

融合词、句层级信息的抽取式摘要优化框架

林心宜¹ 严睿^{2,†} 赵东岩²

1. 北京大学信息科学技术学院, 北京 100871; 2. 北京大学计算机科学技术研究所, 北京 100080;

† 通讯作者, E-mail: ruiyan@pku.edu.cn

摘要 在抽取式摘要生成工作中, 单词和句子是两个最常用于刻画文档信息的粒度。然而, 由于词、句分别源于不同的语义空间, 难以用统一的目标函数刻画其数学关系, 所以即使词、句信息能够互为补充, 至今仍旧很少有使用非监督学习方法的工作将两个层级的信息进行综合考虑。本文推出一个混合的抽取式摘要优化框架: 在优化单词层级信息的同时, 将句子层级信息作为优化约束。在约束条件下, 该优化框架将迭代地进行摘要文本中单元的替换, 以求不停地逼近目标函数的最优解。对比其他传统方法, 该框架在 DUC 数据集上获得 ROUGE 评测的高分, 证明了该框架的有效性。

关键词 抽取式摘要生成工作; 词层级信息; 句层级信息; 混合迭代优化框架;

A Hybrid Optimization Framework Fusing Word- and Sentence-Level Information for Extractive Summarization

LIN Xinyi¹, YAN Rui^{2,†}, ZHAO Dongyan²

1. School of Electronic Engineering and Computer Science, Peking University, Beijing 100871; 2. Institute of Computer Science and Technology, Peking University, Beijing, 100080;

†Corresponding Author, E-mail: ruiyan@pku.edu.cn

Abstract Word and sentence units are two granularities to characterize document information in automatic summarization. Yet, few unsupervised studies take both factors into account due to the difficulties in fusing word-level and sentence-level information from different semantic spaces. In this paper, we propose a hybrid optimization framework to optimize word-level information while simultaneously incorporate sentence-level information as constraints. The optimization is conducted by iterative unit substitutions. The performance on DUC benchmark datasets demonstrates the effectiveness of our proposed framework in terms of ROUGE evaluation.

Key words Extractive Summarization; Word-level information; Sentence-level information; Hybrid Optimization Framework;

1 引言

多文档摘要生成工作被认为是自然语言处理领域的核心任务之一, 十几年来吸引了众多的研究者投入到该任务中。其目标是利用计算机, 自动为一组文档生成一段限定字数内精简、浓缩的摘要, 辅助读者快速领悟文档集合的要点, 捕捉重要信息。

完成这个任务的一个很自然的想法, 是将文档拆解成一组语义单元, 然后根据预设的标准, 从中抽取单元集合的子集进行重组, 形成摘要。而组成文档最常用的语义单元是句子和单词, 所以摘要工作通常也基于这两个粒度。之前大多数工作主要关注句子层级, 例如质心法 (centroid-based method) (e.g. MEAD^[1] 和 NeATS^[2]), 基于图排序法 (graph-based method) (e.g. PageRank^[3-4], HITS^[5], ClusterCMRW^[6]) 和子模方法 (Submodular functions^[7]); 基于单词级别的主要是整数线性规划法 (Integer Linear Programming process^[8])。Gillick^[8]将单词看作是赋有不同权重的语义单元, 将摘要工作抽象为一个整数线性规划问题,

在长度限制内抽取句子集合形成摘要，使之包含的单词的总权重最大；然而，研究者们很少将句子级别和单词级别的信息进行融合，主要是因为单词和句子来自于不同的语义空间，在大部分情况下，这两个空间非线性相关。Xie^[9]在融合句子和单词信息的时候，对句子和单词的分数分别进行了归一化，然后将两个分数进行线性叠加。尽管这个方法效果欠佳，但不失为融合词句信息的一个简单直接的解决思路。

