OTExtSum: Extractive Text Summarisation with Optimal Transport

Peggy Tang[†], Kun Hu[†], Rui Yan[⋆], Lei Zhang[◇], Junbin Gao[‡], Zhiyong Wang[†]

[†]School of Computer Science, The University of Sydney
[⋆]Gaoling School of Artificial Intelligence, Renmin University of China
[◇]International Digital Economy Academy
[‡]The University of Sydney Business School, The University of Sydney
{peggy.tang,junbin.gao,zhiyong.wang}@sydney.edu.au,
kuhu6123@uni.sydney.edu.au,
ruiyan@ruc.edu.cn,leizhang@idea.edu.cn

Abstract

Extractive text summarisation aims to select salient sentences from a document to form a short yet informative summary. While learning-based methods have achieved promising results, they have several limitations, such as dependence on expensive training and lack of interpretability. Therefore, in this paper, we propose a novel non-learning-based method by for the first time formulating text summarisation as an Optimal Transport (OT) problem, namely Optimal Transport Extractive Summariser (OTExtSum). Optimal sentence extraction is conceptualised as obtaining an optimal summary that minimises the transportation cost to a given document regarding their semantic distributions. Such a cost is defined by the Wasserstein distance and used to measure the summary's semantic coverage of the original document. Comprehensive experiments on four challenging and widely used datasets - MultiNews, PubMed, BillSum, and CNN/DM demonstrate that our proposed method outperforms the state-of-the-art non-learning-based methods and several recent learning-based methods in terms of the ROUGE metric. [1]
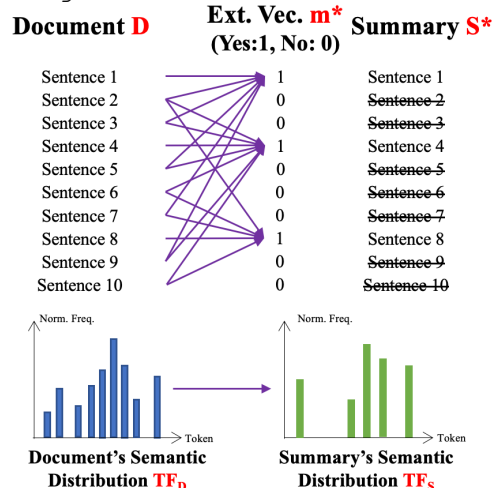
Figure 1: Illustration of Optimal Transport Extractive Summariser (OTExtSum): the formulation of extractive summarisation as an optimal transport (OT) problem. Optimal sentence extraction is conceptualised as obtaining the optimal extraction vector $\mathbf{m}^*$, which achieves an OT plan from a document $\mathbf{D}$ to its optimal summary $\mathbf{S}^*$ that has the minimum transportation cost. Such a cost is defined as the Wasserstein distance between the document's semantic distribution $\mathrm{TF_D}$ and the summary's semantic distribution $\mathrm{TF_S}$ and is used to measure the summary's semantic coverage.

1 Introduction

Text summarisation aims to condense a given document into a short and succinct summary that best covers the semantics of the document with the least redundancy. It helps users quickly browse and understand long documents by focusing on their most important sections (Mani, 2001; Nenkova and McKeown, 2011). A common practice for text summarisation is extractive summarisation which aims to select the salient sentences of a given document to form its summary. Extractive summarisation ensures the production of grammatically and factually

---

[1]Our code is publicly available for research purpose in https://github.com/peggypytang/OTExtSum/

correct summaries, though the output summaries could be inflexible. Since abstractive summaries are highly prone to contain contents that are unfaithful and nonfactual to the original document (Maynez et al., 2020), extractive summaries are more practical for real-world scenarios, especially for the domains requiring formal writing such as legal, science, and journalism documents.

Existing methods (Yao et al., 2017) often first score the importance of individual sentences of a given document and then combine the top-ranked ones to form a summary. However, the sentences with high importance scores may not well represent the document from a global perspective, which results in a sub-optimal summary. Recently, learning-

based methods, especially those based on supervised and unsupervised deep learning techniques (Narayan et al., 2018; Zheng and Lapata, 2019; Zhang et al., 2019; Narayan et al., 2020; Xu et al., 2020; Zhong et al., 2020; Padmakumar and He, 2021) can significantly improve summarisation performance. However, training deep learning models is computationally expensive, and it can be difficult to apply those models learned from a particular domain to other domains with different distributions. Moreover, deep learning methods generally lack interpretability for the summarisation process.

