of sharing information to model text with task-specific and shared layers. The entire network is trained jointly on all these tasks. Experiments on four benchmark text classification tasks show that our proposed models can improve the performance of a task with the help of other related tasks.

摘要基于神经网络的方法在各种自然语言处理任务中取得了巨大进步。 然而，在大多数以前的工作中，基于单任务监督目标来学习模型，其经常遭受不足的训练数据。 在本文中，我们使用多任务学习框架来联合学习跨多个相关任务。 基于循环神经网络，我们建议三种不同的机制共享信息模型文本与任务特定和共享层。 整个网络在所有这些任务上联合训练。 关于四个基准文本分类任务的实验表明，我们提出的模型可以在其他相关任务的帮助下提高任务的性能。

【解决的问题：利用相关任务之间的相关性，同步进行多任务监督学习，同时优化系统参数，使得共享的底层部分共同促进不同任务的学习

使用的方法：

达到的效果：】

1. [**DrMAD: Distilling Reverse-Mode Automatic Differentiation for Optimizing Hyperparameters of Deep Neural Networks**](http://xueshu.baidu.com/s?wd=paperuri%3A%28ae3bdfa4b9545e61bad874ad91fc5730%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Farxiv.org%2Fabs%2F1601.00917&ie=utf-8&sc_us=8035307294130185275)**【较相关---涉及到超参数的优化问题】**

The performance of deep neural networks is well-known to be sensitive to the setting of their hyperparameters. Recent advances in reverse-mode automatic differentiation allow for optimizing hyperparameters with gradients. The standard way of computing these gradients involves a forward and backward pass of computations. However, the backward pass usually needs to consume unaffordable memory to store all the intermediate variables to exactly reverse the forward training procedure. In this work we propose a simple but effective method, DrMAD, to distill the knowledge of the forward pass into a shortcut path, through which we approximately reverse the training trajectory. Experiments on two image benchmark datasets show that DrMAD is at least 45 times faster and consumes 100 times less memory compared to state of-the-art methods for optimizing hyperparameters with minimal compromise to its effectiveness. To the best of our knowledge, DrMAD is the first research attempt to make it practical to automatically tunethousands of hyperparameters of deep neural networks. The code can be downloaded from <https://github.com/bigaidream-projects/drmad/>

深层神经网络的性能是众所周知的，对其超参数的设置是敏感的。反向模式自动微分的最新进展允许用梯度优化超参数。计算这些梯度的标准方法涉及计算的向前和向后传递。然而，反向传递通常需要消耗不可负担的存储器来存储所有中间变量以精确地反转前向训练过程。在这项工作中，我们提出了一个简单而有效的方法，DrMAD，将正向传递的知识提炼成快捷路径，通过它我们大约颠倒了训练轨迹。对两个图像基准数据集的实验表明，与用于优化超参数的最先进的方法相比，DrMAD至少快45倍并消耗100倍的内存，而对其有效性的影响最小。据我们所知，DrMAD是第一个研究尝试，以使其实用于自动调节深部神经网络的超参数。

【解决的问题：一种新的超参数优化模型，有效降低内存需求

使用的方法：DrMAD，取消反向传播

达到的效果：速度快，内存消耗低】

1. [**Towards Convolutional Neural Networks Compression via Global Error Reconstruction**](http://ijcai.org/Proceedings/16/Papers/251.pdf)**【较相关】**

In recent years, convolutional neural networks (CNNs) have achieved remarkable success in var- ious applications such as image classification, ob- ject detection, object parsing and face alignment. Such CNN models are extremely powerful to deal with massive amounts of training data by using mil- lions and billions of parameters. However, these models are typically deficient due to the heavy cost in model storage, which prohibits their usage on resource-limited applications like mobile or em- bedded devices. In this paper, we target at com- pressing CNN models to an extreme without sig- nificantly losing their discriminability. Our main idea is to explicitly model the output reconstruction error between the original and compressed CNNs, which error is minimized to pursuit a satisfactory rate-distortion after compression. In particular, a global error reconstruction method termed GER is presented, which firstly leverages an SVD-based low-rank approximation to coarsely compress the parameters in the fully connected layers in a layer- wise manner. Subsequently, such layer-wise initial compressions are jointly optimized in a global per- spective via back-propagation. The proposed GER method is evaluated on the ILSVRC2012 image classification benchmark, with implementations on two widely-adopted convolutional neural networks, i.e., the AlexNet and VGGNet-19. Comparing to several state-of-the-art and alternative methods of CNN compression, the proposed scheme has demonstrated the best rate-distortion performance on both networks.