本文推出了一个混合词层级、句层级信息的迭代优化摘要框架。之所以要融合词层级和句层级的信息，是因为词层级和句层级从不同角度满足了摘要的需求，并且功能存在互补。句子拥有完整的语法结构，其大部分字词组成的作用是连接，使之形成语法上可读的句子，而与语义内容无关；相比之下，单词粒度小，表达的含义更为凝练、精准，以单词为粒度能够更加充分的利用有限的摘要长度，包含更多的文档重要信息。然而，虽然单词体积小、以它为粒度能够提高摘要的有效信息率，但摘要还需考虑其可读性、连续性、逻辑完整性。简单的将一个个单词实体拼接起来无法形成可读的文段，相比之下，句子的优势在于，它是更为完整的语义单元，包含了必需的语法结构、语义连接、以及实体间的逻辑关系，保证了可读性，并且保留了一定的连续性、时间的顺序性、和逻辑的完整性。因此，更优的策略应该是，结合两个层级的优势，在抽取重要的句子的同时，兼顾使之包含尽可能多的重要的单词。

相比于直接将句子和单词的权重进行线性组合的简单方法，我们考虑到两个语义空间的数学关系不明确，为避免直接刻画句子层级和单词层级信息的关系，我们使用主任务和副任务的类似双任务的框架去实现句子层级和单词层级信息的融合。我们将任务抽象为带依赖关系的背包问题，用迭代框架对该优化问题进行求解。一方面，每个单词由于重要程度不同，有不同的权重得分。任务的主要目标是在长度限制内，迭代更新选择的句子集合，使之包含单词的总得分最高。另一方面，句子作为更高层次的语义单元，也有其句子层级的得分，其特征包括句子的重要性、句子组合的多样性、信息覆盖率等。这些特征的得分会作为迭代替换句子单元时的约束条件。因此，该框架在寻求单词层级信息的最优解的同时，保证了句子层级信息的不断优化，最终得到一个句子层级得分尽可能高的单词层级上的近似最优解。具体而言，我们通过最大化单词级别的分数获得一个初始化摘要，这个初始的摘要虽然是单词层级上的近似最优解，但是它在句子层级上的分数可能较低，所以我们在满足长度限制的约束下，迭代的进行句子集合的替换，使之满足句子层级评分的约束，并且保证该替换得到的摘要在单词级别上表现与最优解相差不超过一个阈值，最终获得一个单词级别和句子级别表现双高的平衡点。实验结果显示，该框架在标准数据集 DUC 上使用 ROUGE^[10]进行评测，获得了明显的效果提升。

综上，我们推出一个创新的混合词、句层级信息的带约束迭代优化框架，用于抽取式摘要生成工作。

2 方法

2.1 词层级优化

如果将单词看作是原子单元，句子则看作是由多个单词构成的有序元组。我们初步将任务刻画为求解单词层级上得分最大化的一个最优问题求解。我们的目标是选取句子集合构成摘要，使得摘要中的单词总分最高。这里需要注意的是，我们目前虽然以句子为单元进行抽取，但是只是保证了摘要的可读性，优化目标还是在于使得单词得分最大化，即，句子层面信息依旧未被考虑进来。

为了获取单词的得分，我们选用最经典的单词评分策略 tf-isf 方法为每个单词进行打分^[11]。tf 表示的是单词在句子中的词频，isf 表示逆向句子频率，是出现该单词的句子占有所有句子的比例的倒数再取对数值。给定一个多次出现的单词 w_i （在实验中，我们筛去出现次数小于 5 的单词，以省去不必要的处理），我们将其对应所有句子的 tf-isf 的分数进行求和，并取平均值作为该单词的得分，记做 $\pi(w_i)$ 。

通常，基于单词的摘要有一个基本的假设，即单词间是相互独立的，所以我们可以简单的将单词的得分进行直接相加，用来评价一个摘要在单词级别上的质量。

$$O(S) = \sum_{w_i \in S} \pi(w_i) \quad (1)$$

其中，S 是摘要，O(S)是该摘要在单词级别上的得分。为了避免冗余，摘要中重复出现的单词不进行重复的计分。我们的目标是，通过抽取包含尽可能多重要单词的句子，以达到最大化 O(S)的目的。我们通

过贪心法获得 $\text{argmax } O(S)$ 的近似解，将其作为下述迭代优化框架的初始摘要。

2.2 句层级优化

通过贪心法求得的近似解只是在单词层级上获得了近似最优，却没有考虑句子层级上的信息。句子的特征对于形成一篇逻辑关系完整合理、信息覆盖率高、冗余度低的摘要十分的关键^[7]。所以没有考虑句子层级信息的初始解显然无法更好地满足摘要的需求。