Motivated by these issues, we propose a novel non-learning based extractive summarisation method, namely Optimal Transport Extractive Summariser (OTExtSum). As illustrated in Figure 1, we formulate extractive summarisation based on the optimal transport (OT) theory (Peyré et al., 2019). A candidate summary can be evaluated by an OT plan regarding the optimal cost to transport between the semantic distributions of the summary and its original document. Then a Wasserstein distance can be obtained with this optimal plan to measure the discrepancy between the two distributions. To this end, it can be expected that a summary of high quality minimizes this Wasserstein distance. Moreover, a common assumption in the formulations of the OT problem is that the source and target distributions are fixed. In OTExtSum problem formulation, we relax this assumption by adding an extraction vector $\mathbf{m}^*$ to indicate which document sentences would be extracted to form the summary's semantic distribution, thus making the target distribution variable.

The semantic distributions of a given document and its candidate summary can be formulated in line with the frequency of their tokens. Inspired by Word Mover's Distance (Kusner et al., 2015), summarisation can be conceptualized as moving the "semantics" of a given document to its summary, and the ideal summary is obtained at the minimal transportation cost. This ensures the highest semantic coverage of the given document and the least redundancy in the summary without explicitly modelling conventional criteria such as relevance and redundancy. Thus, under the OT plan, the Wasserstein distance indicates the candidate summary's semantic coverage of the given document.

We design two optimisation strategies to approximate the extraction vector $\mathbf{m}^*$, namely beam search strategy (Tillmann and Ney, 2003), which itera-

tively evaluates the semantic coverage scores of a set of candidate summaries to obtain the optimal extraction, and binary integer programming strategy, which approximates the optimal extraction given the constraints of the Wasserstein distance and extraction budget. As a non-learning based method, OTExtSum does not require any training and is applicable to different document domains. Furthermore, it provides explainable results in terms of the semantic coverage of the summary.

There have been some studies on OT in NLP, such as document distance (Kusner et al., 2015; Yurochkin et al., 2019), text generation (Chen et al., 2018), text matching (Swanson et al., 2020), and machine translation (Xu et al., 2021). These methods generally focus on deriving similarities between words, sentences, and documents. On the contrary, we for the first time formulate text summarisation as an OT problem that optimally transports the semantic distributions between two texts (e.g., source document and summary candidate).

Overall, the key contributions of this paper are:

- We propose a non-learning based extractive summarisation method - OTExtSum by treating the text summarisation task as an optimal transport problem for the first time.

- We design two optimisation strategies for OTExtSum: beam search strategy and binary integer programming strategy.

- We present an interpretable visualisation of the semantic coverage of a generated summary by visualising the transport plan between summary tokens and document tokens.

- Comprehensive experimental results on four widely used datasets, including CNN/DM, MultiNews, BillSum and PubMed, demonstrate that OTExtSum outperforms the state-of-the-art non-learning based methods.

2 Related Work

Generally, text summarisation methods can be categorized as extractive, abstractive, and hybrid ones. While abstractive and hybrid summarisation methods (Lebanoff et al., 2019; Zhang et al., 2020) aim to mimic human beings in summarisation by paraphrasing a given document, extractive summarisation generally produces more factual summaries. In this section, we review existing extractive sum-

marisation methods in two categories: non-learning based and learning-based methods.

2.1 Non-learning based Methods

Most of the non-learning based methods conceptualise text summarisation as a sentence ranking task. Each sentence in a given document is scored in terms of various sentence importance criteria, which measure how well the sentence could represent the document. The top-ranked sentences are combined to form a summary. These methods often heavily rely on handcrafted features in regards to linguistic knowledge by focusing on local and/or global contexts.

