

双通道深度主题特征提取的文章推荐模型^①



井明强, 房爱莲

(华东师范大学 计算机科学与技术学院, 上海 200062)

通信作者: 房爱莲, E-mail: alfang@cs.ecnu.edu.cn

摘要: 针对传统的文章推荐方法存在的冷启动、用户反馈稀疏以及相似度衡量准确性欠佳等问题, 本文提出了融合主题模型和预训练模型 BERT 的文章相似度计算模型 (contextualized topic BERT, ctBERT). 给定查询, 该算法会计算查询与相关文章之间的相似度分数, 文章经过预处理分别输入独立的子模块进行特征抽取并计算相似度得分, 相似度得分与支撑集的个性化得分相结合以获得最终分数, 该方法将单样本学习整合进推荐框架中, 进一步取得了显著的改进. 本文在 3 个不同的数据集上的实验结果表明, 所提出方法在 3 个数据集上的 NDCG 标准均有提升, 例如在 Aminer 数据集上 NDCG@3 和 NDCG@5 标准对比方法分别提高了 6.1% 和 7.2%, 验证了该方法的有效性.

关键词: 个性化推荐; 主题模型; 文本相似度; 单样本学习; 潜在狄利克雷

引用格式: 井明强, 房爱莲. 双通道深度主题特征提取的文章推荐模型. 计算机系统应用, 2022, 31(10): 323-328. <http://www.c-s-a.org.cn/1003-3254/8735.html>

Article Recommendation Model with Two-channel Deep Topic Feature Extraction

JING Ming-Qiang, FANG Ai-Lian

(School of Computer Science and Technology, East China Normal University, Shanghai 200062, China)

Abstract: For the cold start, sparse user feedback, and poor accuracy of similarity measurement in traditional article recommendation methods, this study proposes contextualized topic BERT (ctBERT), an article similarity calculation method that combines BERT with the topic model. The algorithm calculates the similarity scores between the given query and the related articles. The preprocessed articles are input into separate sub-modules for feature extraction and similarity score calculation. The similarity score is combined with the personalization score of the support set to obtain the final score. The algorithm is further improved by integrating single-sample learning into the recommendation framework. The experimental results from three different datasets show that the proposed method improves the NDCG criteria on all three datasets. For example, the NDCG@3 and NDCG@5 criteria improve by 6.1% and 7.2% respectively compared with other methods on the Aminer dataset, which validates the effectiveness of the method.

Key words: personalized recommendation; topic model; text-similarity; one-shot learning; LDA

现在在线文章的数量已经非常庞大, 而且每年都在急剧增加, 根据 NCSES 的统计, Scopus S&E 出版物数据库的数据, 2020 年的出版物产量达到 290 万篇. 过去 4 年 (2017-2020 年) 以每年 6% 的速度增长, 如何向用户推荐他们喜欢并且适合他们的文章成为数字图书馆服务提供商面临的一个关键问题. 许多学术搜索网

站在某一特定领域的信息页面上提供文章推荐来帮助用户查找相关文章. 然而, 这些建议往往是基于关键字的推荐, 如周兴美^[1]提出的论文检索系统, 通过将文章的段落、语句、常用词、关键词、结构框架内容等检索筛选、做出对比来推荐文章, 这并不能个性化地向用户推荐文章.

① 收稿时间: 2021-12-30; 修改时间: 2022-01-31; 采用时间: 2022-02-24; csa 在线出版时间: 2022-07-07

通常,一篇文章会涵盖多个不同的主题,例如本文中,如关键字所示,包含:个性化推荐、主题模型、文本相似度.不同背景和兴趣的用户可能更喜欢与不同主题相关的文章.个性化推荐如果不考虑用户的多样性,也不能满足多数用户.良好的个性化对于一个进行文章推荐的网站来说是一个挑战,很多用户都是冷启动用户,他们的个人资料并不完善,不能提供很多有用的信息.典型的方法是基于用户反馈,然而用户反馈可能非常稀疏.对于新用户,只有来自同一会话的隐式反馈可用.因此很难利用传统的推荐方法,如基于内容的推荐^[2]或协同过滤^[3]等.