因此，为了提高摘要在句子层级上的表现，我们需要为摘要设定句子层级上优化的方向。

首先，我们考虑了句子层级上最突出的几个特征，包括信息覆盖度、信息冗余度、代表性。通过对这几个特征进行刻画，我们能够近似评价摘要在句子层级上的表现。

2.2.1 句子集合的信息覆盖度

我们引入了 Lin 和 Bilmes^[7]在 2010 年的一项工作中刻画句子集合信息覆盖度的公式。将文档集合看作一个图结构，图中的顶点表示文档集合中的句子，顶点间的边表示句子之间的相似度。我们使用如下公式来刻画摘要 S 和文档剩余的句子集合的相似度，表示摘要对原始文档集合的信息的覆盖程度，该指标表明摘要的内容是否全面。

$$C(S) = \sum_{s_j \in S} \min \{ \text{sim}(s_i, s_j), \sum_{s_k \in D \setminus S_i} \text{sim}(s_i, s_k) \} \quad (2)$$

其中， s_i 、 s_j 分别表示句子 i 和句子 j ， $\text{sim}(s_i, s_j)$ 表示句子 i 和句子 j 的相似度， $C(S)$ 表示摘要 S 的信息覆盖度。其值等于摘要对剩余句子覆盖程度的叠加。其中求取最小值是为了避免摘要对不重要的句子过分重视。两个句子的相似度可以用余弦相似度表示。我们把每个句子表示成一个向量，该向量的维数等于所有被计数的单词总数，向量的第 i 维是第 i 个单词在该句子中的 tf-isf 值。然后通过求解两个句子向量的余弦相似度来表示这两个句子的相似程度。

$$\text{sim}(s, s') = \frac{s \cdot s'}{|s| \times |s'|} \quad (3)$$

2.2.2 句子集合的信息多样性

由于自动摘要生成有长度限制，所以如何减少冗余信息，使得在限定长度内，摘要能尽可能表达更多的信息，提高摘要的有效信息率是自动摘要生成的重点研究问题。在此，我们使用如下的等式，定义一个句子集合的多样性：

$$D(S) = - \sum_{s_i \in S} \max_{s_j \in S \setminus s_i} \text{sim}(s_i, s_j) \quad (4)$$

我们叠加摘要 S 中句子间的相似性，取其负数作为摘要的多样性的评分。 $D(S)$ 的评分越高，说明该摘要 S 中包含的句子的相似度越低，进一步说明摘要 S 的冗余信息少，空间利用率提升。

2.2.3 句子集合的代表性

虽然在一开始将句子看作单词的有序集合，但句子并不是简单的单词的拼接，单词间的组合顺序赋予了句子高于单词层级的语义信息。所以，作为更高层级完整的语义单元，应当为句子对文档主旨的贡献力度有另外的评测。

PageRank^[3-4]方法广泛应用于多文档摘要工作对句子重要性的评价，到目前仍旧十分有效。而对句子重要性进行评估的图方法中，大多数也是 PageRank 方法的变种和演化。我们借用 PageRank 方法的思路，为给定的文档集合构建一副无向图，图的顶点是文档中的句子，顶点间的边是句子对的相似性。用 PageRank 算法处理该无向图，便可获得每个句子的重要性得分。对这些得分进行归一化，再对摘要 S 中包含的句子的重要性得分进行叠加，便获得了该摘要的总体代表性得分 $\varphi(s)$ 。

$$I(S) = \sum_{s \in S} \varphi(s) \quad (5)$$

该得分越高，说明摘要中包含的“高质量”句子越多，对文档的重要信息刻画的越好。

2.2.4 句子级别的总体得分

为了综合刻画摘要在句子级别上的表现，我们定义了一个公式，该公式通过线性组合的方式对句子集合在三个特征上的得分进行叠加，从而使用叠加获得的和作为对句子集合的综合打分（所有的项都进行了归一化）。

$$U(S) = \lambda_c \cdot C(S) + \lambda_i \cdot I(S) + \lambda_d \cdot D(S) \quad (6)$$