近年来，卷积神经网络（CNN）在各种应用中取得了显着的成功，例如图像分类，对象检测，对象解析和面对齐。这种CNN模型非常强大，能够通过使用数百万和数十亿的参数来处理大量的培训数据。然而，由于模型存储中的高成本，这些模型通常是不足的，这限制了它们在诸如移动或嵌入式设备的资源有限的应用上的使用。在本文中，我们的目标是将CNN模型压缩到极限，而不会显着失去其可辨别性。我们的主要思想是明确建模原始和压缩CNN之间的输出重建误差，该误差被最小化以在压缩之后追求满意的速率失真。特别地，提出了称为GER的全局误差重建方法，其首先利用基于SVD的低秩近似来以分层方式粗略地压缩完全连接的层中的参数。随后，通过反向传播在全局视觉中联合优化这种层式初始压缩。所提出的GER方法在ILSVRC2012图像分类基准上进行评估，在两个广泛采用的卷积神经网络（即AlexNet和VGGNet-19）上实现。与CNN压缩的几种现有技术和替代方法相比，所提出的方案已经证明了在两个网络上的最佳速率失真性能。

【解决的问题：对CNN模型的压缩方法，节省内存空间

使用的方法：SVD等

达到的效果：相比其他压缩方式，已经证明在AlexNet和VGGNet-19上有最佳的失真性能】

1. [**Neural Enquirer: Learning to Query Tables in Natural Language**](http://www.researchgate.net/publication/306093738_Neural_Enquirer_Learning_to_Query_Tables_in_Natural_Language)**【不太相关---因为我们的项目在问题匹配过程中并没有考虑答案信息，所以我认为我们的系统更像一个客服系统，而不是问答系统】**

We propose NEURAL ENQUIRER — a neural network architecture for answering natural language (NL) questions given a knowledge base (KB) table. Unlike previous work on end-to-end training of semantic parsers, NEURAL ENQUIRER is fully “neuralized”: it gives distributed representations of queries and KB tables, and executes queries through a series of differentiable operations. The model can be trained with gradient descent using both endto-end and step-by-step supervision. During training the representations of queries and the KB table are jointly optimized with the query execution logic. Our experiments show that the model can learn to execute complex NL queries on KB tables with rich structures.

我们建议神经ENQUIRER - 一个神经网络架构，用于回答自然语言（NL）问题给定的知识库（KB）表。 与以前关于语义解析器的端到端训练的工作不同，NEURAL ENQUIRER是完全“神经化的”：它给出查询和KB表的分布式表示，并通过一系列可区分的操作执行查询。 该模型可以使用端到端和逐步监控的梯度下降来训练。 在训练期间，查询和KB表的表示与查询执行逻辑联合优化。 我们的实验表明，该模型可以学习在具有丰富结构的KB表上执行复杂的NL查询。

【解决的问题：一种神经网络架构：用于在给定知识库情况下回答自然语言问题

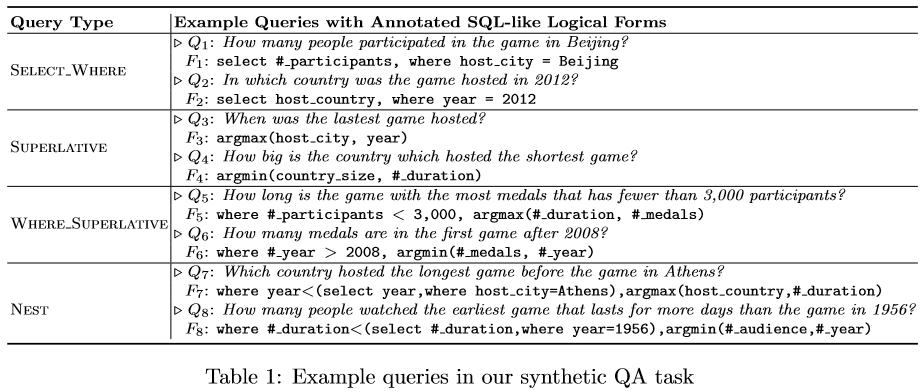
使用的方法：分成问题向量生成模块、答案向量生成模块、操作执行模块。严格依照知识库中的信息进行解答。