文本相似度衡量在推荐系统和信息检索领域中起着非常关键的作用,也是一个挑战.大多数传统方法都是基于词袋模型,丢失了有关词语顺序以及上下文的信息.因此这些方法无法捕捉到短语、句子或更高级别的文意信息.近年来,词嵌入和神经网络在各个领域都得到了广泛的应用,在自然语言处理领域也提出了许多可以直接处理词序列的神经相似度计算模型^[4,5],但它们经常不加区分的处理文章中的所有单词,因此它们无法将文章的重要部分与文章的一些固定化表达(比如“我们发现”“这说明了”)区分开来.

为了应对这些挑战,本文提出一个基于双通道深度主题特征提取的文章推荐模型,将相关主题模型CTM^[6]与预训练模型BERT^[7]相结合,文章经过预处理后,分别输入到独立的子模块中进行高维信息抽取并计算相似性得分,结合个性化得分给出最终推荐.我们发现,在BERT中添加相关主题信息,在3个数据集上均有明显提升,尤其是在特定领域的数据集上有优秀的表现.本文的主要贡献有:

(1) 提出了基于双通道深度主题特征提取的文本相似度计算模型ctBERT,融合了相关主题模型CTM和深度预训练模型BERT作为文本语义相似度预测工具.

(2) 我们在不同来源和尺度上的数据集上进行实验.实验结果表明,我们的模型相比于其他的基线模型能有效提高性能.

1 文章个性化推荐系统

1.1 个性化推荐

个性化推荐系统旨在为特定用户在给定的上下文中提供最合适的推荐项目,基于内容的方法^[1]将项目

描述和用户的画像文件进行比较,以得到推荐内容.经典的协同过滤方法^[8]利用当前用户或者与当前用户类似的其他用户的历史行为数据进行预测.混合推荐方法^[9]则结合了上述两种方法的优点以提高建模精度.

杜政晓等人^[10]将个性化文章推荐问题定义如定义1.

定义1. 个性化文章推荐问题

输入: 查询文章 d_q , 候选集 $D=\{d_1, d_2, \dots, d_N\}$, 与用户 u 相关的支撑集 $S=\{(\vec{d}_i, \vec{s}_i)\}_{i=1}^T$.

输出: 一个为用户 u 生成的前 k 个完全排序的集合 $R(d_q, S) \subset D$, 其中 $|R|=k$.

1.2 文本相似度

文本相似度是指通过某种方法获得一个值来描述文本对象之间的相似度,如单词、句子、文档等.衡量两篇文档之间的相似度的传统方法如BM25^[11]和TF-IDF^[12]都是基于词袋模型,这些方法通常将两篇文章中匹配单词的权重之和作为相似度分数,忽略了语义信息,所以在衡量短语和句子的匹配程度方面表现不佳.

基于神经网络的方法可以大致分为两种类型:基于表示的模型和基于交互的模型.基于表示的代表模型有DSSM, LSTM-RNN以及MV-LSTM等,他们的主要思想是使用神经网络获取文章的语义表示,然后将两篇文章的表示之间的相似性(通常是余弦相似性)作为相似性分数.然而这些模型通常缺乏识别特定匹配信号的能力.基于交互的模型通常是依赖于词嵌入,使用神经网络来学习两篇文章的单词级交互模式,比如DRMM在单词相似直方图上使用多层感知器来获得两篇文章的相似度分数,其他的代表模型有MatchPyramid以及K-NRM等.这些模型主要依赖于单词的嵌入,忽略了文章中的潜在主题信息.

1.3 相关主题模型CTM

主题模型的核心思想是以文本中所有字符作为支撑集的概率分布,表示该字符在该主题中出现的频繁程度,即与该主题关联性高的字符有更大概率出现.在文本拥有多个主题时每个主题的概率分布都包括所有字符,但一个字符在不同主题的概率分布中的取值是不同的,它可以发现隐藏在大规模文档集中的主题结构.

相关主题模型(CTM)^[6]是建立在Blei等人^[13]的潜在狄利克雷(LDA)模型之上,LDA假设每个文档的单词都来自不同的主题,其中每个主题都有一个固定的词汇表,文档集中的所有文档均共享这些主题,但

是文档之间的主题比例会有不同的差异,因为他们是从狄利克雷分布中随机抽取的. LDA 模型是一个典型的词袋模型,无法直接对主题之间的相关性进行建模. CTM 用更灵活的逻辑正态分布取代了狄利克雷,该正态分布在组件之间包含协方差结构,这给出了一个更现实的潜在主题结构模型,其中一个潜在主题可能与另一个潜在主题相关.