为简化说明，我们设置 $\lambda_c + \lambda_i + \lambda_d = 1$

2.3 约束条件下的迭代优化框架

考虑到词、句层级信息的非线性关系，我们避免了直接描述其数学关系的方法，而是采用了单词级别优化作为主优化目标、句子级别优化作为迭代求解主优化过程中约束条件的混合迭代优化框架。首先，我们通过贪心法获得单词级别优化的近似解，将该解作为迭代优化的初始值。由于该解既非单词级别上能获得的最优解，又非句子级别上的最优解，所以我们从该解起步，进行迭代的双方优化。在每一轮迭代过程中，我们根据句子层级的约束条件，进行摘要句子的更换，该更换一方面保证了句子层级上的优化，另一方面在目标函数的指导下，在多轮迭代中保证了单词级别的近似最优，通过这种方式，达到对词、句层级信息的融合，获得一个在词、句层级上表现双优的摘要。更具体地， $S^{(t-1)}$ 是第 $(t-1)$ 轮迭代结束时获得的摘要解； $D \setminus S^{(t-1)}$ 是第 $(t-1)$ 轮迭代结束时文档剩余句子的集合； $S^{(0)}$ 是由贪心法获得的初始化句子集合。在第 t 轮迭代过程中，我们会根据一定规则，从 $S^{(t-1)}$ 中选择一部分句子与 $D \setminus S^{(t-1)}$ 中的一部分句子进行交换，以获得 $S^{(t)}$ ，即该轮迭代对摘要解的更新。

替换句子组首先需要满足句子级别上的约束。我们将满足句子级别上约束条件的替换句子组放入一个替换候补集合中，该集合中的替换句子组都会带来 $S^{(t)}$ 在句子级别上表现的提升。

$$\Omega_t = \left\{ \langle X^{(t)}, Y^{(t)} \rangle \mid X^{(t)} \subseteq S^{(t-1)}, Y^{(t)} \subseteq D \setminus S^{(t-1)} \right\}$$

获得替换候补集合 Ω_t 后，我们从中选择一个替换句子组，在该候补集合中，选中的替换组获得的摘要在单词层级上是候补集合的所有替换中表现最优的，并且该替换仍旧使得摘要在单词集合上满足近似最优。换言之，我们每次迭代在保证单词级别表现的近似最优同时，还换来了句子级别上的提升。

迭代框架的公式如下：

$$\arg \max_{\langle X^{(t)}, Y^{(t)} \rangle} O((S^{(t-1)} - X^{(t)}) \cup Y^{(t)}) - O(S^{(t)}) \quad (7)$$

使得

$$(1) \quad L((S^{(t-1)} - X^{(t)}) \cup Y^{(t)}) < L$$

$$(2) \quad U((S^{(t-1)} - X^{(t)}) \cup Y^{(t)}) - U(S^{(t-1)}) > 0$$

$$(3) \quad O(S^{(\max)}) - O((S^{(t-1)} - X^{(t)}) \cup Y^{(t)}) < \varepsilon$$

约束条件 1 表示每一轮替换都必须满足摘要的长度限制。约束条件 2 表示每一轮替换都必须带来句子级别上的优化。我们记录了迭代过程中产生过的在单词级别上最优的解，记作 $S^{(\max)}$ 。在迭代过程中，我们允许摘要的单词层级表现相对于 $S^{(\max)}$ 的表现有一定阈值内的损失，一方面是因为优化的目标函数非凸，可能存在多个局部最优解，所以允许一定幅度内的临时下降是为了跳过部分的局部最优解，以增加达到全局最优解的可能性；另一方面，相对于最优解存在小幅度单词级别性能的损失对摘要在单词级别上的表现影响是微小的，然而这小部分的损失，能够换取摘要在句子级别上的性能的提升。该框架在多轮迭代后将收敛，并且最终获得在单词级别和句子级别上表现双优的摘要。

3 实验

3.1 实验环境设置

本实验使用数据集是自动摘要生成工作的标准数据集 DUC2003 和 DUC2004。DUC2003 和 DUC2004 包含 5 个任务，其中任务 2 是多文档摘要生成任务。DUC2003 任务 2 提供了 30 个文档集合，DUC2004 任务 2 提供了 50 个文档集合。DUC2003 与 DUC2004 的每个文档集合都包含 10 篇长度为 200-500 个词的文章，任务要求为每个文档集生成 100 词的摘要。本实验分别抽取 DUC2003 和 DUC2004 的 1/2 的文档集合作为训练集，剩余部分作为测试集合。

