网络出版时间：2016-09-20 21:04:43
网络出版地址：http://www.cnki.net/kcms/detail/11.1826.TP.20160920.2104.006.html

深度文本匹配综述

庞亮 [1),2)3)] 兰艳艳 [1)2)] 徐君 [1)2)] 郭嘉丰 [1)2)] 万圣贤 [1),2)3)] 程学旗 [1)2)]

[1)](中国科学院网络数据科学与技术重点实验室 北京 100190)

[2)](中国科学院计算技术研究所,北京 100190)

[3)](中国科学院大学,北京 100190)

摘 要 自然语言理解的许多任务，例如信息检索、自动问答、机器翻译、对话系统、复述问题等等，都可以抽象成文本匹配问题。过去研究文本匹配主要集中在人工定义特征之上的关系学习，模型的效果很依赖特征的设计。最近深度学习自动从原始数据学习特征的思想也影响着文本匹配领域，大量基于深度学习的文本匹配方法被提出，我们称这类模型为深度文本匹配模型。相比于传统方法，深度文本匹配模型能够从大量的样本中自动提取出词语之间的关系，并能结合短语匹配中的结构信息和文本匹配的层次化特性，更精细地描述文本匹配问题。根据特征提取的不同结构，深度文本匹配模型可以分为三类：基于单语义文档表达的深度学习模型、基于多语义文档表达的深度学习模型和直接建模匹配模式的深度学习模型。从文本交互的角度，这三类模型具有递进的关系，并且对于不同的应用，具有各自性能上的优缺点。本文在复述问题、自动问答和信息检索三个任务上的经典数据集上对深度文本匹配模型进行了实验，比较并详细分析了各类模型的优缺点。最后本文对深度文本模型未来发展的若干问题进行了讨论和分析。

关键词 文本匹配；深度学习；自然语言处理；卷积神经网络；循环神经网络

中图法分类号 TP18

论文引用格式：

庞亮，兰艳艳，徐君，郭嘉丰，万圣贤，程学旗，深度文本匹配综述，2016，Vol.39,在线出版号 No. 128

Pang Liang，Lan Yanyan，Xu Jun，Guo Jiafeng，Wan Shengxian ，Cheng Xueqi，A Survey on Deep Text Matching，2016，Vol.39,Online Publishing No.128

A Survey on Deep Text Matching

Pang Liang [1),2)3)] Lan Yanyan [1)2)] Xu Jun [1)2)] Guo Jiafeng [1)2)] Wan Shengxian [1),2)3)] Cheng Xueqi [1)2)]

[1)](CAS Key Lab of Network Data Science and Technology, Beijing100190)

[2)](Institute of Computing Technology, Chinese Academy of Sciences, Beijing100190)

[3)](University of Chinese Academy of Sciences, Beijing 100190)

Abstract Many problems in natural language processing, such as information retrieval, question answering, machine translation, dialog system, paraphrase identification and so on, can be treated as a problem of text

本课题得到国家重点基础研究发展计划(973)(No. 2014CB340401, 2013CB329606)、国家自然科学基金重点项目(No.61232010, 61472401, 61425016, 61203298)、中国科学院青年创新促进会(No. 20144310，2016102)资助.庞亮(**通讯作者**)，男，1990年生，博士，学生，计算机学会(CCF)学生会员（59709G），主要研究领域为深度学习与文本挖掘.E-mail: pangliang@software.ict.ac.cn.兰艳艳，女，1982年生，博士，副研究员，计算机学会(CCF)会员（28478M），主要研究领域为统计机器学习、排序学习和信息检索.E-mail: lanyanyan@ict.ac.cn.徐君，男，1979年生，博士，研究员，计算机学会(CCF)会员，主要研究领域为信息检索与数据挖掘.E-mail: junxu@ict.ac.cn.郭嘉丰，男，1980年生，博士，副研究员，计算机学会(CCF)会员，主要研究领域为信息检索与数据挖掘.E-mail: guojiafeng@ict.ac.cn.万圣贤，男，1989年生，博士，学生，主要研究领域为深度学习与文本挖掘.E-mail: wanshengxian@software.ict.ac.cn.程学旗，男，1971年生，博士，研究员，计算机学会(CCF)会员，主要研究领域为网络科学、互联网搜索与挖掘和信息安全等.E-mail: cxq@ict.ac.cn.

matching. The past researches on text matching focused on defining artificial features and learning relation between two text features, thus the performance of the text matching model heavily relies on the features designing. Recently, affecting by the idea of automatically feature extraction in deep learning, many text matching models based on deep learning, namely Deep Text Matching model, have been proposed. Comparing to the traditional methods, Deep Text Matching models can automatically learn relations among words from big data and make use of the information from phrase patterns and text hierarchical structures. Considering the different structures of Deep Text Matching models, we divide them into three categories: Single semantic document representation based deep matching model, Multiple semantic document representation based deep matching model and Matching pattern based deep matching model. We can see the progressive relationship among three kinds of models in modelling the interaction of texts, while which have their own merits and defects based on a specific task. Experiments conduct on the typical datasets of paraphrase identification, question answering and information retrieval. We compare and explain the different performance of three kinds of deep text matching models. Finally, we give the key challenge and the future outlook of the deep text matching models.

Key words Text Matching; Deep Learning; Natural Language Processing; Convolutional Neural Network; Recurrent Neural Network

1 引言

文本匹配是自然语言理解中的一个核心问题。对文本匹配的研究可以应用到大量已知的自然语言处理任务中，例如信息检索[1]、自动问答[2]、机器翻译[3]、对话系统[4]、复述问题[5]等等。这些自然语言处理的任务都可以在一定程度上抽象成文本匹配问题，比如信息检索可以归结为查询项和文档的匹配，自动回答可以归结为问题和候选答案的匹配，机器翻译可以归结为两种语言间的匹配，对话系统可以归结为前一句对话和回复的匹配，复述问题则可以归结为两个同义词句的匹配。这些匹配需要关注的特性具有很大不同，如何利用一个较好的文本匹配模型，针对不同任务找到最适合的匹配方式，成为研究文本匹配这个核心问题最大的挑战。

文本匹配面临的挑战主要来源于以下几个方面：

（1）**词语匹配的多元性**

不同的词语可能表示的是同一个语义，比如同义词，“荷花”、“莲花”、“水芙蓉”、“芙蕖”，它们表示的都是同一种植物；同理一个相同的词在不同的语境下会有不同的语义，比如“苹果”既可以是一种水果，也可以是一家公司，亦可以是一个品牌。

（2）**短语匹配的结构性**

多个词语可以按照一定的结构组合成短语，匹配两个短语需要考虑短语的结构信息。比如“机器学习”和“机器学习”是两个词顺序匹配的，而“机器学习”和“学习机器”只有词语是匹配的，而顺序是打乱的。这两种情况的匹配程度是不一样的。

（3）**文本匹配的层次性**

文本是以层次化的方式组织起来的，词语组成短语，短语组成句子，句子组成段落，段落组成篇章。这样一种特性使得我们在做文本匹配的时候需要考虑不同层次的匹配信息，按照层次的方式组织我们的文本匹配信息。

最近文本匹配问题的研究，渐渐从传统文本匹配模型向深度文本匹配模型转移。由于传统的文本匹配模型需要基于大量的人工定义和抽取的特征[6-8]，而且可以学习调整的参数相对较少，所以这些特征总是根据特定的任务（信息检索，或者自动问答）人工设计的，很大程度上限制了模型的泛化能力。在信息检索方面很多工作是基于传统的检索模型的改进，例如将文档模型和 LDA 模型融合[9]，将词向量的结果作为特征[10]；自动问答的大量工作是基于知识库检索[11-13]，也既是对结构化的数据的检索问题；而对话系统的研究还处于起步阶段，部分利用传统模型的工作是基于句型模式识别和语义提取[14, 15]。传统模型在一个任务上表现很好的特征很难用到其他文本匹配任务上。而利用深度学习方法[16]，可以自动从原始数据中抽取特征，免去了大量人工设计特征的开销。首先特征的抽取过程是模型的一部分，根据训练数据的不同，可以方便适配

到各种文本匹配的任务当中。以此同时，深度文本匹配模型结合上词向量（Word2Vec[17, 18]）的技术，更好地解决了词语匹配的多元性问题。最后得益于神经网络的层次化设计原理，深度文本匹配模型也能较好地符合短语匹配的结构性和文本匹配的层次性的特性。

根据特征提取的不同方式，结合近期大量的相关工作，本文将深度文本匹配模型划分成三大类：基于单语义文档表达的深度学习模型、基于多语义文档表达的深度学习模型和直接建模匹配模式的深度学习模型。基于单语义文档表达的深度学习模型主要思路是，首先将单个文本先表达成一个稠密向量（分布式表达），然后直接计算两个向量间的相似度作为文本间的匹配度；基于多语义的文档表达的深度学习模型认为单一粒度的向量来表示一段文本不够精细，需要多语义的建立表达，更早地让

表 1 深度文本匹配模型分类

	基于单语义文档表达	基于多语义文档表达	直接建模匹配模式
全连接网络	DSSM[19]	---	DeepMatch$_{Tree}$[20]
卷积网络	CDSSM[21], ARC-I[22], CNTN[23]	MultiGranCNN[24]	DeepMatch[25], ARC-II[22], MatchPyramid[26]
递归网络	---	uRAE[27]	---
循环网络	LSTM-RNN[28]	MV-LSTM[29]	Match-SRNN[30]

两段文本进行交互，然后挖掘文本交互后的模式特征，综合得到文本间的匹配度。表格 1 展示了深度文本匹配模型的分类。也就是分别提取词、短语、句子等不同级别的表达向量，再计算不同粒度向量间的相似度作为文本间的匹配度；而直接建模匹配模式的深度学习模型则认为匹配问题需要更精细的建模匹配的模式，也就是需要

本文的内容组织结构如下：第二章简单介绍文本匹配问题，得到一个抽象简洁的表述，并回顾了传统文本匹配模型的发展。第三章先形式化了现有深度学习框架的数学形式，然后分三个类别详细讲解近期深度文本匹配模型的发展和他们之间的联系。第四章分别在复述问题、自动问答和信息检索三个任务对各类模型的表现做了分析。最后我们对深度文本匹配模型做一个总结和未来发展的展望。

2 文本匹配问题简介

2.1 问题描述

本节将文本匹配问题利用数学符号进行形式化，这样我们可以得到一个更为抽象简洁的问题描述。首先给定标注训练数据集合$\mathbb{S}_{train} = \{(s_1^{(i)}, s_2^{(i)}, r^{(i)})\}_{i=1}^{N}$，其中$s_1^{(i)} \in S_1$，$s_2^{(i)} \in S_2$为两段文本（例如在搜索引擎中，两者分别为查询项和文档；而在问答系统中，两者分别为问题和答案）；$r^{(i)} \in R$表示对象x_i和y_i的匹配程度（如在搜索引擎和问答系统中代表相关程度）。文本匹配的目标是在训练数据上，自动学习匹配模型$f: S_1 \times S_2 \mapsto R$，使得对于测试数据$\mathbb{S}_{test}$上的任意输入$s_1 \in S_1, s_2 \in S_2$，能够预测出$s_1$和$s_2$的匹配度$r$，然后通过匹配度排序得到结果。