1.4 单样本学习

单样本学习可以将分类任务转化为差异评估问题,当深度学习模型针对单样本学习模型进行调整的时候,它会获取两个目标并返回一个值,这个值显示两个目标之间的相似性,如果两个目标有包含相同的特征,神经网络则返回一个小于特定阈值(例如 0)的值,如果两个目标差异较大,返回的值将会高于阈值. 单样本学习目前主要被应用于图像任务中,用于少量样本情况下的图片分类任务. 它可以将在其他类中学到的知识建模为与模型参数有关的先验概率函数. 当给出一个新类的示例时,它可以更新知识并生成一个后验密度来识别新的实例.

2 方法

2.1 基于单样本学习的推荐框架

为了得到有序集 R ,我们的模型为候选集 D 中的每篇文章 d_i 计算了一个分数 $s(d_i|d_q, S)$,然后选取 D 中获得最高分数的前 k 篇文章作为 top-k 推荐. 针对特定用户 u 的推荐问题可以看作用户 u 是否接受这篇文章,以此将问题转化为二分类问题,即对于支撑集 S 中的 $(\hat{d}, \hat{y}) \in S$, \hat{y} 是二进制的(1表示用户接受该文章,0表示不接受). 可以将 S 视为分类的训练集,其中 \hat{d} 是训练实例, \hat{y} 是相应的标签,由于前期用户的交互数据非常少, S 的规模很小甚至为空,我们将单样本学习与我们的问题类比. 受文献[14]的启发,我们的模型计算 $s(d_i|d_q, S)$ 的方式如下:

$$s(d_i|d_q, S) = \begin{cases} t(d_q, d_i), & S = \emptyset \\ t(d_q, d_i) + \frac{1}{|S|} \sum_{(\hat{d}, \hat{y}) \in S} t(\hat{d}, d_i) \hat{y}, & S \neq \emptyset \end{cases} \quad (1)$$

其中, $t(\cdot, \cdot)$ 是用于计算文本相似性的 ctBERT 模型,将在下面的部分中讨论. 从式(1)中可以看出, s 由两部分组成,第1部分是与查询文章的相似度得分. 第2部分

是个性化得分,是 S 中反馈的归一化线性组合,以文本相似度为系数, S 为空时等于 0. 整个推荐框架如图 1 所示.

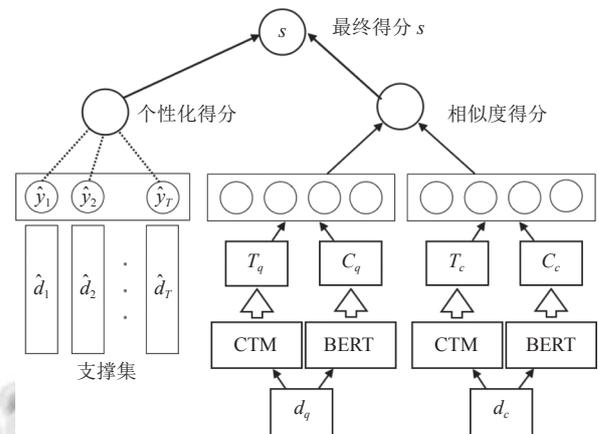


图 1 ctBERT 模型的整体架构

2.2 文本语义相似度得分计算

首先,需要对文章进行预处理,包括切词、去除停用词等. 由于 BERT 模型支持的最大输入长度只有 512, 本文采用的方法是按固定长度进行截断, 即对于每篇文档 d_i , 选取去除停用词后的前 512 个字作为输入. 为了提高性能和减少计算开销, 本文根据文献[15]的启发, 采用了双塔模型结构. 通过独立的子模块进行高维信息抽取, 再用抽取得到的特征向量的余弦距离来得到文本语义相似度得分.