Local Context based Methods. Local context-based methods rank a sentence based on the features obtained from the sentence itself. Sentence features such as frequency-based and topic-based were studied. Frequency-based features (Edmundson, 1969; Hovy and Lin, 1998) assume that the occurrence of high-frequency terms in a sentence is associated with their importance. Topic-based features (Kupiec et al., 1995; Nobata and Sekine, 2004; Lin and Hovy, 2000) assume that the density of a set of topic terms is highly correlated to the topic of a document.

Global Context based Methods. As local context features could overlook the correlations between sentences and lead to redundant summaries involving similar sentences, global context-based methods rank individual sentences from the perspective of the entire document. Discourse-based methods (Marcu, 1999) construct a document's rhetorical structure and extract the sentences on the longest chain of the semantic structure, i.e. the main topic. Centroid-based methods (Radev et al., 2000) cluster the sentences of a document through similarity measures and rank the sentences based on their distances to the cluster centroids. TextRank (Mihalcea and Tarau, 2004), as a graph-based method, is the state-of-the-art non-learning based method. A graph among document sentences is first formed by connecting sentences using sentence similarity scores, then the sentence connectivity can be used to score the importance of a sentence. Nonetheless, the nature of these sentence based scoring methods could miss summary-level or document-level patterns.

2.2 Learning-based Methods

Instead of utilising handcrafted features, due to the great success of deep learning in many natural language processing tasks, recent studies on extractive summarisation aim to learn sentence features from the corpus in a data-driven manner.

Supervised Methods. Most of these methods follow the sentence ranking conceptualisation, and an encoder-decoder scheme is generally adopted (Nallapati et al., 2017; Zhang et al., 2019; Narayan et al., 2020; Xu et al., 2020). An encoder formulates document or sentence representations, and a decoder predicts a sequence of sentence importance scores with the supervision of ground-truth sentence labels.

Reinforcement Learning based Methods. Reinforcement learning (RL) can be utilised for extractive summarisation by directly optimising the ROUGE metric, which is used as the training reward. The RL based summarisation task can be treated as a sentence ranking problem similar to the aforementioned methods (Narayan et al., 2018) or as a contextual-bandit problem (Luo et al., 2019).

Unsupervised Methods Various unsupervised methods have also been proposed to leverage pre-trained language models to compute sentence similarities and select important sentences. Some methods (Zheng and Lapata, 2019) use these similarities to construct a sentence graph and select sentences based on their centrality. Some methods (Padmakumar and He, 2021) use these to score relevance and redundancy of sentences as selection criteria.

Although these learning-based methods have significantly improved summarisation performance, computationally expensive training costs are inevitable, and it is challenging to generalise the trained models to documents from other domains that have distributions different from the training dataset. In addition, it is difficult to explain the correspondence and the coverage between a summary and a source document using these deep models. Therefore, to address these limitations, we revisit the non-learning based approach and propose a novel summarisation method by exploring the optimal transport theory for the first time.

3 Methodology

As shown in Figure 1, OTExtSum utilizes a text OT approximation to obtain the optimal extraction vector $\mathbf{m}^* = [m_1, ..., m_n]^T$, where $m_i \in \{0, 1\}$ denotes whether the i-th sentence is to be extracted (denoted by 1) or not (denoted by 0). The optimal extraction vector $\mathbf{m}^*$ achieves an OT plan from the semantic distribution of the document to that

of its optimal candidate summary which has the minimum total transportation cost.

The OT approximation consists of four components: 1) a tokeniser & embedding procedure that formulates token level representations and a semantic distribution estimation that computes the frequency of each token within a summary or a document ; 2) a transportation cost matrix that measures the cost using one token to represent another based on their Euclidean distances; 3) an OT solver that approximates Wasserstein distance and semantic coverage of the candidate summaries; and 4) an optimisation strategy that obtains the optimal extraction vector by choosing the summary with the minimum Wasserstein distance, and thus with the highest semantic coverage of the source document.