例如这样一个复述问题：

s_1:*从古至今，面条和饺子是中国人喜欢的食物。*

s_2: *从古至今，饺子和面条在中国都是人见人爱。*

那么判断s_1和s_2是否匹配就是我们的结论r。在复述问题中，是否匹配就是看两个句子是否表达同一个意思；在问答系统中，是否匹配就是看第二个句子是否是第一个句子的答案。因此针对不同的任务匹配的定义是不同的。

上面这个复述问题的例子，我们发现有许多关键词可以直接匹配上，例如（*从古至今-从古至今*）、（*面条-面条*）、（*饺子-饺子*）等。还有些词是语义相似的，例如（*中国人-中国*）、（*喜欢-人见人爱*），这体现了词语本身的多元性。而在短语级别，我们发现（*面条和饺子-饺子和面条*）这两个短语是一个意思，由相同的词组成，但是组织结构上有所不同，这里反映了短语的结构对匹配的影响。而上升到整段文本来看，我们首先匹配词语，进而匹配不同短语的结构，然后再在更长的短语级别来归纳匹配的信息，这样就能层次化得到两段文本的匹配度。

对于一个实际任务，我们通常会抽象成一个排序问题。给定一段文本s_1，和另一段文本的一个列

表$\{s_2^{(i)}\}$，目标是在这个列表中筛选出与给定文本s_1匹配的文本。文本匹配模型会计算所有的文本对$(s_1, s_2^{(i)})$的匹配度$r^{(i)}$，$\{r^{(i)}\}$列表排序靠前的文本和s_1的匹配度越高。

2.2 评价指标

衡量一个排序结果优劣的评价指标[31]主要包括：P@k（Precision at k），R@k（Recall at k），MAP（Mean average precision），MRR（Mean reciprocal rank）以及 nDCG（normalized Discounted cumulative gain）。

定义真实排序前 k 个文本中，匹配文本的数量为G_k，而在预测排序中前 k 个文本中，匹配文本的数量为Y_k。评价指标 P@k 和 R@k 的定义如下：

$$P@k = \frac{Y_k}{k} R@k = \frac{Y_k}{G_k}$$

假设预测排序中的真实匹配的文本的排序位置分别为$k_1, k_2, \ldots, k_r$，其中 r 为整个列表中所有匹配文本的数量。那么指标 MAP 的定义如下：

$$MAP = \frac{\sum_{i=1}^{r} P@k_i}{r}$$

依据同样的思路，如果我们只考虑排名最靠前的真实匹配的文本k_1，就可以导出指标 MRR 的定义：

$$MRR = P@k_1$$

以上的评价指标对于匹配的度量都是二值化的，也就是说，只有匹配上（匹配值为 1），或者不匹配（匹配值为 0）。而对于有些问题匹配是有等级的。我们进一步引入匹配程度，例如匹配程度分为 0、1、2 三个等级。给定最优排序，按照顺序每个位置的文档对应的匹配度为$\hat{rel}_1, \hat{rel}_2, \ldots, \hat{rel}_n$。给定预测的排序，按照顺序每个位置的文档对应的匹配度为$rel_1, rel_2, \ldots, rel_n$。基于这个设定，nDCG 评价指标的定义如下：

$$IDCG = \hat{rel}_1 + \sum_{i=2}^{n} \frac{\hat{rel}_i}{\log_2 i}$$

$$DCG = rel_1 + \sum_{i=2}^{n} \frac{rel_i}{\log_2 i}$$

$$nDCG = \frac{DCG}{IDCG}$$

2.3 传统文本匹配学习模型

传统的文本匹配研究主要基于人工提取的特征，因此问题的焦点在于如何设置合适的文本匹配学习算法来学习到最优的匹配模型。以互联网搜索为例，查询项与网页被认为是两个异质空间中的对象，多种匹配学习模型被提出来去计算查询与网页的相关度。Berger 和 Lafferty [32]提出使用统计机器翻译模型计算网页词和查询词间的“翻译”概率，从而实现了同义或者近义词之间的匹配映射；Gao 等人[33]在词组一级训练统计机器翻译模型并利用用户点击数据进行模型训练，获得了很好的效果。进一步地说，典型相关分析 (CCA, canonical correlation analysis) [34]和偏最小二乘 (PLS, partial least square) [35]等隐空间模型试图为两种对象建立一个公共的隐空间，任意给定的查询和文档都可以被映射到此隐空间中，且在隐空间中查询和文档有一致的表达方式和特征维度，从而可以方便地计算两者的相似度或者距离，进而对其是否具有相同的“语义”做出判断。例如，Wu 等人[36]提出正则化隐空间映射(regularized mapping to latent space, RMLS)把查询项和网页映射到同一隐空间中，并在模型训练中引入了正则化因子以避免奇异解，Bai 等人[37]提出有监督学习语义索引模型(supervised semantic indexing, SSI)，Gao 等人[38]扩展了话题模型提出双语话题模型(bilingual topic model, BLTM)，对隐空间模型进行概率化建模。

尽管这些模型已经在诸如网络搜索，推荐和问答等应用中取得了良好的效果，然而还是存在许多问题。（1）人工提取特征的代价很大。需要花费大量人力物力才能提取到少量的比较有效的特征，这其中不仅需要有经验的工程师来设计，还需要大规模的特征选择过程。（2）基于主题模型的隐空间模型还比较粗糙，难以克服文本匹配中的语义鸿沟问题。（3）传统模型很难发掘一些隐含在大量数据中，含义不明显的特征，然而往往有些特殊情况需要这样的特征才能提高性能。

3 基于深度学习的文本匹配学习模型

近年来，随着深度学习在计算机视觉[39, 40]，语音识别[41]等领域取得的突破性进展，自然语言处理成为深度学习研究的下一个应用热点。深度学习用于自然语言处理的优势主要体现在：(1) 深度学习模型可以将单词表示为语义空间中的向量，利用向量之间的运算可以更准确地描述两个单词之间的语义关系；(2) 深度学习模型自身的结构是层次化和序列化的，能够比较自然地描述自然语言中的层次结构、序列结构和组合操作；(3) 深度学习模型

很好地利用大规模数据的优势和日益发展的高性能计算的能力，将神经网络的灵活结构，匹配上复杂的自然语言的知识表示。直接从大量数据学习既可以模拟人们定义规则（特征）来描述规范的一般的语言规律，又可以刻画例外的、特殊的语言现象，从而大幅提高语言处理的精度。

在传统的自然语言处理领域深度学习已经有了很多突破性的进展，如词性标注[42]、语法分析[43]、情感分析[44]、关系分类[45]等。相关的工具包括卷积神经网络[46]和循环神经网络[47]等。卷积神经网络的卷积核的结构能够建模局部化信息，并有平移不变性的特性[48]，堆叠起来的卷积层可以很方便地模拟语言层次化的特性。而循环神经网络更偏向于序列化建模，类似人类阅读文本的方式每次将历史的信息压缩到一个向量，并作用于后面的计算，符合建模文本的序列性。

3.1 深度文本匹配形式化

考虑到深度学习技术强大的特征表示学习能力，不少研究人员纷纷把深度学习用于完成复杂的文本匹配任务，提出了一些基于深度学习的文本匹配学习算法[49]。我们将深度学习模型分为了基于单语义文档表达的深度学习模型、基于多语义文档表达的深度学习模型以及直接建模匹配模式的深度学习模型。已有的模型可以按照表格 1 进行划分。

针对深度文本匹配模型，需要关注如下的几个关键点。首先我们定义$s_1 = \{x_i\}_{i=1}^{n}, s_2 = \{y_i\}_{i=1}^{m}$表示文本样本$s_1$和$s_2$中的单词序列，其中$n$和$m$表示句子长度，$x_i$，$y_i$表示句子中的单词。

（1）**单词表达**：函数$w_i = \phi(x_i)$，$v_i = \phi(y_i)$表示单词x_i，y_i到词向量w_i，v_i的一个映射。整个句子映射后得到矩阵$\boldsymbol{w}$和$\boldsymbol{v}$。

（2）**短语/句子表达**：利用函数$\boldsymbol{p} = \Phi(\boldsymbol{w})$，$\boldsymbol{q} = \Phi(\boldsymbol{v})$，得到短语或者整个句子的表达。

到这一步为止，我们都是处理的单个句子的表示问题，所以最后得到的表达，或者中间的表达不仅仅可以用于文本匹配问题，还可以用于文本分类聚类问题。接下来就是文本匹配的特殊步骤了。

（3）**文本交互**：用$\mathbf{M_0}$表示两段文本交互后的结果，我们定义$\mathbf{M_0} = f(\boldsymbol{p}, \boldsymbol{q})$。

（4）**匹配空间内的模式提取**：在得到基本交互信息的基础上进一步提取匹配空间的模式信息，可以表示为函数$\mathbf{M}_k = g(\mathbf{M}_{k-1})$。这里的函数$g$可以由多个函数级联而成。

（5）**匹配程度得分**：最后一步就是综合前面的信息，得到一个匹配程度的打分，也即$r = h(\mathbf{M}_n)$。

3.2 基于单语义文档表达的深度学习模型

将文档表达成一个向量，这个向量就称为文档的表达，广义地来说，传统方法得到的只基于一个文档的特征就可以看做一个文档的表达。例如文档中每个词的词频，文档的长度等等。而这里的文档表达则是利用深度学习的方法来生成的。考虑到深度学习的优势在于特征表示能力，因此这些算法首先将待匹配的两个对象通过深度学习表达成两个向量。目前利用深度学习的方式生成句子表达的方法有很多，例如基于扩展 Word2Vec 的方法[50]，卷积神经网络的方法[51, 52]，循环神经网络的方法[53, 54]以及树状的递归神经网络的方法[55, 56]。

得到两个句子的表达之后，通过计算这两个向量之间的相似度度量便可输出两者的匹配度。这个框架可以统一表达为图 1 所示的流程。

这类工作的灵感主要来自于 Siamese 框架[57]，利用同质的网络得到两个对象的表达，然后通过表达的相似度来衡量两个对象的匹配度。

有代表性的几个基于单语义文档表达的深度学习模型，主要是在构建单个文档表达和如何计算表达之间的相似度上面进行了研究。由于这类方法的核心是构建单个文档表达的方法的差异，我们将他们分成了基于全连接神经网络、卷积神经网络和循环神经网络这样的三类构造文档表达的方法。