对于查询文档 d_q 和候选文档 d_c , 分别输入 BERT 模型, 这里的 BERT 模型选用 BERT-BASE, 最终会得到 BERT 最后一层 [CLS] 向量输出, 将其作为文档的全局语义特征表示 C , 形式化如式(2):

$$C = \text{BERT}(d) \quad (2)$$

关于主题模型部分, 本文使用 Bianchi 等人^[16]的 CTM 模型对文档进行主题建模, 使用 Word2Vec 模型对词向量进行训练, 其计算方式如式(3):

$$x = \{e(w_1), e(w_2), \dots, e(w_v)\} \quad (3)$$

其中, x 为所有词汇的词向量表征形式, v 是词汇字典中的词汇总数, $e(w_i)$ 为词汇 w_i 的词向量表示形式. 通过主题模型训练结果, 得到文档 d 对应的主题分布, 取概率最大的主题 T_{\max} , 在根据主题词文件选择主题 T_{\max} 下的前 n 个词 (t_1, t_2, \dots, t_n) 及其概率值 (p_1, p_2, \dots, p_n) , 将概率值归一化处理作为 n 个词语的权重信息, 公式如式(4):

$$q_i = \frac{p_i}{\sum_{a=1}^r p_a} \quad (4)$$

其中, q_i 表示 p_i 归一化后的值, (q_1, q_2, \dots, q_n) 即为前 n 个词的权重大小. 通过训练好的 Word2Vec 模型得到词向量 $(e(t_1), e(t_2), \dots, e(t_n))$, 再对其进行加权求和, 得到主题扩展特征, 公式如式 (5):

$$D = \sum_{b=1}^r q_b \times e(t_b) \quad (5)$$

接下来, 对全局语义特征表示 C 和主题扩展特征 D 进行组合以得到文档向量与文档主题表示相结合的表现形式:

$$F = [C; D] \quad (6)$$

对于查询文档和候选文档的最终表示 F_q 和 F_c , 通过一个隐藏层统一长度, 最后计算他们的余弦相似度得出相似度得分.

2.3 优化与训练

整个模型在目标任务上进行端到端的训练, 我们将较链损失函数作为训练的目标函数. 给定三元组 $\{(d_q^{(i)}, d_+^{(i)}, d_-^{(i)})\}_{i=1}^N$, 其中对于查询文档 $d_q^{(i)}$, $d_+^{(i)}$ 的排名要高于 $d_-^{(i)}$, 损失函数表达式为:

$$Loss = \sum_{i=1}^N \max(0, 1 - t(d_q^{(i)}, d_+^{(i)}) + t(d_q^{(i)}, d_-^{(i)})) \quad (7)$$

其中, $t(d_q, d)$ 表示查询文章 d_q 和文章 d 的相似度得分. 通过标准反向传播和小批量随机梯度下降法进行优化, 同时采用随机丢弃和早期停止策略避免过拟合.

3 实验

在本节中, 为了评估所提出的模型, 我们在 3 个数据集上对文章推荐问题进行了实验, 并与传统方法和神经模型做了对比.

3.1 数据集

第 1 个数据集是来自 Aminer^[17] 的论文集, 由 188 篇查询文档组成, 每个查询文档有 10 篇候选论文, 这些文档均是计算机相关的专业论文. 第 2 个数据集是基于美国专利和商标局的专利文件 (Patent), 由 67 个查询组成, 每个查询有 20 个候选文档. 第 3 个数据集是来自 Sowiport 的相关文章推荐数据集 (RARD), 这是一个社会科学文章的数字图书馆, 向用户提供相关文章. 此数据集包含 63 932 个带有用户点击记录的不同查询. 每篇查询文章平均有 9.1 篇文章显示, 我们选取了 100 个点击率最高的查询进行测试, 其他的查询则用于训练.

3.2 实验对比方法

下面是本文中用来做对比的几种方法, 其中前两

种方法为传统方法, 后 3 种方法为神经模型.

(1) TF-IDF^[12]: 将查询文档和候选文档中共同出现的词语的权重相加来得到文档之间的相似度得分. 一个词的权重是其词频和逆文档频率的乘积.

(2) Doc2Vec^[18]: 通过段落向量模型得到每一篇文章的分布表示, 然后通过计算他们之间的余弦相似性来得到文档之间的相似度得分.

(3) MV-LSTM^[5]: 由 Bi-LSTM 生成的不同位置的句子表征形成了一个相似性矩阵, 由这个矩阵来生成相似度得分.

(4) DRMM^[19]: 基于查询文档和候选文档的词对向量构建局部交互矩阵, 根据此矩阵转化为固定长度的匹配直方图, 最后将直方图发送到 MLP 以获得相似度得分.