3.1 Optimal Transport

Consider a transportation problem that transports goods from a collection of suppliers $\mathbf{D} = \{d_i | i = 1, ..., N\}$ to a collection of customers $\mathbf{S} = \{s_j | j = 1, ..., N\}$, where d_i and s_j indicate the supply quantity of the i-th supplier and the order quantity of the j-th customer, respectively. Note that, in this study, we consider the number of suppliers to be the same as the customers. By defining t_{ij} as the quantity transported from the i-th supplier to the j-th customer, a transport plan $\mathbf{T} = \{t_{ij}\} \in \mathbf{R}^{N \times N}$ can be obtained. Given a cost matrix $\mathbf{C} = \{c_{ij}\} \in \mathbf{R}^{N \times N}$, where c_{ij} is the cost to deliver a unit of goods from the i-th supplier to the j-th supplier, the cost of the transport plan $\mathbf{T}$ can be calculated. Particularly, an OT plan $\mathbf{T}^* = \{t_{i,j}^*\} \in \mathbf{R}^{N \times N}$ in pursuit of minimising the transportation cost can be obtained by solving the following optimisation problem:

$$\mathbf{T}^* = \underset{\mathbf{T}}{\operatorname{argmin}} \sum_{i,j=1}^{N} t_{ij} c_{ij},$$

$$\text{s.t.} \sum_{j=1}^{N} t_{ij} = d_i, \quad \forall i \in \{1, ..., N\},$$

$$\sum_{i=1}^{N} t_{ij} = s_j, \quad \forall j \in \{1, ..., N\},$$

$$t_{ij} \geq 0, \quad \forall i, j \in \{1, ..., N\},$$

where the first two constraints indicate the quantity requirements for both suppliers and customers and the last constraint proves a non-negative order quantity. Mathematically, this OT problem is to find a joint distribution $\mathbf{T}$ with respect to a cost $\mathbf{C}$,

of which the marginal distribution is $\mathbf{D}$ and $\mathbf{S}$. In particular, Wasserstein distance can be defined as:

$$d_W(\mathbf{D}, \mathbf{S} | \mathbf{C}) = \sum_{i,j} t_{i,j}^* c_{i,j}. \qquad (2)$$

It can be viewed as the distance between the two probability distributions $\mathbf{D}$ and $\mathbf{S}$, if they are normalized, in line with the cost $\mathbf{C}$.

3.2 Semantic Distribution

In the context of text summarisation, denote $\mathbf{D} = \{\mathbf{s}_1, ..., \mathbf{s}_n\}$ to represent a document, where $\mathbf{s}_i$ denotes the i-th sentence contained in the document. The sentence $\mathbf{s}_i$ has a semantic distribution $\text{TF}_i \in \mathbb{R}^N$ computed by the normalised bag-of-tokens with removal of stop-words:

$$\text{TF}_i = [TF_{i1}, ..., TF_{iN}]^T,$$
$$TF_{ij} = \frac{d_j}{\sum_{k=1}^{N} d_k}, \qquad (3)$$

where d_j indicates the count of the j-th token in a vocabulary of size N.

A document $\mathbf{D}$ has a semantic distribution $\text{TF}_\mathbf{D}$:

$$\text{TF}_\mathbf{D} = \frac{\text{TF}_1 + ... + \text{TF}_n}{n}. \qquad (4)$$

For a summary $\mathbf{S} \subset \mathbf{D}$ with its corresponding extraction vector $\mathbf{m}$, of which the i-th element m_i is an indicator ($m_i = 1$ if $\mathbf{s}_i \in \mathbf{S}$, $m_i = 0$ otherwise), it has a semantic distribution $\text{TF}_\mathbf{S}$:

$$\text{TF}_\mathbf{S} = \frac{m_1 \times \text{TF}_1 + ... + m_n \times \text{TF}_n}{m_1 + ... + m_n}. \qquad (5)$$

In our proposed method, a normalization step is introduced to approximate the semantic distributions of $\mathbf{D}$ and $\mathbf{S}$ with term frequency. Note that after the normalization $\text{TF}_\mathbf{D}$ and $\text{TF}_\mathbf{S}$ have an equal total good quantities of 1 and can be *completely transported from one to the other*. In addition, $\text{TF}_\mathbf{D}$ and $\text{TF}_\mathbf{S}$ satisfy the property of discrete probability distributions, of which the sum should be 1.