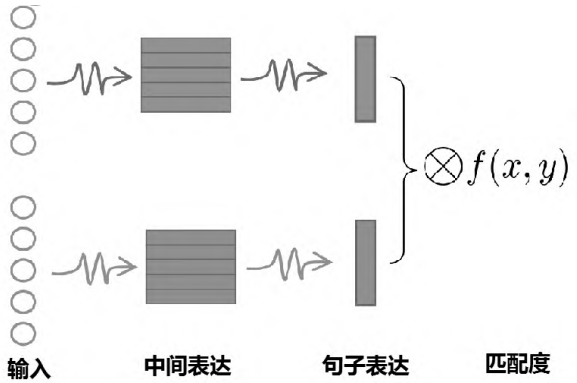

图 1 基于单语义文档表达的深度学习模型统一框架

3.2.1 基于全连接神经网络

深度语义结构模型（Deep Semantic Structured Model，DSSM）[19]是最早将深度模型应用在文本匹配的工作之一，该模型主要针对查询项和文档的匹配度进行建模，相对于传统文本匹配的模型，该方法有显著的提升。深度语义结构模型是个典型的 Siamese 网络结构，每个文本对象都是由五层的网络单独进行向量化的，最后计算两个文本向量的余

弦相似度来决定这两段文本的相似程度。

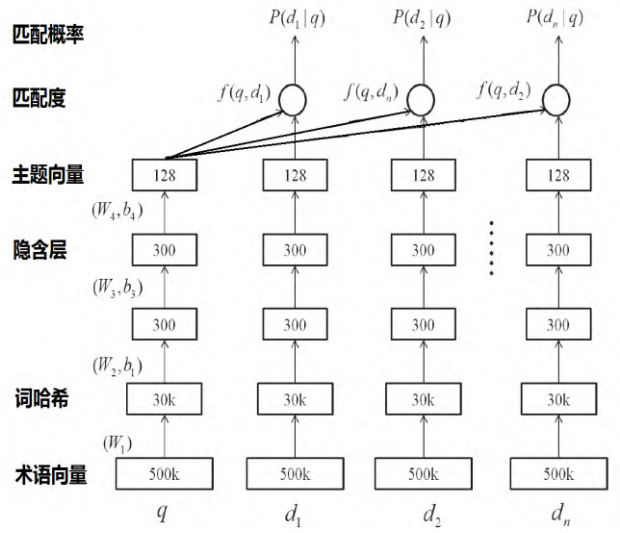

图 2 深度语义结构模型

深度语义结构模型在文本向量化的时候分成了两个主要部分，第一部分是将文本中的每个单词，或者是由三个字母组成的片段做成一个最小单元，通过哈希方式映射到一个单词（字母片段）级别的向量，见图 2 中的词哈希层。基于得到的单词向量，深度语义结构模型接上了三层的全连接来表达整个句子的主题向量，这个向量有 128 个维度。整个模型在训练的时候采用了搜索系统的点击日志，点击的作为正样本，并从没有点击的里面随机抽样一定量的负样本。然后正负样本组成一组，通过 Softmax 函数计算每个文档和查询项的匹配概率（加和为 1），然后最大化所有正例的匹配概率的似然函数（公式 1）。

$$P(d|q)=\frac{\exp(\gamma f(q,d))}{\sum_{d'\in\mathbf{D}}\exp(\gamma f(q,d'))},$$
$$L(\Lambda)=-\log\prod_{(q,d^+)}P(d^+|q). \quad (1)$$

其中γ是 Softmax 函数的平滑参数，$f(q,d)$表示一个查询项q与文档d之间的匹配度。$\mathbf{D}$表示所有文档的集合，在实际应用中我们一般采样若干正例d^+，以及采样若干负例d^-，来取代整个集合$\mathbf{D}$。

3.2.2 基于卷积神经网络

微软的研究团队在成功提出深度语义结构模型之后，发现全连接的神经网络的参数太多，不利于优化，而且构造输入数据利用的是词袋模式，忽略了词与词之间序的关系，对于匹配这种局部信息很强的任务，没法将一些学到的局部匹配信息应用到全局。于是进一步改进模型，从而提出了基于单词序列的卷积深度语义结构模型（Convolutional Deep Semantic, CDSSM）[21, 58]。卷积深度语义结构模型相对于深度语义结构模型，将中间生成句子向量的全连接层换成了卷积神经网络的卷积层（Convolutional Layer）和池化层（Pooling Layer），如图 3，其他的结构和深度语义结构模型是一样的。

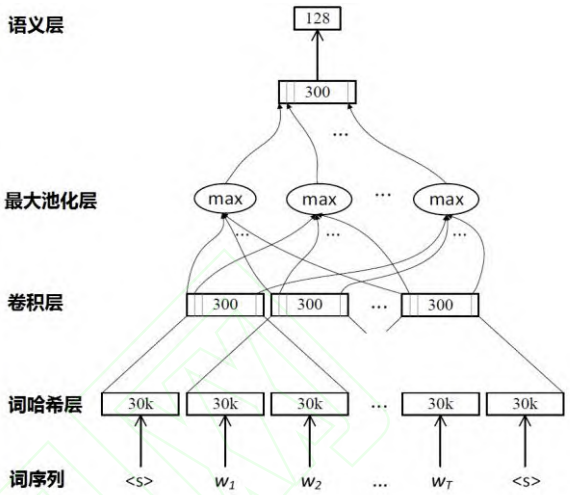

图 3 卷积深度语义结构模型

卷积深度语义结构模型首先将查询项与文档中的每个单词（字母片段）都表示为一个词向量，然后对每个固定长度的窗口内的词向量进行卷积操作，得到针对这个窗口内短语的一个向量表达，之后卷积深度语义结构模型在这些卷积得到的向量上进行全局的池化操作，即对所有窗口输出的向量的相同位置取最大值。由于卷积的滑动窗口的结构形式考虑到了句子中的单词顺序信息,在相关度判断的准确度相对于深度语义结构模型方面有了一定的提升。

此后，在其他文本匹配的任务方面也提出了思想类似的模型。在对话方面，华为诺亚方舟实验室的李航等人参考在文本分类任务中，Kim 等人提出的卷积神经网络来建模句子表达[52]的思想，在[22]中提出的 ARC-I 模型也使用了卷积神经网络的结构来进行文本匹配。ARC-I 直接将两个待匹配的句子表达为两个定长的向量，然后拼接两个向量并输入一个全连接的多层神经网络，从神经网络的输出得到最终的匹配值（图 4）。复旦大学邱锡鹏等人同样使用了卷积神经网络的结构来进行文本的表达，然后提出使用张量神经网络（Neural Tensor Network）[23]作为相似度度量来建模两个文本向量之间的关系，提出了 CNTN 模型（图 5），从而能够刻画更加复杂的匹配关系，在社区问答（Community Question Answering）等应用场景中取

得了良好的效果。

这里的张量神经网络是由 Socher 等人在[59]中提出的，具体的形式为：

$$s = f(\phi(x)^{\top} T[1:c]\phi(y) + W\begin{bmatrix}\phi(x)\\ \phi(y)\end{bmatrix} + b). \qquad (2)$$

其中x和y分别表示两个单词，$\phi(x)$、$\phi(y)$表示两个单词的词向量。$T[1:c]$是由c个矩阵组成的张量，表示两个对象之间的二维交互，而W则表示基于词向量的一维特征，b是偏移量。张量神经网络有很大的自由度，可以学习并模拟很多种相似度的计算公式，例如余弦相似度、点积或者双线性（bilinear）相似度。

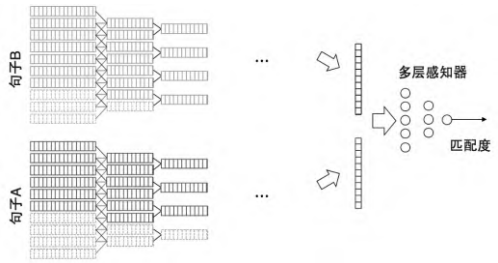

图 4ARC-I 模型

ARC-I 和 CNTN 这两个模型在训练的时候也和卷积深度语义结构模型有所不同，他们使用了基于排序的损失函数，旨在拉大正负样本之间的匹配度数值的差距，而并不在意匹配度的绝对值的大小，

这个损失函数更接近排序的应用场景。

$$L(\Lambda) = \max(0, 1 + f(q, d^{-}) - f(q, d^{+})). \qquad (3)$$

然而这些基于卷积神经网络的深层匹配结构只考虑了滑动窗口内单词的顺序，所以无法表达句子中远距离的依存关系和复杂语义。

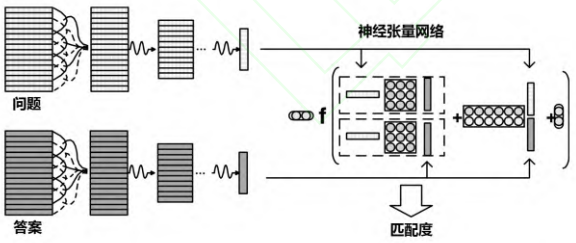

图 5CNTN 模型

3.2.3 基于循环神经网络

为解决基于卷积神经网络没法捕捉句子长距离的依存关系的问题，微软的研究团队 Palangi 等人在[28]中提出基于长短时记忆（Long Short Term Memory）[60]的文本匹配模型（LSTM-RNN）。具体地来说，查询项和文档分别经由长短时记忆的循环神经网络（Recurrent Neural Network）表达为一个向量，然后计算两个向量表达的余弦距离作为相似度的度量，输出最终的匹配值。该方法在网页搜索的在线日志数据上得到了目前深度模型所能取得的最好结果。

长短时记忆的循环神经网络由于门机制的作用，可以在顺序扫描句子的时候，将长距离的依赖关系保存到存储单元中，并通过门的方式控制读写。这样的设计用公式表示为：

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i),\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f),\\
u_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c),\\
c_t &= f_t \odot c_{t-1} + i_t \odot u_t,\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o),\\
h_t &= o_t \odot \tanh(c_t).
\end{aligned} \qquad (4)$$

其中i, f, o分别表示输入门、遗忘门和输出门，c表示一个存储单元，h表示每个时刻的输出表达，u表示当前输入x_t和上一时刻输出h_{t-1}组合的表达。这些方程中的σ表示 sigmoid 函数，而$\odot$表示矩阵元素乘法（Hadamard product）。

Palangi 在文章中用实验的方式展示了长短时记忆网络能够保持并利用长距离信息的特性，其中门的机制可以抽取出匹配文本中的关键词。这篇文章的其他部分和基于单词序列的卷积深度语义结构模型是类似的。