(5) LTM-B^[20]: 采用分层的思想将文档切分成多个分段, 使用 BERT 将文本向量化, 将得到的矩阵表示与 BiLSTM 产生的位置矩阵求和之后输入到 Transformer 中进行特征提取, 最后将两个文档矩阵进行交互、池化拼接后经全连接层分类得到匹配分数.

3.3 实验设计与结果分析

(1) 实验 1: 与其他方法的对比实验

表 1 展示了不同方法在归一化折损累计增益 (NDCG) 指标方面的排名精度. 为了比较的公平性, 所有模型都不涉及用户反馈, 这一点会在后面讨论.

表 1 不同方法在 NDCG 指标方面的排名精度 (%)

方法	Aminer		Patent		RARD	
	NDCG	NDCG	NDCG	NDCG	NDCG	NDCG
	@3	@5	@3	@5	@3	@5
TF-IDF	73.4	76.3	52.1	55.8	33.6	42.2
Doc2Vec	60.0	66.7	44.8	45.6	27.4	32.3
MV-LSTM	56.4	60.3	60.2	59.0	24.2	31.7
DRMM	74.2	78.0	55.3	56.3	33.2	35.3
LTM-B	70.0	74.2	62.4	62.8	35.1	35.3
ctBERT	80.3	85.2	63.8	64.5	40.2	44.4

从表 1 中的评估结果可以看出, 我们提出的模型性能优于其他的方法, 在 NDCG@3 标准上优于对比方法 3.6%–19.6%, 在 NDCG@5 标准上优于对比方法 2.2%–24.9%, 在每个数据集上 NDCG 平均提高分别为: 14.6%、5.8%、9.8%. 我们注意到, 在 Aminer 数据集的提升效果要明显优于其他两个数据集, 说明我们的模型在涉及特定专业领域的数据集上有更好的表现.

(2) 实验 2: 消融实验

为了探究主题模型和预训练模型分别对 ctBERT

的性能影响,我们为消融实验建立基线模型,分别是:

(1) ctBERT-ctm: 去掉主题模型, 仅用预训练模型 BERT 对文章进行建模.

(2) ctBERT-BERT: 去掉预训练模型, 仅用主题模型 CTM 对文章进行建模.

为了比较的公平性, 本实验均不涉及用户反馈. 它们与 ctBERT 模型的性能比较如表 2 所示.

表 2 消融实验结果对比 (%)

方法	Aminer		Patent		RARD	
	NDCG	NDCG	NDCG	NDCG	NDCG	NDCG
	@3	@5	@3	@5	@3	@5
ctBERT-ctm	66.2	68.4	48.7	54.0	31.3	38.0
ctBERT-BERT	44.4	48.3	32.1	42.3	18.9	19.4
ctBERT	80.3	85.2	63.8	64.5	40.2	44.4

实验结果可以看出, 仅用预训练模型 BERT 对文章进行建模的方法 ctBERT-ctm 相比表 1 中的 Doc2Vec 以及 MV-LSTM 方法有部分性能提升, 但是效果仍不如其他对比的基线模型, 说明此种方式对文章语义信息的表示有限. 去掉预训练模型的 ctBERT-BERT 模型仅用概率最大的主题词词向量作为文章表示, 实验结果表明此方法在 3 个数据集上均无法达到很好的效果.

(3) 实验 3: 单样本学习对整个框架的影响

为了探究单样本学习对整个框架的影响, 我们设置了以下实验, 探究加入单样本学习之后 ctBERT 模型的表现. 由于以上 3 个数据集均没有用户个性化标签, 我们利用数据集中的数据模拟个性化问题. 首先选择有多个正面标记的候选项的查询, 对于每个查询, 我们将标记的候选项随机分为两部分. 第 1 部分用于支撑集, 第 2 部分用作将要推荐的候选集. 然后我们将单样本学习推荐框架 (称为 ctBERT-os) 与前面不涉及用户反馈的最佳模型 ctBERT 进行比较. 结果如表 3 所示. 可以看到, 通过加入支撑集, 可以提高模型性能, 平均而言, ctBERT-OS 模型的表现 NDCG@3 和 NDCG@5 标准上要比 ctBERT 高出 4% 和 3.9%.