本实验搭建在 PKUSUMSUM^[12]平台上，该平台提供了多个非监督学习摘要方法的实现，而且这些方法都是目前表现最优的传统方法之一。这些方法中，一部分是基于句子级别的信息，比如 MEAD^[1]，ClusterCMRW^[6]，LexPageRank^[3]，和子模方法^[7]；一部分基于单词层级信息，比如 ILP^[8]。为了说明融合词句层级信息的优势，我们列出了这些方法在对应数据集上公布的最优结果，和本文框架的表现进行对比；另一方面，为了测试带约束的迭代优化框架的有效性，我们实现了同样是融合词句层级信息的线性组合算法^[9]（抽取句子来最大化句子和单词的线性加和评分）来与我们的框架进行对比。另外，为了实验的公平起见，我们对所有的算法使用相同的文本预处理步骤。

我们使用 ROUGE^[10]作为实验的评测标准。ROUGE 是 DUC 数据集官方规定的评测标准，也是摘要工作最常用的评价指标。ROUGE 通过计算生成摘要和标准摘要的单元的重合度来对摘要的质量进行打分。本次实验使用并公布了 ROUGE 评测结果中如下指标的 Recall 分数和 F-measure 分数：ROUGE1, ROUGE2, ROUGE-W-1.2 和 ROUGE-SU4。

表 1 于 DUC2003 数据集上多文档摘要任务的 ROUGE 评分
Table 1 ROUGE scores for multi-document summarization task on DUC2003 dataset

DUC2003	R scores				F scores			
	R-1	R-2	R-W-1.2	R-SU4	R-1	R-2	R-W-1.2	R-SU4
MEAD	0.3075	0.0695	0.1020	0.0848	0.3343	0.0759	0.1353	0.0990
ILP	0.3515	0.0840	0.1197	0.1115	0.3526	0.0844	0.1496	0.1124
LexPageRank	0.3268	0.0705	0.1078	0.0982	0.3582	0.0774	0.1439	0.1165
ClusterCMRW	0.3784	0.0947	0.1255	0.1263	0.3830	0.0958	0.1579	0.1291
Submodular	0.3717	0.0893	0.1226	0.1219	0.3765	0.0904	0.1543	0.1249
Linear Comb.	0.3540	0.0853	0.1200	0.1141	0.3552	0.0856	0.1500	0.1149
Our Model	0.3824	0.0952	0.1260	0.1299	0.3861	0.0960	0.1583	0.1323

表 2 于 DUC2004 数据集上多文档摘要任务的 ROUGE 评分
Table 2 ROUGE scores for multi-document summarization task on DUC2004 dataset

DUC2004	R scores				F scores			
	R-1	R-2	R-W-1.2	R-SU4	R-1	R-2	R-W-1.2	R-SU4
MEAD	0.3316	0.0783	0.1004	0.0979	0.3668	0.0876	0.1357	0.1268
ILP	0.3619	0.0743	0.1121	0.1154	0.3601	0.0743	0.1438	0.1185
LexPageRank	0.3830	0.0920	0.1128	0.1048	0.3607	0.0755	0.1407	0.1202
ClusterCMRW	0.3829	0.0899	0.1179	0.1326	0.3760	0.0908	0.1517	0.1308
Submodular	0.3935	0.0940	0.1190	0.1333	0.3890	0.0943	0.1527	0.1340
Linear Comb.	0.3599	0.0742	0.1122	0.1150	0.3607	0.0743	0.1439	0.1155
Our Model	0.3867	0.0946	0.1191	0.1347	0.3875	0.0948	0.1528	0.1352

*粗体表示在该列得分中该项分值最大

3.2 实验结果分析

在表格 1 和表格 2 中,我们展示了本文框架和其他框架在 DUC2003 和 DUC2004 数据集上分别的表现。结果可见,本文推出的框架在几乎所有的得分中都超过其他的方法。