3.3 Transport Cost between Tokens

We define the unit transportation cost between two tokens by measuring their semantic similarity. Intuitively, the more semantically dissimilar a pair of tokens are, the higher the "transport cost" of transporting one token to another. Given a pre-trained tokeniser and token embedding model with N tokens, define $\mathbf{v}_i$ to represent the feature embedding

of the i-th token. The transport cost from the i-th token to the j-th token c_{ij} in $\mathbf{C}$ can be written as:

$$c_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|_2, \qquad (6)$$

which is based on the Euclidean distance. [1]

3.4 Semantic Coverage of Candidate Summaries

Intuitively, a good summary $\mathbf{S}$ is supposed to be close to the document $\mathbf{D}$ in terms of their semantic distributions. OTExtSum utilizes the Wasserstein distance to measure the distance between the two associated semantic distributions $\mathrm{TF_D}$ and $\mathrm{TF_S}$ with the OT cost. The computation of the Wasserstein distance has time complexity of $O(p^3\log(p))$ (Altschuler et al., 2017), where p denotes the number of unique words in the document.

In detail, it can be obtained with Eq. (2) as $d_W(\mathrm{TF_D}, \mathrm{TF_S}|\mathbf{C})$ with a pre-defined cost matrix $\mathbf{C}$. Then a semantic coverage score of the summary $\mathbf{S}$ in respect to the document $\mathbf{D}$ can be further defined based on the Wasserstein distance:

$$g(\mathbf{D}, \mathbf{S}) = 1 - d_W(\mathrm{TF_D}, \mathrm{TF_S}|\mathbf{C}). \qquad (7)$$

Therefore, OTExtSum aims to search for an extraction vector $\mathbf{m}$, of which the corresponding summary $\mathbf{S}$ minimises the Wasserstein distance, i.e. maximising the semantic coverage score for the given document $\mathbf{D}$ by solving OT problems.

3.5 Optimisation Strategy

The remaining problem for OTExtSum is to search for the optimal extraction vector $\mathbf{m}^*$ which achieves the minimum total transportation cost from the semantic distribution of the document $\mathrm{TF_D}$ to that of the optimal summary $\mathrm{TF_S}$, given a budget B which is the number of sentences can be extracted to create a summary:

$$\mathbf{m}^* = \underset{\mathbf{m}}{\arg\min}\, d_W(\mathrm{TF_D}, \mathrm{TF_S}|\mathbf{C}), \\ \text{s.t. } m_1 + ... + m_n \leq B. \qquad (8)$$

In search of optimal extraction vector $\mathbf{m}^*$, we design two optimisation strategies, namely beam search strategy to achieve better coverage approximation, and binary integer programming strategy to achieve better computational efficiency.

---

[1] We investigated the effect of different distance measurements. As discussed in Section 4.3, cost matrix based on the Euclidean distance and the cosine distance yield similar ROUGE scores.

Algorithm 1: Optimisation of OTExtSum with Beam Search Strategy

Input : $\mathbf{D}$ the document, B the budget of the number of extracted sentences, K the beam width.
Output: $\mathbf{S}^*$ the optimal extractive summary.

1 Compute the cost matrix $\mathbf{C}$, and the document's semantic distribution $\mathrm{TF_D}$;
2 Initialise $\mathbf{m} = \mathbf{0}$, i.e. the candidate summary set $\mathbb{S} = \varnothing$;
3 **while** *# of sentences in candidate summary $\leq B$;* **do** // Beam search
4   **for** $k = 1, ..., |\mathbb{S}|$ **do**
5   *Generate the successor set $\mathbb{S}_b^k$ for $\mathbf{S}^k \in \mathbb{S}$;*
6   **end**
7      $\mathbb{S} \leftarrow \bigcup_k \mathbb{S}_b^k$;
8   **for** $k = 1, ..., |\mathbb{S}|$ **do**
9   *Compute the semantic distribution $\mathrm{TF}_{\mathbf{S}^k}$ of $\mathbf{S}^k \in \mathbb{S}$;*
10   *Compute the Wasserstein distance $d_W(\mathrm{TF_D}, \mathrm{TF}_{\mathbf{S}^k}|\mathbf{C})$ and the semantic coverage $g(\mathrm{TF_D}, \mathrm{TF}_{\mathbf{S}^k}|\mathbf{C})$;*
11   **end**
12   Keep the top K candidate summaries with the highest $g(\mathrm{TF_D}, \mathrm{TF}_{\mathbf{S}^k}|\mathbf{C})$ and prune the rest in $\mathbb{S}$;
13 **end**
14 $\mathbf{S}^* = \underset{\mathbf{S}^k \in \mathbb{S}}{\arg\max}\, g(\mathrm{TF_D}, \mathrm{TF}_{\mathbf{S}^k}|\mathbf{C})$;