达到的效果：】

知识库存储的信息如下：



进行的QA举例：



1. [**Bag-of-Embeddings for Text Classification**](http://www.ijcai.org/Proceedings/16/Papers/401.pdf)**[不太相关]**

Words are central to text classification. It has been shown that simple Naive Bayes models with word and bigram features can give highly competitive accuracies when compared to more sophisticated models with part-of-speech, syntax and semantic features. Embeddings offer distributional features about words. We study a conceptually simple classification model by exploiting multiprototype word embeddings based on text classes. The key assumption is that words exhibit different distributional characteristics under different text classes. Based on this assumption, we train multiprototype distributional word representations for different text classes. Given a new document, its text class is predicted by maximizing the probabilities of embedding vectors of its words under the class. In two standard classification benchmark datasets, one is balance and the other is imbalance, our model outperforms state-of-the-art systems, on both accuracy and macro-average F-1 score.

词语是文本分类的核心。 已经表明，与具有词性，语法和语义特征的更复杂的模型相比，具有词和双元语法特征的简单朴素贝叶斯模型可以给出高度竞争的准确性。 嵌入提供关于单词的分布特征。 我们通过利用基于文本类的多原型字嵌入来研究概念上简单的分类模型。 关键假设是在不同的文本类下，单词表现出不同的分布特征。 基于这个假设，我们训练不同文本类的多原型分布词表示。 给定一个新文档，它的文本类通过最大化其类的词的嵌入向量的概率来预测。 在两个标准分类基准数据集中，一个是平衡，另一个是不平衡，我们的模型在精度和宏观平均F-1分数上胜过最先进的系统。

【解决的问题：文本分类：用词语的分布式表示模型来预测文本分类

使用的方法：

达到的效果：】

**KDD-2016**

1. **On-line Active Reward Learning for Policy Optimisation in Spoken Dialogue Systems【不太相关】**

The ability to compute an accurate reward function is essential for optimizing a dialogue policy via reinforcement learning. In real-world applications, using explicit user feedback as the reward signal is often unreliable and costly to collect. This problem can be mitigated if the user’s intent is known in advance or data is available to pre-train a task success predictor off-line. In practice neither of these apply for most real world applications. Here we propose an on-line learning framework whereby the dialogue policy is jointly trained alongside the reward model via active learning with a Gaussian process model. This Gaussian process operates on a continuous space dialogue representation generated in an unsupervised fashion using a recurrent neural network encoder-decoder. The experimental results demonstrate that the proposed framework is able to signiﬁcantly reduce data annotation costs and mitigate noisy user feedback in dialogue policy learning.

计算准确的报酬函数的能力对于通过强化学习优化对话政策至关重要。 在现实世界的应用中，使用明确的用户反馈作为奖励信号通常是不可靠的并且收集昂贵的。 如果用户的意图是预先已知的或者数据可用于离线预先训练任务成功预测器，则可以减轻这个问题。 在实践中，这些都不适用于大多数真实世界的应用。 在这里我们提出一个在线学习框架，对话政策与奖励模型一起联合训练，通过主动学习与高斯过程模型。 该高斯过程对使用循环神经网络编码器 - 解码器以无监督方式产生的连续空间对话表示进行操作。 实验结果表明，所提出的框架能够显着降低数据注释成本，并减少对话策略学习中嘈杂的用户反馈。

【解决的问题：在线学习框架，获取用户反馈并降低反馈噪声

使用的方法：RNN

达到的效果：】

1. [**Compressing Neural Language Models by Sparse Word Representations**](http://www.researchgate.net/publication/306093382_Compressing_Neural_Language_Models_by_Sparse_Word_Representations)**【不太相关】**

Abstract Neural networks are among the state-of-the-art techniques for language modeling. Existing neural language models typically map discrete words to distributed, dense vector representations. After information processing of the preceding context words by hidden layers, an output layer estimates the probability of the next word. Such approaches are time- and memory-intensive because of the large numbers of parameters for word embeddings and the output layer. In this paper, we propose to compress neural language models by sparse word representations. In the experiments, the number of parameters in our model increases very slowly with the growth of the vocabulary size, which is almost imperceptible. Moreover, our approach not only reduces the parameter space to a large extent, but also improves the performance in terms of the perplexity measure.