首先对比本文方法(融合词句信息)与其他方法(只考虑其一)的实验结果可见,融合词句层级信息能够带来效果上的提升,这说明词句层级的信息对摘要的贡献互补,能够互相提高。其中子模方法(Submodular Function)是最具挑战难度的,这是因为它在对句子特征进行评分时,采用了更为复杂有效的方法,从而获得了对句子级别信息更加准确的评价。由于本文的重点在于研究融合词句信息的有效性、以及如何有效地融合词、句信息,对句子信息及单词信息分别的刻画都只采用了最直接的特征,因此分别在词、句的信息抽取性能上会略输专注抽取其一的方法。然而,在采用最简单直接的特征对词、句分别刻画的情况下,本文框架仍旧在绝大部分指标上超过了其他方法,这一点更加说明了融合词、句层级信息这一思路的有效性。

第二,我们对比了融合词、句信息的不同方法,即对比本文框架与线性组合框架的实验结果可知,迭代优化框架明显优于线性组合框架。这主要是因为单词的语义空间和句子的语义空间不是简单的线性关系,所以直接通过线性组合来融合词句信息是低效甚至有损性能的。更合理的方式是避开确切的词句信息的关系刻画,将其分别处理。本文将单词信息作为优化目标,句子信息作为优化约束,在迭代优化过程中达到了词、句层级都得到优化的效果。实验结果也显示了分开处理、迭代优化的框架在处理词句信息融合上的有效性。

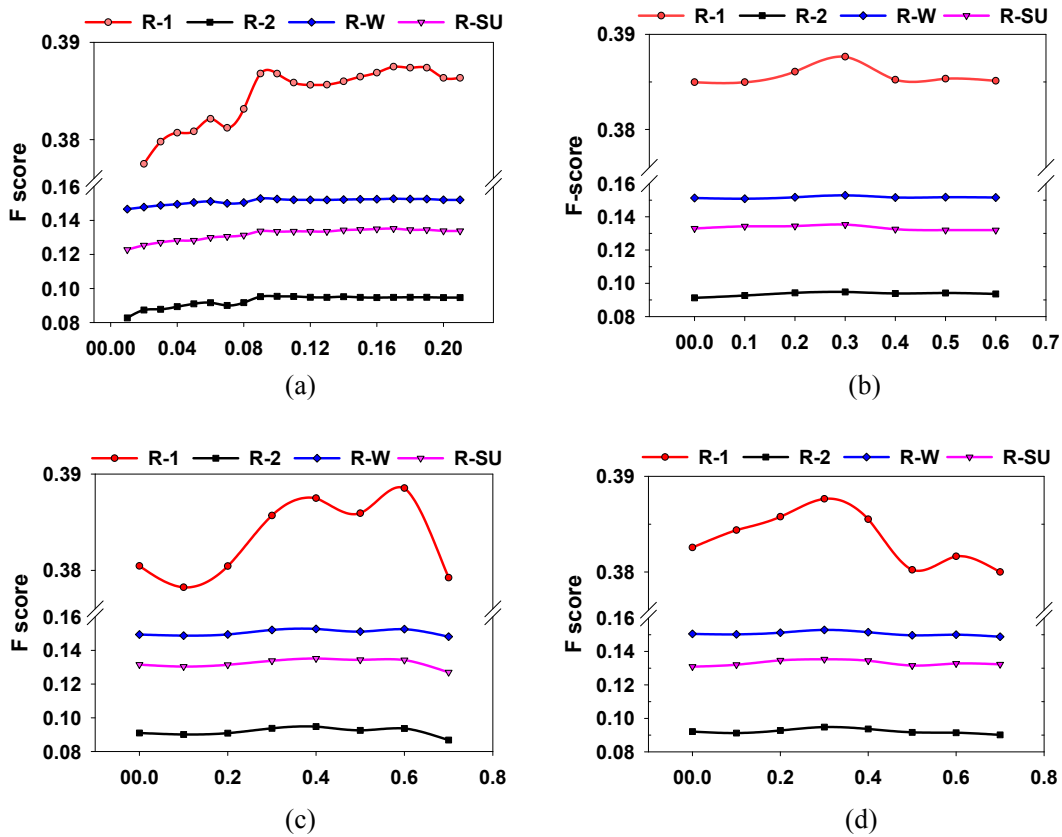


图 1 DUC2004 数据集上的调参结果
Fig.1 Parameter Tuning on DUC2004 dataset.