3.2.4 小结

基于单语义文档表达的深度学习算法本质上是把重心放在求句子表达上，就是我们 3.1 提出的$\boldsymbol{p}$和$\boldsymbol{q}$。然后文本交互函数f定义的比较直接，比如余弦相似度，或者更复杂点的神经张量网络。本章的许多方法的提出，也影响了后面的许多工作。相似度的计算方式和损失函数的定义都是后面方法经常借鉴的对象。

以上基于单语义文档表达的深度学习算法具有三个优点：（1）将文本映射为一个简洁的表达，便于储存；（2）匹配的计算速度快，可以和一些加速方法如位置敏感哈希（Locality Sensitive Hashing，LSH）技术[61]结合，进一步提高计算速度；（3）模型可以用大量无监督的数据进行预训练，尤其是在匹配监督数据很少的时候，用大量文本进行预训练是相当有效的方法。因此，该模型非常适合于信息检索这种对存储和速度要求都比较高的任务。

然而，该方法也存在很大的缺点：首先，很多匹配问题不具备传递性（例如问答系统中的匹配），因此不适合用一个度量空间来描述；其次，文本的表示学习本身是非常困难的问题，需要有效捕捉与

描述对匹配有用的局部化（细节）信息。

3.3基于多语义文档表达的深度学习模型

针对基于单语义文档表达的深度学习模型存在的缺点，一些新的深度匹配模型被提出来去综合考虑文本的局部性表达（词，短语等）和全局性表达（句子）。这类模型不仅会考虑两段文本最终的表达向量的相似程度，也会生成局部的短语或者更长的短语的表达进行匹配。这样多粒度的匹配可以很好地补充基于单语义文档表达的深度学习模型在压缩整个句子过程中的信息损失，而达到更好的效果。

3.3.1 可伸展递归自动编码器

美国斯坦福大学 Socher 等人[27]提出了一种有递归特性的神经网络，称之为可伸展递归自动编码器（Unfolding Recursive Auto-Encoder, uRAE），见图 6。该算法首先利用现有的工具对两段文本进行句法分析，并自动构建句法树[62]。得到的句法树作为递归自动编码器（Recursive Autoencoder）树状连接的结构，并利用大量的无监督样本进行与训练得

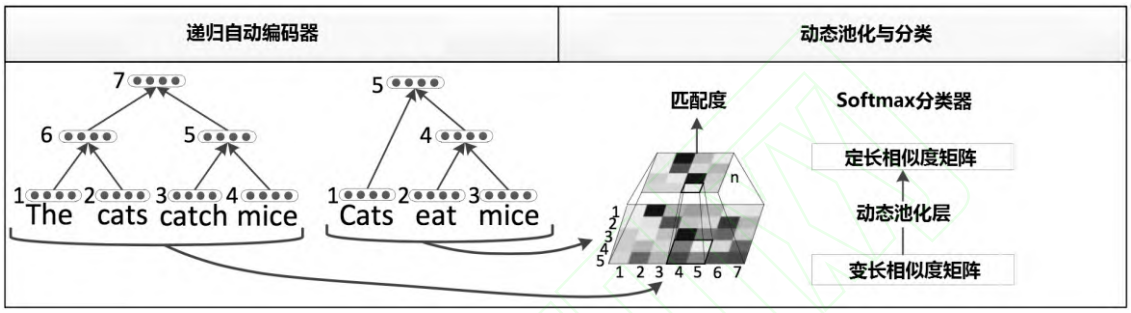

图 6 可伸展递归自动编码器模型

到一个能够编码句子，短语，词的模型。进而把文本中不同级别的片段（词、短语以及句子）变换为向量。由于使用递归自动编码器，这些向量处在同一个语义空间中，所有片段之间的匹配度可以直接计算得到，从而按照语法树的遍历得到两段文本之间的匹配相似度矩阵。该矩阵反映了两个句子的匹配关系，其中的每个元素表示两个句子的一个片段对在语义空间里的欧式距离。

考虑到不同句子可以分成的片段数量并不一致，所以得到的匹配矩阵的大小是动态变化的。利用动态池化技术（dynamic pooling）可以将变长的相似度矩阵变换为定长的相似度矩阵，之后通过一个简单的神经网络模型计算出两个句子的最终匹配值。动态池化技术简单来讲，首先需要确定输出大小，然后反向从对应的位置找最近的输出赋值，即相当于针对不同样本使用变尺度的池化核进行池化操作（公式 5）。

$$\mathbf{z}_{i,j}^{(2,k)} = \max_{0\le s<d_k} \max_{0\le t<d_k'} \mathbf{z}_{i\cdot d_k+s,\, j\cdot d_k'+t}^{(1,k)}, \tag{5}$$

其中d_k和d_k'表示当前池化核的大小，他们是由上一层的特征矩阵大小n和m决定的，计算方法是$d_k = \lceil n/n' \rceil, d_k' = \lceil m/m' \rceil$，最后输出的是$n' \times m'$固定大小的特征矩阵。

3.3.2 多粒度卷积神经网络

Yin 等人在[24, 63]中提出使用卷积神经网络来分别得到词，短语和句子等几个不同层面的文本表达，然后将这些向量拼接到一起或者建模这些向量之间的相似度来得到最终的匹配值。如图 7 所示，多粒度卷积神经网络（MultiGranCNN）将一个句子拆解成四个层次，单词级别、短语级别、长短语级别和句子级别，之后将两个句子不同级别的特征进行两两的相似度计算，得到一个相似度矩阵，进行动态最大值池化（3.3.1）得到的就是两个句子的相似度得分。目前这类工作通过加入丰富的细节信息，从词、短语和句子三个层次来表示句子，使得句子表示更加丰富，可以描述更加抽象的内容，在复述问题等任务上取得了很好的效果。

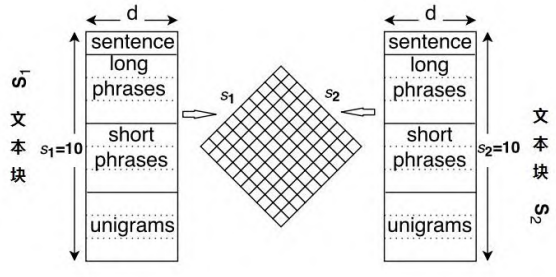

图 7 多粒度卷积神经网络

实验表明长短语级别和短语级别的特征对改

述任务十分关键，而整个句子和仅仅单词的特征相比较就要弱一些，将所有特征信息拼起来可以得到最好的实验结果。仿照可伸展递归自动编码器模型，多粒度卷积神经网络，也进行了大量无监督样本的预训练，在 Yin 等人的文章[63]中指出，预训练起到了很重要的作用。

3.3.3 多视角循环神经网络

我们在分析循环神经网络的时候发现，循环神经网络在扫描一个句子的过程中会在不同位置分别输出一个表达，这个表达表示的是从句子开始到当前位置内容的一个整合。基于这样的发现我们提出了多视角循环神经网络（MV-LSTM）[29]。

多视角循环神经网络基于长短时记忆网络（LSTM）设计，由于长短时记忆网络的特殊的基于门的神经元设计，使得它能够同时捕获长距离和短距离的依赖。这种处理数据的方式，使得长短时记忆网络存在位置偏见，会倾向于离当前位置之前较近的单词[64]。为了得到对整个句子在当前单词位置的表达，多视角循环神经网络从两个方向同时扫描，也即用双向的循环神经网络[65]，在同一个位置会得到两个表达，分别是从句子开始扫描过来得到的表达和从句子结尾扫描过来得到的表达（图 8 左）。然后将一个位置的两个表达拼接到一起作为当前位置为中心的整个句子的表达，这样每个句子都可以看做是由不同中心词产生的多个视角表达的集合。然后将两个句子不同视角的句子表达两两计算相似度，得到一个相似度矩阵，通过动态最大值池化操作加上全连接网络就能得到最后的相似度得分。

双向循环神经网络得到的每个位置的表达也可以类似看做一个变长的窗口[54]的卷积操作，在不同的位置可以有不同粒度的窗口，也就是说双向循环神经网络可以生成多粒度的表达，这个和多粒度卷积神经网络的功能是类似的，而且更加灵活的是窗口的大小是一个软的边界，不像卷积核的大小是固定的，并且可根据输入的数据不同灵活地变化。

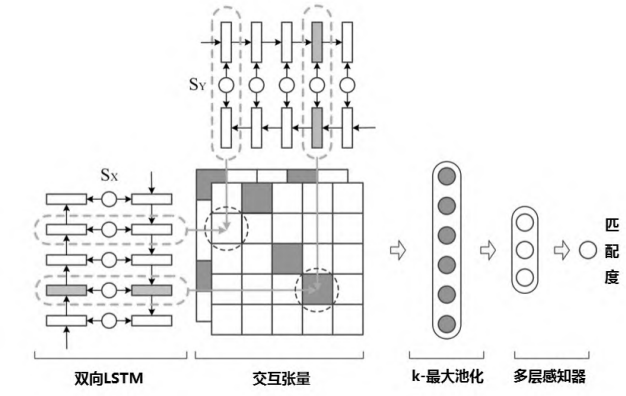

图 8 多视角循环神经网络模型

3.3.4 小结

基于多语义的文档表达的深度学习模型和基于单语义文档表达的深度学习模型类似，都是将两个对象分开进行表达，最后在计算两个表达的相似度。不同的是多语义的文档表达会考虑不同粒度的表达，不仅仅是句子级别的，还有短语和单词级别的表达。也就是将 3.1 里面的$\boldsymbol{p}$和$\boldsymbol{q}$表达成多个粒度的向量。细粒度的表达带来了更丰富的信息，所以能够比基于句子级别的表达的深度模型得到更好的效果。而且这两种单个对象进行表达的模型都可以进行大量无监督的预训练，例如在可伸展循环自动编码器模型和多粒度卷积神经网络模型。这样在数据量不够的情况下，我们依旧能够得到较为不错的效果。

但基于多语义文档表达的深度学习模型还存在如下缺点：1）可伸展循环自动编码器模型依赖于一个给定的句法树，而语法树算法自身准确性不高，因此算法鲁棒性不足。2）无法区分不同上下文中局部化信息的重要性，在语言多义的挑战下，很难将局部化信息与全局化信息进行有效地整合利用。3）匹配不仅仅是一元的一一对应，而且是有层次、有结构的，分别从两个对象单独提取特征，很难捕获匹配中的结构信息。

3.4直接建模匹配模式的深度学习模型

与集中于文本表达（局部化或者全句化）的思路不同，直接建模匹配模式的深度学习模型旨在直接捕获匹配的特征：匹配的程度和匹配的结构。这样更接近匹配问题的本质，也更加契合人们面对两段文本进行匹配分析的方法。当进行两段文本的匹配时，我们会先看是不是有匹配的关键词，然后再看关键词之间的相对位置是不是匹配的，最后整合整个句子的意思给两段文本匹配的程度进行打分。