摘要神经网络是语言建模的最先进的技术之一。 现有的神经语言模型通常将离散字映射到分布的密集向量表示。 在通过隐藏层对前面的上下文字进行信息处理之后，输出层估计下一个字的概率。 这样的方法是时间和存储器密集的，因为用于字嵌入和输出层的大量参数。 在本文中，我们建议通过稀疏词表示来压缩神经语言模型。 在实验中，我们模型中的参数数量随着词汇量的增长而增长非常缓慢，这几乎是不可察觉的。 此外，我们的方法不仅在很大程度上减少了参数空间，而且还改善了在困惑度测量方面的性能。

【解决的问题：用稀疏词表示法压缩神经语言模型

使用的方法：

达到的效果：】

1. **Structural Deep Network Embedding【不太相关】**

Network embedding is an important method to learn low-dimensional representations of vertexes in networks, aiming to capture and preserve the network structure. Almost all the existing network embedding methods adopt shallow models. However, since the underlying network structure is complex, shallow models cannot capture the highly non-linear network structure, resulting in sub-optimal network representations. Therefore, how to find a method that is able to effectively capture the highly non-linear network structure and preserve the global and local structure is an open yet important problem. To solve this problem, in this paper we propose a Structural Deep Network Embedding method, namely SDNE. More specifically, we first propose a semi-supervised deep model, which has multiple layers of non-linear functions, thereby being able to capture the highly non-linear network structure. Then we propose to exploit the first-order and second-order proximity jointly to preserve the network structure. The second-order proximity is used by the unsupervised component to capture the global network structure. While the first-order proximity is used as the supervised information in the supervised component to preserve the local network structure. By jointly optimizing them in the semi-supervised deep model, our method can preserve both the local and global network structure and is robust to sparse networks. Empirically, we conduct the experiments on five real-world networks, including a language network, a citation network and three social networks. The results show that compared to the baselines, our method can reconstruct the original network significantly better and achieves substantial gains in three applications, i.e. multi-label classification, link pre-diction and visualization.

网络嵌入是学习网络中顶点的低维表示的重要方法，旨在捕获和保留网络结构。几乎所有现有的网络嵌入方法都采用浅层模型。然而，由于底层网络结构是复杂的，浅层模型不能捕获高度非线性网络结构，导致次优网络表示。因此，如何找到能够有效捕获高度非线性网络结构并保持全局和局部结构的方法是一个开放但重要的问题。为了解决这个问题，在本文中，我们提出了一种结构深度网络嵌入方法，即SDNE。更具体地，我们首先提出一种半监督深度模型，其具有多层非线性函数，从而能够捕获高度非线性网络结构。然后，我们建议利用一阶和二阶接近共同保持网络结构。第二级接近由无监督组件使用以捕获全局网络结构。虽然第一级接近度被用作受监视组件中的监督信息以保留局部网络结构。通过在半监督深度模型中联合优化它们，我们的方法可以保留局部和全局网络结构，并且对稀疏网络是鲁棒的。根据经验，我们在五个真实世界网络上进行实验，包括语言网络，引文网络和三个社交网络。结果表明，与基线相比，我们的方法可以显着更好地重建原始网络，并在三个应用程序，即多标签分类，链接预测和可视化中实现实质性增益。

【解决的问题：结构化的深度模型嵌入表示

使用的方法：

达到的效果：】

1. **Compressing Convolutional Neural Networks in the Frequency Domain【较相关】**

Convolutional neural networks (CNN) are increasingly used in many areas of computer vision. They are particularly attractive because of their ability to “absorb” great quantities of labeled data through millions of parameters. However, as model sizes increase, so do the storage and memory requirements of the classifiers, hindering many applications such as image and speech recognition on mobile phones and other devices. In this paper, we present a novel network architecture, Frequency-Sensitive Hashed Nets (FreshNets), which exploits inherent redundancy in both convolutional layers and fully-connected layers of a deep learning model, leading to dramatic savings in memory and storage consumption. Based on the key observation that the weights of learned convolutional filters are typically smooth and low-frequency, we first convert filter weights to the frequency domain with a discrete cosine transform (DCT) and use a low-cost hash function to randomly group frequency parameters into hash buckets. All parameters assigned the same hash bucket share a single value learned with standard back-propagation. To further reduce model size, we allocate fewer hash buckets to high-frequency components, which are generally less important. We evaluate FreshNets on eight data sets, and show that it leads to better compressed performance than several relevant baselines.