3.3 参数作用分析

分析参数对实验结果的影响,可以获悉每一项特征的作用。实验在调整参数的时候,每次只改变一个参数,其他参数保持固定,记录实验结果。参数调整的结果显示在图表 1 中。

第一组参数是约束条件 (2) 中的 $\lambda_c, \lambda_d, \lambda_l$, 其结果分别显示在图表 1.b, 1.c, 1.d 中。从实验曲线可知,

句子集合的多样性、代表性都对文档摘要的效果有较大的影响，相比之下，句子集合的覆盖度影响力度较小。从三个参数的走势可知，句子集合的信息覆盖度、多样性、代表性都是构成句子级别表现的重要因素，所以其参数过高或过低都会影响其他两个特征发挥作用，所以它们在一个中间点达到平衡。在实验中，我们设置 $\lambda_c = 0.3, \lambda_d = 0.4, \lambda_t = 0.3$ 。

另一个关键参数是约束 3 中的 ϵ ，其结果显示在图表 1.a 中。当 ϵ 增大时，表明迭代过程中对单词性能表现的临时性降低的容忍阈增大，该容忍程度增大有助于跳出局部最优点，并且为句子性能提升提供更大的空间。该参数在一定区间内显示出不稳定的升降幅度，当 ϵ 到达 0.17 之后，随着 ϵ 的增大，性能降低，这是因为该降低幅度过大损害了单词级别的表现。所以实验中，我们设置 $\epsilon = 0.17$ 。

4 总结

我们考虑了词、句层级信息对于摘要贡献的互补，推出了一个创新的融合词、句层级信息的抽取式摘要生成框架。该方法将单词信息作为优化目标，句子信息作为优化求解过程中的约束，以迭代替换的框架进行优化的求解，是一个新颖有效的融合词、句层级信息的模型。实验结果也显示了本文优化框架的优越性。

致谢 感谢 PKUSUMSUM 平台对本项目的帮助，该项目得到国家自然科学基金[No.61672058]，国家高技术研究发展计划[No.2015AA015403]，CCF-腾讯科研基金资助。

参考文献

- [1] Radev D, Allison T, et al. MEAD-A platform for multidocument multilingual text summarization. In Proceedings of the 4th International Conference on Language Resources and Evaluation. 2004: 699-702
- [2] Lin Chin-Yew, Hovy E. From single to multi-document summarization: A prototype system and its evaluation. In Proceedings of the 40th Annual Meeting of Association for Computational Linguistics. 2002: 457-464
- [3] Erkan G, Radev D. LexRank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research, 2004, 22: 457-479
- [4] Mihalcea R. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In Proceedings of the ACL Interactive Poster and Demonstration Sessions. 2004: 170-173
- [5] Wan Xiaojun. Document-based HITS model for multi-document summarization. In Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence. 2008: 454-465
- [6] Wan Xiaojun, Yang Jianwu. Multi-document summarization using cluster-based link analysis. In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2008: 299-306
- [7] Lin Hui, Bilmes J. Multi-document summarization via budgeted maximization of submodular functions. In Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2010: 912-920
- [8] Gillick D, Favre B. A scalable global model for summarization. In Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing. 2009: 10-18
- [9] Xie S S, Favre B, Hakkani-Tür D, et al. Leveraging sentence weights in a concept-based optimization framework for extractive meeting summarization. In Proceedings of the 10th Annual Conference of the International Speech Communication Association. 2009: 1503-1506
- [10] Lin Chin-Yew. ROUGE: A package for automatic evaluation of summaries. In Proceedings of Workshop on Text Summarization Branches Out, Post-Conference Workshop of the Annual Meeting of the Association for Computational Linguistics. 2004: 74-81
- [11] Neto J L, Santos A D, Kaestner C AA, et al. Generating text summaries through the relative importance of topics. 2000:

300-309

- [12] Zhang Jianmin, Wang Tianming, Wan Xiaojun. PKUSUMSUM: A Java platform for multilingual document summarization. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations. 2016: 287-291