实验显示这些模型能在相对复杂的问题上表现得更优秀。

3.4.1 主题深度匹配模型

主题深度匹配模型（DeepMatch）[25]包含局部匹配层和综合层两个部分（图 9），局部匹配层包含多个局部匹配模型，用于将输入的文本对表达为多个局部匹配的结果，其中每个局部匹配模型都是一个双语主题模型，而综合层是一个多层神经网络，将得到的局部匹配的结果进一步综合得到最终匹配结果。

具体而言局部匹配层希望模拟卷积核能够提取局部块的方式来构造第一层的局部连接，经过分析，主题深度匹配模型认为局部块是主题连续的，为了达到这样的设计，模型首先根据预训练好的主题模型筛选局部块，同一个局部块的单词会统一连接到下一层的一个隐节点上。而隐含层的多层全连接网络能够在细粒度的主题上进行进一步的抽象和提取，最后得到匹配分数。

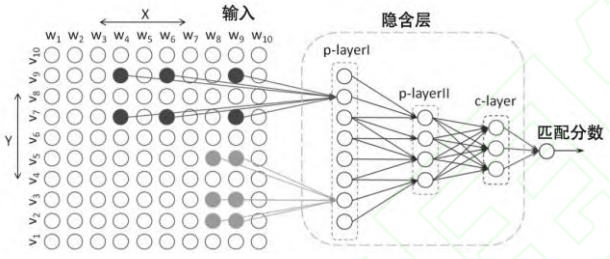

图 9 主题深度匹配模型

由于主题深度匹配模型用词袋（bag of words）来表示句子，忽略了词在句子里的顺序，它与 DSSM 一样虽然善于捕捉主题层面上的匹配，但并不适合表达相对更精细的语义。

3.4.2 树深度匹配模型

树深度匹配模型（DeepMatch$_{Tree}$）[20]采用依存树作为文本（一般为短文本）的表示，如图 10。具体地说，树深度匹配模型由局部匹配模型和综合层两部分决定，其中局部模型是千万量级的基于依存树的二值匹配模型，每一个局部匹配模型都对应一个子树对，而匹配模型的输出（0 或 1）取决于输入的句子对是否含有这两个子树对所表示的依存结构。对于给定的两个文本，树深度匹配模型首先检查其对应的依存树，并根据其是否包含表中的有效子树对来得到一个稀疏的二值表示，然后由一个深层神经网络来综合这个表示，最后得到全局的匹配值。它可以看做是主题深度匹配模型在两方面的推广：首先，从词到包含词的子树结构；其次，从词的集合到单个子树对应的精度上的提升。这些推广帮助准确地捕捉表示大量的精细匹配模型，如“价格-调控”，“政府-调节”等，使得匹配精度得到大幅度的提升。作为代价，树深度匹配模型需要大规模的图挖掘和深层神经网络作为支撑。

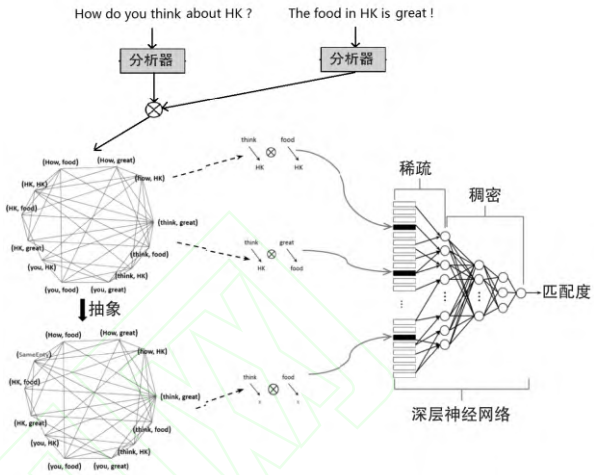

图 10 树深度匹配模型

3.4.3 卷积网络深度匹配模型

卷积网络深度匹配模型（ARC-II）[22]首先把句子表达成为句中单词的向量序列，然后用滑动窗口来选择词向量组作为基本单元进行卷积操作，得到一个三维的张量，作为两个句子相互作用的一个初步表示。随后的卷积以这个三维张量为基础进行“卷积+池化”的操作若干次，最后得到一个描述两个句子整体关联的向量，最终由一个多层神经网络来综合这个向量的每个维度从而得到匹配值（图 11）。

与主题深度匹配模型相比，卷积网络深度匹配模型考虑了句子中词的顺序，从而可以对两个句子的匹配关系进行相对完整的描述；与树深度匹配模型相比，卷积网络深度匹配模型的学习框架更加灵活，然而还缺乏对于细微匹配关系的捕捉，在精确匹配上面还存在缺陷。

3.4.4 MatchPyramid

虽然卷积深度匹配模型 ARC-II 更早地让两段文本进行了交互，但是这个交互的意义其实并不明确，层次化的过程也比较模糊。我们提出的 MatchPyramid 模型[26]重新定义了两段文本交互的方式——匹配矩阵（Matching Matrix），然后基于匹配矩阵这个二维的结构进行二维卷积提取匹配空

间的模式，最后通过全连接的网络得到两个句子之间的相似度。

MatchPyramid 模型的核心思想是层次化的构建匹配过程，与树深度匹配模型不同的是 MatchPyramid 模型不需要依赖于构建好的语法树。首先定义的匹配矩阵是基于最细粒度的两个句子中词和词之间的匹配程度。模型利用两个词的词向量之间的同或关系、余弦相似度或者点积来定义词之间的相似度，然后句子之间两两词之间都会计算相似度，根据词在句子中的空间位置刚好可以构建出一个二维的结构，我们称之为匹配矩阵。匹配矩阵包含了所有最细粒度的匹配的信息，类似一副图像，如果值的大小和像素的深浅相关，那么可以很容易做出这样一个匹配对应的图像（图 12）。

之后我们把匹配问题看做是在这个二维匹配矩阵上的图像识别问题，利用在图像识别中成熟使用的卷积神经网络进行建模（图 13）。我们可以在第一层卷积中学到类似于 n-gram 和 n-term 的匹配层面的特征，而后的卷积层将底层的 n-gram 和 n-term 信号进行组合，类似于图像识别中，后面的卷积是将第一层得到的边缘信号进行组合。最终经过全连接得到句子之间的相似度。已经能够达到十分好的效果。而在机器翻译、对话系统、自动问答等需要语义匹配的任务上则需要更加灵活地匹配矩阵构建方式，而且提取的匹配特征就是在语义层面的 n-gram 和 n-term 信息，结合这些语义信息，所得到的模型在这些数据上也能有很好的表现。

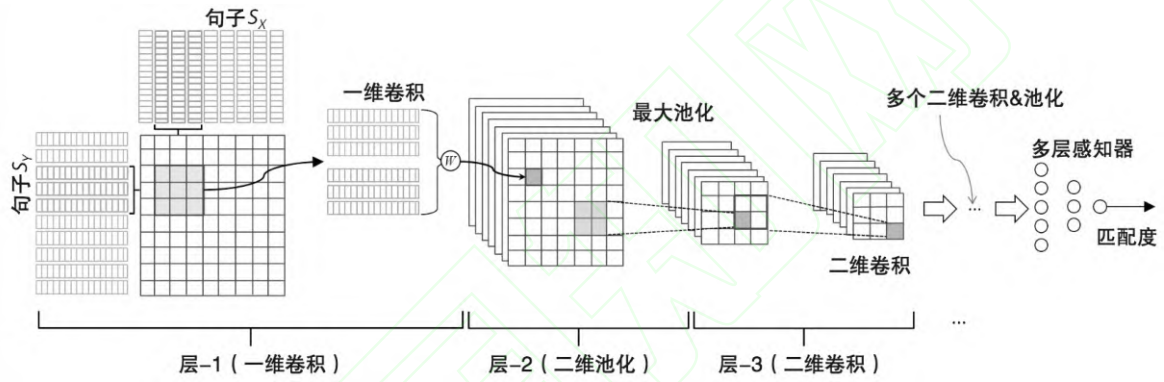

图 11 卷积网络深度匹配模型

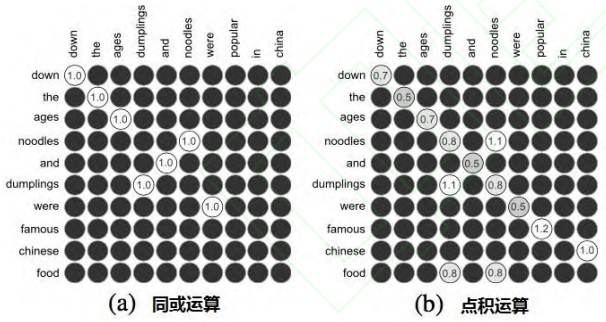

图 12 匹配矩阵

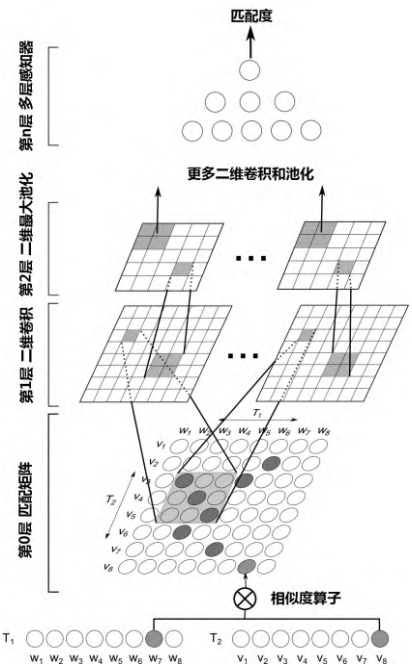

图 13MatchPyramid 模型

3.4.5 Match-SRNN

在得到匹配矩阵之后，我们发现利用二维的循环神经网络来建模特征空间的模式更为合理。二维的循环神经网络[66-68]，尤其是 2D-GRU 网络能够模拟最长公共子序列的计算过程，最长公共子序列也是快速求解字符串精确子串匹配的动态规划算法。因此我们提出了 Match-SRNN 模型[30]。

在 Match-SRNN 中,利用了前面提到过的神经张量网络，来捕获两段文本之间的基础交互信息，也就是单词级别的交互。具体地说，每个单词首先被表达成一个分布式向量。给定任意两个单词，x_i和y_j，以及他们的向量$\phi(x_i)$和$\phi(y_j)$，他们之间的交互信息也被表达成一个向量。因此我们得到一组匹配矩阵，堆叠起来就称为词级别的交互张量（Word Interaction Tensor）。词级别的交互张量可以看做一个二维结构，每个二维上的点是一个能够度量两个单词匹配度的向量，模型整体结构如图 14 所示。