卷积神经网络（CNN）越来越多地用于计算机视觉的许多领域。它们特别有吸引力，因为它们能够通过数百万个参数“吸收”大量的标记数据。然而，随着模型尺寸增加，分类器的存储和存储器需求也增加，阻碍了诸如移动电话和其他设备上的图像和语音识别的许多应用。在本文中，我们提出一个新的网络架构，频率敏感散列网络（FreshNets），利用卷积层和深层学习模型的完全连接层的固有冗余，导致显着节省内存和存储消耗。基于学习的卷积滤波器的权重通常是平滑和低频的关键观察，我们首先利用离散余弦变换（DCT）将滤波器权重转换到频域，并使用低成本哈希函数来随机地对频率参数进行分组到哈希桶。分配了相同散列桶的所有参数共享使用标准反向传播学习的单个值。为了进一步减小模型大小，我们将较少的哈希桶分配给高频分量，这通常不太重要。我们评估FreshNets上的八个数据集，并表明它导致比几个相关基线更好的压缩性能。

【解决的问题：压缩CNN神经网络（主要是全连接层参数冗余）

使用的方法：FreshNets、hash、DCT

达到的效果：在八个数据集上上验证优于其他压缩方式】

1. **Large-Scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks【不太相关】**

Precise item categorization is a key issue in e-commerce domains. However, it still remains a challenging problem due to data size, category skewness, and noisy metadata. Here, we demonstrate a successful report on a deep learning-based item categorization method, i.e., deep categorization network (DeepCN), in an e-commerce website. DeepCN is an end-to-end model using multiple recurrent neural networks (RNNs) dedicated to metadata attributes for generating features from text metadata and fully connected layers for classifying item categories from the generated features. The categorization errors are propagated back through the fully connected layers to the RNNs for weight update in the learning process. This deep learning-based approach allows diverse attributes to be integrated into a common representation, thus overcoming sparsity and scalability problems. We evaluate DeepCN on large-scale real-world data including more than 94 million items with approximately 4,100 leaf categories from a Korean e-commerce website. Experiment results show our method improves the categorization accuracy compared to the model using single RNN as well as a standard classification model using unigram-based bag-of-words. Furthermore, we investigate how much the model parameters and the used attributes influence categorization performances.

精确的项目分类是电子商务领域的一个关键问题。然而，由于数据大小，类别偏斜和嘈杂的元数据，它仍然是一个具有挑战性的问题。在这里，我们演示了一个成功的报告，基于深度学习的项目分类方法，即深度分类网络（DeepCN），在一个电子商务网站。 DeepCN是一种端到端模型，使用专用于元数据属性的多个循环神经网络（RNN），用于从文本元数据和完全连接的层生成特征，以从生成的特征中分类项目类别。分类误差通过完全连接的层传播回到RNN以用于学习过程中的重量更新。这种基于深度学习的方法允许将各种属性集成到一个通用表示中，从而克服稀疏性和可伸缩性问题。我们从韩国电子商务网站评估DeepCN的大型真实世界数据，包括超过9400万个项目，大约4,100个叶类别。实验结果表明，与使用单一RNN的模型相比，我们的方法提高了分类精度，以及使用基于单字母词袋的标准分类模型。此外，我们调查多少模型参数和使用的属性影响分类性能。

【解决的问题：电子商务产品分类问题

使用的方法：多个RNN

达到的效果：】

自然科学基金：

1. 需为热点方向
2. 解决科学问题（算法运行中遇到的问题、算法调度、CPU内存硬盘创建虚机，并非现实生活中的需求）=>自然语言处理的挑战、科学建模成数学问题
3. CNN、问答系统=>调研、海量泛读（方向一致的文章标题）=>论文精读=>先确定问题，再寻求方法
4. 国内外研究状况综述（CNN、NLP、问答系统领域=>算法精度、复杂性、内存加载效率、特定知识库场景下的应用）
5. 课程：斯坦福ML、斯坦福cs224n
6. 周一（1/16）上午开会，word汇报【标题+摘要，标题注明是否有用，描述论文的问题+解决方案】
7. 主要关键词：神经网络、文本处理、问答系统