表 3 原始模型与加入单样本学习后的实验结果对比 (%)

方法	Aminer		Patent		RARD	
	NDCG	NDCG	NDCG	NDCG	NDCG	NDCG
	@3	@5	@3	@5	@3	@5
ctBERT	78.2	83.5	60.1	63.2	36.6	42.6
ctBERT-os	81.6	86.7	63.8	67.0	41.2	47.3

4 结论

本文研究了个性化文章推荐问题, 在前人研究的

基础上提出了一个新的模型来解决它, 相较于其他的特征工程语义相似性模型, 创新性的加入主题信息, 将主题信息与预先训练好的上下文表示 (BERT) 相结合作为文本表示. 实验结果表明, 该模型优于传统的和最新的神经网络模型基线, 尤其是在特定专业领域 (如计算机领域) 的数据集上能显著的提高性能.

本文提出的模型虽然有一定的效果提升, 但也有些局限, 比如目前主题信息与文本表示的集合是在两个模型外部进行的融合, 对于模型的一致性而言还不够高, 未来的工作方向可能集中在如何不破坏预训练信息的情况下直接将主题信息融合到 BERT 内部中, 从而指导文本语义相似度的计算.

由于本文中用到的数据集均没有用户个性化标签, 在一定程度上会影响单样本学习框架的性能表现, 未来的工作方向会将 ctBERT 部署到线上环境, 收集到足够的用户反馈数据, 再根据训练结果调整模型和超参数, 从而更好地指导模型性能.

参考文献

- 周兴美. 一种论文检索系统: 中国, 107784100A. 2018-03-09.
- Pazzani MJ, Billsus D. Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W., eds. The Adaptive Web. Berlin: Springer, 2007. 325-341.
- Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. arXiv: 1301.7363, 2013.
- Pang L, Lan YY, Guo JF, et al. Text matching as image recognition. Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016. 2793-2799
- Wan SX, Lan YY, Guo JF, et al. A deep architecture for semantic matching with multiple positional sentence representations. Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016. 2835-2841.
- Blei DM, Lafferty JD. A correlated topic model of science. The Annals of Applied Statistics, 2007, 1(1): 17-35.
- Devlin J, Chang MW, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2018. 4171-4186.
- Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet

- Computing, 2003, 7(1): 76–80.
- 9 Li L, Wang DD, Li T, *et al.* SCENE: A scalable two-stage personalized news recommendation system. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. Beijing: ACM, 2011. 125–134.
 - 10 Du ZX, Tang J, Ding YH. POLAR: Attention-based CNN for one-shot personalized article recommendation. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Dublin: Springer, 2018. 675–690.
 - 11 Robertson S, Walker S, Jones S, *et al.* Okapi at TREC-3. Overview of the 3rd Text REtrieval Conference (TREC-3). Gaithersburg: NIST, 1995. 109–126.
 - 12 Salton G, Fox EA, Wu H. Extended Boolean information retrieval. Communications of the ACM, 1983, 26(11): 1022–1036. [doi: [10.1145/182.358466](https://doi.org/10.1145/182.358466)]
 - 13 Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993–1022.
 - 14 Vinyals O, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: ACM, 2016. 29: 3637–3645.
 - 15 Yang L, Zhang MY, Li C, *et al.* Beyond 512 tokens: Siamese multi-depth transformer-based hierarchical encoder for long-form document matching. Proceedings of the 29th ACM International Conference on Information & Knowledge Management. ACM, 2020. 1725–1734.
 - 16 Bianchi F, Terragni S, Hovy D. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. ACL, 2020. 759–766.
 - 17 Tang J, Zhang J, Yao LM, *et al.* ArnetMiner: Extraction and mining of academic social networks. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas: ACM, 2008. 990–998.
 - 18 Le Q, Mikolov T. Distributed representations of sentences and documents. Proceedings of the 31st International Conference on International Conference on Machine Learning. Beijing: ACM, 2014. II-1188–II-1196.
 - 19 Guo JF, Fan YX, Ai QY, *et al.* A deep relevance matching model for Ad-hoc retrieval. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. Indianapolis: ACM, 2016. 55–64.
 - 20 刘龙, 刘新, 蔡林杰, 等. 长文本匹配 LTM-B 模型. 计算机系统应用, 2022, 31(2): 291–297. [doi: [10.15888/j.cnki.csa.008313](https://doi.org/10.15888/j.cnki.csa.008313)]

(校对责编: 牛欣悦)