在数学以及计算机科学中，当面临一个复杂对象的时候，一种常用的简化方式是去把一个复杂的问题分解为同样类型的一些子问题，然后通过递归的方式来解决这些问题，也就是递归(recursion)的思想。给定两段文本$s_1 = \{w_1, \ldots, w_m\}$和$s_2 = \{v_1, \ldots, v_n\}$，两个前缀$s_1[1:i] = \{w_1, \ldots, w_i\}$和$s_2[1:j] = \{v_1, \ldots, v_j\}$之间的交互(记做$h_{ij}$)是由他们子序列之间的交互以及当前位置单词级别的交互位置一起构成,也就是通过如下的方式：

$$h_{ij} = f(h_{i-1,j}, h_{i,j-1}, h_{i-1,j-1}, s(w_i, v_j)) \quad (6)$$

其中$s(w_i, v_j)$是两个单词w_i和v_j之间的交互信息。

根据二维 RNN 的计算模式，给定前缀之间的交互$s_1[1:i-1] \sim s_2[1:j]$, $s_1[1:i] \sim s_2[1:j-1]$, 以及$s_1[1:i-1] \sim s_2[1:j-1]$,分别记作$h_{i-1,j}$，$h_{i,j-1}$和$h_{i-1,j-1}$。那么$s_1[1:i]$和$s_2[1:j]$前缀之间的交互可以被表达为：

$$h_{ij} = f(h_{i-1,j}, h_{i,j-1}, h_{i-1,j-1}, s_{ij}) \quad (7)$$

我们可以看到二维 RNN 可以很自然地建模公式 6 中定义的递归匹配结构。

基于以上分析，二维 RNN 在词级别的交互张量上扫描。从左上角遍历到右下角就类似于阅读完两个句子，并找到了一条匹配的通路。匹配度从左上角开始累积，两个句子越匹配，这个积累的值越大。

实验表明，对于结构性明显的数据，例如自动问答数据，该模型有明显的提升。

3.4.6 小结

基于单语义文档表达和多语义文档表达的深度模型都是将重点放在单个文本表达成一个向量，与这些模型不同的是直接建模匹配模式的模型中并不存在单个文本的表达，从模型的输入开始两段

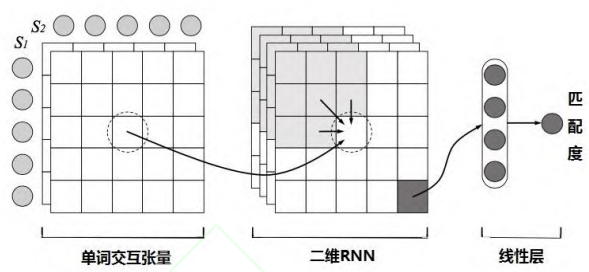

图 14Match-SRNN 模型

文本就进行了交互，得到细粒度的匹配信息。这就是说我们是在单词级别的表达上来构造$\mathbf{M_0}$的，而后会基于这个细粒度的匹配做更复杂的变换g和h。和基于单语义文档表达的深度模型相比，这样的好处在于保持细粒度的匹配信息，避免在一段文本抽象成一个表达时，细节的匹配信息丢失。虽然多语义文档表达模型在一定程度上缓解了这个问题，将不同粒度的表达拼接成最终的表达，希望细粒度的匹配信息在最后计算匹配得分的时候能够被考虑到，但是在得到表达之后的计算还是偏向于简单的向量相似度计算，缺少了直接建模匹配模式的挖掘，例如 Match-SRNN 这种复杂的可以模拟最长公共子序列的匹配模式。

但是这类模型也有相应的缺点：1）需要大量的有监督的文本匹配的数据训练，没法通过无监督的文本进行预训练；2）预测的时候资源消耗较大，每次都得完全通过一遍网络，没法像基于单语义文档表达或者多语义文档表达的模型可以离线计算好每个文本的特征，预测的时候直接利用算好的特征，并增量地计算新来的文本。因此这类模型一般都是用于类似问答系统、翻译模型、对话系统这种语义匹配程度高、句式变化复杂的任务中。

4 实验结果与分析

上述介绍的三类模型在不同类型的数据上有各自的特性，下面在复述问题、自动问答和信息检索，三个任务的三个经典数据集上，对比不同模型的表现。复述问题的两段文本更多是结构上的匹配，用词上是类似的；自动问答的问题和答案更偏

向于语义匹配，有些时候答案中不会出现问题中的词语；信息检索的查询项基本肯定会出现在匹配的文档中，其中查询项一般都很短，而文档会很长。由于这三个任务的特点，导致没有一个模型能够在所有的任务上都表现得最好。下面对这些任务和不同模型的结果做一个详细的分析。

4.1 复述问题实验

MSRP[5]数据集（Microsoft Research paraphrase）是复述问题的一个经典公开数据集。数据集总共包含 5801 对文本，一对匹配文本的标签为 1，反之为 0，其中标记正例的总共有 3900 对样本。文本的平均长度是 21，最短的文本长度是 7，最长的是 36。数据集切分成训练集总共包含 4076 对文本，测试集总共包含 1725 对文本。

例如，MSRP 数据集里面（*T1, T2*）匹配的一个实例：

T1: PCCW's chief operating officer, Mike Butcher, and Alex Arena, the chief financial officer, will report directly to Mr So.

T2: Current Chief Operating Officer Mike Butcher and Group Chief Financial Officer Alex Arena will report to So.

与前文描述的排序问题不同，MSRP 数据集的评价是按照分类问题来设定的，评价指标是准确率和F_1分值。准确率指的是分类正确的文本对的数量占总量的比例，F_1分值指的是针对匹配类别（标签为 1）的精度和召回率的几何平均：

$$F_1 = \frac{2PR}{P+R},$$

其中P表示精度，R表示召回率。

传统特征 TF-IDF 是文本挖掘领域广泛应用的一个特征，由 Salton 等人在 1983 年提出。TF 表示的是词在文本中出现的频次，而 IDF 表示这个词在整个数据集上的逆文档频率。IDF 的值越大表示这个词在文档中出现的越少，在匹配过程中也就越具有代表性。

表 2 模型在复述问题 MSRP 数据集上的结果
（其中部分实验结果来自已发表论文[24, 26, 27]，MV-LSTM 和 Match-SRNN 是本文实现的结果）

类别	模型	准确率（%）	F_1分值（%）
统计值	全正例	66.50	79.87
传统特征	TF-IDF	70.31	77.62
基于单语义文档表达	DSSM	70.09	80.96
	CDSSM	69.80	80.42
	ARC-I	69.60	80.27
基于多语义文档表达	uRAE	76.80	83.60
	MultiGranCNN	**78.10**	**84.40**
	MV-LSTM	75.40	82.80
直接建模匹配模式	ARC-II	69.90	80.91
	MatchPyramid	75.94	83.01
	Match-SRNN	74.50	81.70

实验结果（表 2）列出了三类算法在复述问题上的表现。总体比较三类算法，基于多语义文档表达和直接建模匹配模式的这两类模型明显优于基于单语义文档表达的模型。主要的原因在于复述问题本身更注重细粒度的匹配和对匹配模式的挖掘，单语义文档表达模型将文档的细节都压缩到了一个定长的向量，而且向量的每一维度都是整个句子信息的压缩，损失了对细节匹配的描述，而这些信息在其他两类模型中都是有保留的。这也是在复述问题上深度模型和传统特征（TF-IDF）差异并不明显的原因。进而我们发现最好的模型是基于多语义文档表达的 uRAE 和 MultiGranCNN，主要的原因是这两个模型在更大的数据集上进行了参数预训练。MSRP 这个数据集的数据量对于深度模型而言是比较小的，复杂的模型直接在这个数据集上很容易出现过拟合的现象。模型预训练使得模型更具鲁棒性，用预训练好的参数在小数据集上微调的结果更具泛化能力。

4.2 自动问答实验

Yahoo! Answers[7]社区问答数据集是在自动问答领域的一个经典公开数据集。整个数据集包含 142,627 个（问题，答案）对，我们只留下问题和答案的长度在 5 至 50 的样本（过滤掉一些噪声）。经过处理之后的数据集包含 60,564 个（问题，答案）对，这些样本作为匹配的正例。为了构造负例，我们利用问题作为查询项，利用现有的 Lucene 工具检索出 1000 个答案，每个问题的负例是在这 1000 个答案中随机抽取的 4 个。针对这个数据集按照 8:1:1 的比例划分了训练集、验证集和测试集，所有超参数在验证集上调整，而最终的结果是测试集上评价的。

Yahoo! Answers 社区问答数据集问题和答案对（Q, A）的一个实例：

Q: How to get rid of memory stick error of my sony cyber shot?
A: You might want to try to format the memory stick but what is the error message you are receiving.

这个任务的评价指标是前文介绍过的指标中的 P@1 和 MRR。因为对于当前任务一个问题只会有一个正确答案，所以我们只需要考虑这一个正确答案的位置。P@1 用来度量模型是否能够将正确答案排在第一个，而 MRR 则来度量模型是否能将正确答案排得更靠前。

传统特征 BM25 是信息检索领域常用的特征之一，由 Robertson 在 1995 年提出。BM25 特征融合了 TF-IDF 特征的信息，并考虑了文档长度的信息。BM25 的定义如下：

$$BM25(Q,D)=\sum_{i=1}^{n} IDF(q_i)\frac{TF_D(q_i)(k_1+1)}{TF_D(q_i)+k_1(1-b+\frac{b\|D\|}{avgdl})}$$

其中IDF表示逆文档频率，TF_D表示在文档 D 上的词频，$avgdl$表示平均文档长度，$\|D\|$表示当前文档长度，k_1和b是 BM25 的两个参数。

表 3 模型在自动问答 Yahoo！Answers 数据集上的结果（实验结果来自已发表论文[30]）

类别	模型	P@1	MRR
统计值	随机	0.200	0.457
传统特征	BM25	0.579	0.726
基于单语义文档表达	ARC-I	0.581	0.756
	CNTN	0.626	0.781
	LSTM-RNN	0.690	0.822
基于多语义文档表达	uRAE	0.398	0.652
	MultiGranCNN	0.725	0.840
	MV-LSTM	0.766	0.869
直接建模匹配模式	DeepMatch	0.452	0.679
	ARC-II	0.591	0.765
	MatchPyramid	0.764	0.867
	Match-SRNN	**0.790**	**0.882**

自动问答任务更偏向于语义匹配，也更需要考虑整体句子含义的匹配度，因此在实验结果（表 3）中，深度模型的效果明显好于传统特征（BM25）。对比三类深度模型，我们发现基于多语义文档表达和直接建模匹配模式的深度匹配模型要优于基于单语义文档表达的模型，主要原因还是细粒度的表达对于匹配问题是很重要的。多语义文档表达直接将细粒度的特征融入到最后的表达中，所以在之后计算匹配度的时候细粒度的匹配信息能够发挥作用；而直接建模匹配模式的模型则是在细粒度匹配完成之后，进一步挖掘匹配模式，也很好的保持了细粒度匹配的信息。由于 uRAE 模型并不是一个端到端的模型，而且自动编码器的训练比较困难，对于预训练的数据集地选择也比较敏感，导致 uRAE 在这个任务上的表现不是很好。DeepMatch 模型的前提假设是词语主题的预测是准确的，对于一个有限的数据集上主题的预测并不能达到很好的效果，影响了 DeepMatch 的表现。ARC-II 模型在细粒度匹配度的定义上比较模糊，近似是用加权平均两个细粒度的表达定义的，这个定义并不能很好反应细粒度上的匹配程度，所以整体的表现不是很好。