3.5.1 Beam Search Strategy

The Beam Search (BS) strategy with the beam width K maintains the candidate summary set $\mathbb{S}$ and searches for the optimal extraction vector $\mathbf{m}^*$, thus the optimal extractive summary $\mathbf{S}^*$. Algorithm 1 presents the steps to obtain the optimal summary with OTExtSum using the BS strategy. The time complexity is $O(BKn(p^3\log(p)))$.

Initially, we have $\mathbf{m} = \mathbf{0}$, where none of the sentences are extracted. Then, each sentence in the document $\mathbf{D}$ is selected as a candidate summary, which derives a set of candidate extraction vectors corresponding to a set of candidate summaries, and its semantic coverage score can be evaluated. The top K candidate summaries in terms of the semantic coverage are kept in the set $\mathbb{S}$ and the rest are pruned. During the b-th iteration of the beam search, by appending each possible sentence

to an existing candidate summary $\mathbf{S}^k \in \mathbb{S}$, where the sentence is not in $\mathbf{S}^k$, a set of new candidate summaries $\mathbb{S}_b^k$ can be obtained. Then $\mathbb{S}$ is updated by combining all these sets of new candidate summaries in regards to k:

$$\mathbb{S} \leftarrow \bigcup_k \mathbb{S}_b^k. \qquad (9)$$

At the end of beam search, a set of final K summary candidates within the budget B is obtained.

Among the K final candidates from the beam search, OTExtSum obtains the optimal extraction vector and thus the optimal summary by choosing the candidate with the highest semantic coverage of the document $\mathbf{D}$.

Algorithm 2: Optimisation of OTExtSum with Binary Integer Programming Strategy

Input : $\mathbf{D}$ the document, B the budget of the number of extracted sentences, T the number of iterations.

Output : $\mathbf{S}^*$ the optimal extractive summary.

1 Compute the cost matrix $\mathbf{C}$, Compute document's semantic distribution $\text{TF}_{\mathbf{D}}$;

2 Initialise $\mathbf{w} \in \mathbb{R}^n$;

3 **for** *iteration* $t \in [1, ..., T]$ **do**

4   *Convert* $\mathbf{w}$ *to probability value* $\mathbf{pr}$ *with Sigmoid function;*

5   *Convert* $\mathbf{pr}$ *to* $\mathbf{b} = [b_i, .., b_n]$ *by hard sampling from the Gumbel-Softmax distribution;*

6   *Construct summary's semantic distribution* $\text{TF}_{\mathbf{S}}$;

7   *Compute the Wasserstein distance* $d_W(TF_{\mathbf{D}}, TF_{\mathbf{S}}|\mathbf{C})$;

8   *Compute the* L_1 *regularisation of* $\mathbf{b}$;

9   *Compute loss by weighted sum of the Wasserstein distance and the squared difference of* B *and* $\mathbf{b}$;

10   *Compute gradients and update* $\mathbf{w}$;

11 **end**

12 Compute $\mathbf{m}^*$ by soft sampling Sigmoid($\mathbf{w}$) from the Gumbel-Softmax distribution;

13 Obtain $\mathbf{S}^*$ by extracting top-B sentences with the highest m_i values for $i = 1, ..., n$;

3.5.2 Binary Integer Programming Strategy

Some prior works showed that integer linear programming is an efficient solution to summarisation problem (McDonald, 2007; Gillick and Favre,

2009).The Binary Integer Programming (BIP) strategy therefore is utilised to search for the optimal extraction vector $\mathbf{m}^*$ with T iterations. Based on the extraction vector, we obtain the optimal extractive summary $\mathbf{S}^*$. Algorithm 2 presents the optimisation steps to obtain the optimal summary with OTExtSum using the BIP strategy. The time complexity is $O(T(p^3 \log(p)))$.