4.3信息检索实验

Robust04 数据集是 TREC（Text REtrieval Conference）的一个常用文档数据集，实验使用了 TREC Robust Track 2004 的所有主题作为查询项。其中包含 60 万左右的单词，50 万篇文档，250 个查询项，查询项的平均长度为 3，文档的平均长度为 447。我们使用已有的 Galago Search Engine 对查询项和文档进行预处理和索引，每一个查询项会索引出 2000 篇文档供模型进行排序。这个数据集由于查询项相对较少，我们使用留一交叉验证的方法检验模型的效果。数据集按 50 个查询项为一份分成了 5 份，每次留一份作为测试，其余的作为训练得到五个结果，然后这五个结果的平均作为最后的评价指标。

Robust04 数据集的查询项和文档对（Q, D）的一个实例：

Q: international organized crime
D: ... country activity the state should propose federal program for preserve and develop national culture and stimulate its diversity our opportunity are limited in the context of the economic crisis but the budgetary obligation which the state is taking on should be carry out society should know what proportion of the budget will be spent on culture science education and public ...

对于信息检索的任务，一个查询项匹配上的文档可能有多个，所以我们使用的评价指标是 MAP 和 nDCG@20。

表 4 模型在信息检索 Robust04 数据集上的结果（实验结果全是本文实现）

类别	模型	MAP	nDCG@20
传统特征	BM25	**0.255**	**0.418**
基于单语义文档表达	DSSM	0.095	0.201
	CDSSM	0.067	0.146
	ARC-I	0.041	0.066
基于多语义文档表达	MV-LSTM	0.119	0.185
直接建模匹配模式	ARC-II	0.067	0.147
	MatchPyramid	0.189	0.330
	Match-SRNN	0.203	0.374

在信息检索的任务中，我们发现深度模型的表现并不好（表 4），较好的模型（MatchPyramid 和 Match-SRNN）和传统 BM25 特征还是有一定差距的。其中主要的原因与信息检索任务的特点有关：1. 查询项和文档的长度差异很大（5.1 节详细分析）；2. 对精确匹配的要求比语义匹配要高很多，仅使用精确匹配信息（例如 BM25），就能达到很好的效果。在信息检索中引入词向量表达，有助于刻画查询项和文档之间语义匹配的关系，但是由于大量不相关的词也会计算得到一个匹配度，引入了太多噪声，再加上数据集本身数据量太小，最终导致深度模型更容易拟合到噪声上，影响了深度模型的效果。MatchPyramid 和 Match-SRNN 模型对细粒度的匹配信号保留的最好，选择合适的相似度计算函数（例如异或函数或者余弦相似度），拉开精确匹配和语义匹配的距离就能得到较好的效果。

5 未来研究方向展望

5.1 变长文本问题

基于深度学习的文本匹配模型能够处理的句子还都是一般长度的（10-500 词）文本。对于过长或者过短的文本，模型处理起来都比较棘手。首先对于过短的文本（小于 10 个词，例如很短的查询项），由于文本包含的词太少，如果进行深度并且复杂的变换和压缩，会导致短文本过度变换，影响模型的效果；而对于过长的文本（大于 500 个词，例如一整篇文章）则会引入大量无关的噪声，例如在信息检索领域，可能文档中只有一部分信息能够匹配上查询项，但是这部分匹配就足以检索出这篇文章[69, 70]。

我们还发现如果两段文本长度差异比较大的时候，比如一段文本是十个词左右，而另一段文本是几千个词，这个时候直接建模匹配模式的深度学习模型就会得到一个很窄的匹配矩阵。在这样的矩阵上进行匹配模式的挖掘，可能就会更偏向于更长的文本方向上的信息积累，而在短文本方向上能获取的信息比较有限。所以类似查询扩展[71]和文档摘要[72]这类的工作可以作为模型数据的预处理，如果能将这两方面的模型和现有模型进行一个整合，构造成一个端对端的可学习模型，将是个很好的改进。

5.2匹配可解释性

模型的可解释性是现在深度模型的一个通病，尤其在文本匹配领域，当两段文本匹配上了，如何解释和分解这个匹配的过程也将是一个重要的问题。本文介绍的部分模型在这个上面做了一些尝试，比如主题深度匹配模型认为两段文本匹配是根据各个区域的主题进行的，MatchPyramid 模型则认为匹配是在语义下的 n-gram 和 n-term 匹配，而 Match-SRNN 模型则认为文本的匹配更类似于在语义下扩展的最长公共子序列问题。这些假设可能并不是两段文本匹配上完全的原因，但至少窥探到了匹配问题的一些可以解释部分。进一步更合理地匹配可解释性的挖掘，希望可以提供更详尽的对于文本为什么能够匹配上的原因。

5.3判别式模型到生成式模型

深度文本匹配模型解决的问题是给定两段文本计算他们的匹配度，也就是一个判别式的模型，更进一步地说，如果我们通过大量样本学习到一段文本对应匹配的文本应该是什么样子的，那么我们是否可以构造出一个生成式的模型呢？在这方面，机器翻译和对话系统都有很多尝试来通过当前的文本生成匹配另一段文本。机器翻译中最成功的就是注意力模型（Attention Model）[64]，也很快用到了图像生成标题的任务中[73]。而在对话系统中也有利用卷积神经网络来构建生成式模型的*gen*CNN[74]。在这些尝试中，还没完全利用上判别式的深度匹配模型的一些发现，如何更好地利用这些，是一个难题，有待我们继续探索。

5.4跨模态文本匹配

本文主要介绍了文本之间的匹配建模，但这些模型其实是可以泛化到其他类型的数据上的，例如声音或者图像。文本和图像的匹配也称为跨模态的匹配，类比文本匹配，传统的跨模态匹配也是针对

文本和图像分别提取特征，然后计算特征之间的相似度，类似的方法有 Hodosh 等人使用的 KCCA（kernel canonical correlation analysis）[75]。最近的研究使用深度学习模型寻找更好的文本和图像的表达，甚至构造了从原始数据到最后结果的端到端的模型。其中包括 Sohcer 等人提出的 SDT-RNN（semantic dependency-tree recursive neural network）[76]，利用语义依存树将图像和文本映射到同一个空间；Karpathy 等人[77]提出的利用图像物体检测模型的结果和文本进行分块匹配；Lin Ma 等人提出的 m-CNN 模型[78]，旨在充分利用图像表达和文本不同粒度表达的匹配信息。

6 结束语

许多自然语言理解的问题我们都可以抽象成文本匹配的问题，本文重点关注了深度学习在文本匹配模型上的应用。面对文本匹配问题的三个挑战：词语匹配的多元性、短语匹配的结构性和文本匹配的层次性，都得到了不同程度的建模和解决。尤其当文本匹配模型遇到复杂的数据的时候，深度文本匹配模型对比于传统模型能够得到更好的效果。我们将深度文本匹配模型分成了基于单语义文档表达的深度学习模型、基于多语义文档表达的深度学习模型和直接建模匹配模式的深度学习模型三大类，并详细讨论了这三类模型之间的联系和优缺点。针对不同应用的特性，我们需要选择最适合的模型来建模。深度文本匹配模型在文本匹配这个基础的任务上取得了很好的效果，当然也有不少问题需要我们进一步改进和优化。

致　谢 本课题得到国家重点基础研究发展计划(973) (No. 2014CB340401, 2013CB329606)、国家自然科学基金重点项目(No. 61232010, 61472401, 61425016, 61203298)、中国科学院青年创新促进会(No. 20144310, 2016102)资助.

参考文献

[1] Li H, Xu J. Semantic matching in search. Foundations and Trends in Information Retrieval, 2014, 7(5): 343-469.

[2] Xue X, Jeon J, Croft W B. Retrieval models for question and answer archives//Proceedings of the Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. Singapore, 2008: 475-482.

[3] Brown P F, Pietra V J D, Pietra S a D, et al. The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 1993, 19(2): 263-311.

[4] Serban I V, Sordoni A, Bengio Y, et al. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models. arXiv preprint arXiv:150704808, 2015,

[5] Dolan W B, Brockett C. Automatically constructing a corpus of sentential paraphrases//Proceedings of the Third International Workshop on Paraphrasing, Jeju Island, Korea, 2005: 9–16.

[6] Das D, Smith N A. Paraphrase identification as probabilistic quasi-synchronous recognition//Proceedings of the Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP.Stroudsburg, USA, 2009: 468-476.

[7] Surdeanu M, Ciaramita M, Zaragoza H. Learning to rank answers to non-factoid questions from web collections. Computational linguistics, 2011, 37(2): 351-383.

[8] Robertson S, Zaragoza H. The probabilistic relevance framework: BM25 and beyond. Foundations and Trends in Information Retrieval, 2009，3(4):333–389

[9] Bu Zhi-Qiong, Zheng Bo-Jin. Ad hoc information retrieval method based on LDA. Application Research of Computers, 2015, 32(5): 1369-72 (in Chinese)
(卜质琼, 郑波尽. 基于 LDA 模型的 Ad hoc 信息检索方法研究. 计算机应用研究, 2015, 32(5): 1369-72).

[10] Ganguly D, Roy D, Mitra M, et al. Word embedding based generalized language model for information retrieval//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2015: 795-798.

[11] Zheng Shi-Fu, Liu Ting, Qin Bing, Li Sheng. Overview of Question Answering. Journal of Chinese Information Processing, 2002, 16(6): 47-53 (in Chinese)
(郑实福, 刘挺, 秦兵, 李生. 自动问答综述. 中文信息学报, 2002, 16(6): 47-53).

[12] Wu You-Zheng, Zhao Jun, Duan Xiang-Yu, Xu Bo. Research on Question Answering & Evaluation : A Survey. Journal of Chinese Information Processing, 2005, 19(3): 2-14 (in Chinese)
(吴友政, 赵军, 段湘煜, 徐波. 问答式检索技术及评测研究综述. 中文信息学报, 2005, 19(3): 2-14).

[13] Zhang Wei-Nan, Zhang Yu, Liu Ting. A Topic Inference Based Translation Model for Question Retrieval in Community-Based Question Answering Services. Chinese Journal of Computers, 2014, 37（1）: 1-8 (in Chinese)
(张伟男, 张宇, 刘挺. 一种面向社区型问句检索的主题翻译模型. 计算机学报, 2014, 37（1）: 1-8).

[14] Huang Pei-Jie, Huang Qiang, Wu Xiu-Peng, et al. Question Understandingby Combining Grammar and Semantic for Chinese Dialogue System. Journal of Chinese Information Processing, 2014, 28(6): 70-8 (in Chinese)
(黄沛杰, 黄强, 吴秀鹏, 等. 语法和语义相结合的中文对话系统问题理解研究. 中文信息学报, 2014, 28(6): 70-8).

[15] Yu Kai, Chen Lu, Chen Bo, Sun Kai, Zhu Su. Cognitive Technology Task-Oriented Dialogue Systems – Concepts, Advances and Future. Chinese Journal of Computers, 2015, 38(12): 2333-48 (in Chinese)
(俞凯, 陈露, 陈博, et al. 任务型人机对话系统中的认知技术——概念, 进展及其未来. 计算机学报, 2015, 38(12): 2333-48).

[16] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436-44.

[17] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality//Proceedings of the Advances in neural information processing systems. Lake Tahoe, USA, 2013: 1-9.

[18] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[19] Huang P-S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data//Proceedings of the 22nd ACM international conference on Conference on information and knowledge management. Amazon, India, 2013: 2333-2338

[20] Wang M, Lu Z, Li H, et al. Syntax-based deep matching of short texts//Proceedings of the International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina, 2015: 1354-1361.

[21] Shen Y, He X, Gao J, et al. A latent semantic model with convolutional-pooling structure for information retrieval//Proceedings of the 23rd ACM international conference on Conference on information and knowledge management. New York, USA, 2014: 101-110.

[22] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences//Proceedings of the Advances in Neural Information Processing Systems, Montreal, Canada, 2014: 2042-2050.

[23] Qiu X, Huang X. Convolutional neural tensor network architecture for community-based question answering//Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), Buenos Aires, Argentina, 2015: 1305-1311.

[24] Yin W, Schütze T, Hinrich. MultiGranCNN: An Architecture for General Matching of Text Chunks on Multiple Levels of Granularity//Proceedings of the 53rd Annual meeting of the association for computational linguistics, Beijing, China, 2015: 63-73.

[25] Lu Z, Li H. A deep architecture for matching short texts//Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, USA, 2013: 1367-1375.

[26] Pang L, Lan Y, Guo J, et al. Text matching as image recognition//Proceedings of the 30th AAAI Conference on Artificial Intelligence.Phoenix, USA, 2016: 2793-2799.

[27] Socher R, Huang E H, Pennin J, et al. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection//Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 2011: 801-809.

[28] Palangi H, Deng L, Shen Y, et al. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, 24(4): 694-707.

[29] Wan S, Lan Y, Guo J, et al. A deep architecture for semantic matching with multiple positional sentence representations//Proceedings of the 30th AAAI Conference on Artificial Intelligence.Phoenix, USA, 2016: 2835-2841.

[30] Wan S, Lan Y, Guo J, et al. Match-SRNN: Modeling the recursive matching structure with spatial RNN//Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, USA, 2016: 1022-1029.

[31] Robertson S. Evaluation in information retrieval//Lectures on information retrieval. Springer Berlin Heidelberg, 2000: 81-92.

[32] Berger A, Lafferty J. Information retrieval as statistical translation//Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. Berkeley, USA, 1999: 222-229.

[33] Gao J, He X, Nie J Y. Clickthrough-based translation models for web search: from word models to phrase models//Proceedings of the 19th ACM international conference on Information and knowledge management. Toronto, Canada, 2010: 1139-1148.

[34] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. Neural computation, 2004, 16(12): 2639-64.

[35] Rosipal R, Krämer N. Overview and recent advances in partial least squares//Subspace, latent structure and feature selection. Berlin Heidelberg :Springer, 2006: 34-51.

[36] Wu W, Lu Z, Li H. Learning bilinear model for matching queries and documents. Journal of Machine Learning Research, 2013, 14(1): 2519-2548.

[37] Bai B, Weston J, Grangier D, et al. Supervised semantic indexing//Proceedings of the 18th ACM conference on information and knowledge management. Hongkong, China, 2009: 187-196.

[38] Gao J, Toutanova K, Yih W. Clickthrough-based latent semantic models for web search//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. Beijing, China, 2011: 675-684.

[39] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 1995, 3361(10): 1995.

[40] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Boston, USA, 2015: 1-9.

[41] Abdel-Hamid O, Mohamed A, Jiang H, et al. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition//2012 IEEE international conference on Acoustics, speech and signal processing. Kyoto, Japan, 2012: 4277-4280.

[42] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch. Journal of Machine Learning Research, 2011, 12(Aug): 2493-2537.

[43] Vinyals O, Kaiser Ł, Koo T, et al. Grammar as a foreign language//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 2773-2781.

[44] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank//Proceedings of the conference on empirical methods in natural language processing.Seattle, Washington,USA,2013: 1631-1642.

[45] Zeng D, Liu K, Lai S, et al. Relation Classification via Convolutional Deep Neural Network//Proceedings of the COLING. Dublin, Ireland, 2014: 2335-2344.

[46] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[47] Levin E. A recurrent neural network: Limitations and training. Neural Networks, 1990, 3(6): 641-650.

[48] LeCun Y. Generalization and network design strategies. Connectionism in perspective, 1989: 143-155.

[49] Lv Zheng-Dong, Li Hang. Apply Deep Matching Learning in Language Matching. Communications of China Computer Federation. 2015, 8(8): 30-38 (in Chinese)
(吕正东, 李航. 深度匹配学习在语言匹配中的应用. 中国计算机学会通讯, 2015, 8(8): 30-38).

[50] Le Q V, Mikolov T. Distributed Representations of Sentences and

Documents//Proceedings of the31st International Conference on Machine Learning. Beijing, China, 2014: 1188-1196.

[51] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, Maryland, 2014: 655-665.

[52] Kim Y. Convolutional neural networks for sentence classification//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014: 1746-1751.

[53] Li J, Jurafsky D, Hovy E. When are tree structures necessary for deep learning of representations?//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing, China, 2015: 2304-2314.

[54] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, USA, 2015 :2267-2273.

[55] Socher R, Lin C C, Manning C, et al. Parsing natural scenes and natural language with recursive neural networks//Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA, 2011:129-136.

[56] Irsoy O, Cardie C. Deep recursive neural networks for compositionality in language//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2014:2096-2104.

[57] Chopra S, Hadsell R, Lecun Y.Learning a similarity metric discriminatively, with application to face verification//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.San Diego, USA, 2005,1:539-546.

[58] Shen Y, He X, Gao J, et al. Learning semantic representations using convolutional neural networks for web search//Proceedings of the 23rd International Conference on WWW. Seoul, Korea, 2014:373-374.

[59] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion//Proceedings of the Advances in neural information processing systems. Lake Tahoe, USA, 2013:926-934.

[60] Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation, 1997, 9(8): 1735-1780.

[61] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions//Proceedings of the Twentieth Annual Symposium on Computational Geometry. Brooklyn, USA,2004:253-262.

[62] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Edinburgh, UK,2011:151-161.

[63] Yin W, Schütze H. Convolutional neural network for paraphrase identification//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver, USA,2015:901-911.

[64] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473, 2014.

[65] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 2005, 18(5): 602-610.

[66] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks//Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing. Vancouver, Canada, 2013: 6645-6649.

[67] Graves A, Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks//Proceedings of the Advances in neural information processing systems.Vancouver, Canada, 2009: 545-552.

[68] Theis L, Bethge M. Generative image modeling using spatial LSTMs//Proceedings of the Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 1927-1935.

[69] Salton G, Allan J, Buckley C. Approaches to passage retrieval in full text information systems//Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval. Pittsburgh, USA, 1993: 49-58.

[70] Liu X, Croft W B. Passage retrieval based on language models//Proceedings of the international conference on information and knowledge management. Mclean, USA, 2002: 375-382.

[71] Xu J, Croft W B. Query expansion using local and global document analysis//Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. Zurich, Switzerland, 1996: 4-11.

[72] Li P, Bing L, Lam W, et al. Reader-aware multi-document summarization via sparse coding//Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina, 2015: 1270-1276.

[73] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015: 2017-2027.

[74] Wang M, Lu Z, Li H, et al. gen CNN: A Convolutional Architecture for Word Sequence Prediction//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. Beijing, China, 2015: 1567-1576.

[75] Hodosh M, Young P, Hockenmaier J. Framing image description as a ranking task: Data, models and evaluation metrics. Journal of Artificial Intelligence Research, 2013, 47: 853-899.

[76] Socher R, Karpathy A, Le Q V, et al. Grounded compositional semantics for finding and describing images with sentences. Transactions of the Association for Computational Linguistics, 2014, 2: 207-218.

[77] Karpathy A, Joulin A, Li F F F. Deep fragment embeddings for bidirectional image sentence mapping//Proceedings of the Advances in neural information processing systems. Montreal, Canada, 2014: 1889-1897.

[78] Ma L, Lu Z, Shang L, et al. Multimodal convolutional neural networks for matching image and sentence//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 2623-2631.

PangLiang, born in 1990, Ph.D. candidate. His research interests include deep learning and text mining.

Lan Yanyan, born in 1982, Ph.D. , associate professor. Her research interests include machine learning, learning to rank and information retrieval.

Xu Jun, born in 1979, Ph.D. , professor. His research interests include information retrieval and data mining.

Guo Jiafeng, born in 1980, Ph.D. , associate professor. His research interests include information retrieval and data mining.

Wan Shengxian, born in 1989, Ph.D. His research interests include deep learning and text mining.

Cheng Xueqi, born in 1971, Ph.D. , professor. His research interests include network science, network and information security, web search and data mining.

Background

Text matching is a key problem in many natural language processing tasks, such as information retrieval, question answering, machine translation, dialog system, paraphrase identification and so on. The past researches on text matching focused on defining artificial features and learning relation between two text features. Recently, Deep Text Matching models have been proposed to tackle this problem and outperform the traditional models. Deep Text Matching models utilize the idea of automatically feature extraction in deep learning, thus comparing to the traditional methods, Deep Text Matching models can automatically learn relations among words from big data and make use of the information from phrase patterns and text hierarchical structures. Considering the different structures of Deep Text Matching models, we divide them into three categories: Single semantic document representation based deep matching model, Multiple semantic document representation based deep matching model and Matching pattern based deep matching model. We can see the progressive relationship among three kinds of models in modelling the interaction of texts, while which have their own merits and defects based on a specific task.

This work was supported by the 973 Program of China under Grant No. 2014CB340401 and 2013CB329606, the National Natural Science Foundation of China under Grant No. 61232010, 61472401, 61425016, and 61203298, and the Youth Innovation Promotion Association CAS under Grant No. 20144310 and 2016102.