As $\mathbf{m}^*$ is a multi-hot vector and is not differentiable, to make the backpropagation work, we optimise a proxy continuous vector $\mathbf{w} \in \mathbb{R}^n$, which is differentiable. Then we hard sample from the Gumbel-Softmax distribution (Maddison et al., 2016) to discretise and compute a multi-hot vector $\mathbf{b}$ during the iterations, and soft sample to compute $\mathbf{m}^*$ at the end.

The BIP strategy optimises the following loss function w.r.t. $\mathbf{w}$, which is a weighted sum of the Wasserstein distance $d_W(\text{TF}_{\mathbf{D}}, \text{TF}_{\mathbf{S}})$ and the L_1 regularisation of $\mathbf{b}$ [2]:

$$d_W(\text{TF}_{\mathbf{D}}, \text{TF}_{\mathbf{S}}|\mathbf{C}) + \alpha |B - \sum_{i=1}^n b_i|, \qquad (10)$$

where α denotes the weight of L_1 regularisation.

4 Experimental Results and Discussions

4.1 Datasets

To validate the effectiveness of the proposed OTExtSum on the documents with various writing styles and its ability to achieve improved summarisation performance, we perform experiments on four widely used challenging datasets collected from different domains.

Dataset	Multi-News	BillSum	PubMed	CNN/DM
Domain	News	Law	Science	News
#Sent./Doc.	80	46	102	33
B	9	7	6	3
Test Set Size	5,622	3,269	6,658	11,490

Table 1: Overview of the datasets. #Sent./Doc. denotes the average number of sentences in the documents, B denotes the budget of number of extracted sentences.

CNN/DailyMail (CNN/DM) (Hermann et al., 2015) is the standard single-document datasets with manually-written summaries. *Multi-News* (Fabbri et al., 2019) is a multi-document dataset which summarises multiple news articles. We concatenate the multiple articles as a single input. *BillSum*

---

[2]We choose L_1 regularisation for sparsity (Ng, 2004).

Method	Multi-News			BillSum			PubMed			CNN/DM		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
LEAD	42.3	14.2	22.4	43.5	25.6	37.8	34.0	8.6	27.1	40.0	17.5	32.9
ORACLE	45.4	20.6	28.1	43.7	25.7	38.0	37.1	15.5	30.4	43.1	23.7	37.5
Non-learning based Methods												
LSA (Gong and Liu, 2001)	-	-	-	32.6	15.7	26.3	33.9	9.9	29.7	-	-	-
LexRank (Erkan and Radev, 2004)	38.3	12.7	13.2	-	-	-	39.2	**13.9**	**34.6**	-	-	-
TextRank (Mihalcea and Tarau, 2004)	38.4	13.1	13.5	34.4	17.8	27.8	-	-	-	<u>34.1</u>	**12.8**	22.5
OTExtSum-BIP (GPT2)	40.6	12.1	20.7	36.6	15.6	30.6	35.4	10.8	28.8	<u>34.1</u>	12.6	**28.1**
OTExtSum-BIP (BERT)	40.6	12.1	20.7	36.6	15.6	30.6	35.4	10.8	28.8	<u>34.1</u>	<u>12.6</u>	**28.1**
OTExtSum-BS (Word2Vec)	42.3	12.8	21.9	**40.1**	<u>19.4</u>	**34.3**	38.2	11.7	30.8	32.3	10.8	25.9
OTExtSum-BS (GPT2)	<u>42.4</u>	**14.2**	**23.2**	36.5	**19.7**	32.0	<u>39.7</u>	<u>13.8</u>	32.3	33.5	12.0	26.7
OTExtSum-BS (BERT)	**43.1**	<u>13.9</u>	<u>22.5</u>	<u>37.5</u>	**19.7**	<u>32.6</u>	**39.8**	13.6	32.3	**34.5**	**12.8**	<u>27.8</u>
Unsupervised Deep Learning based Methods												
PacSum (Zheng and Lapata, 2019)	43.2	14.3	28.5	-	-	-	-	-	-	40.3	17.6	24.9
PMI (Padmakumar and He, 2021)	40.5	13.2	19.8	-	-	-	37.8	13.4	29.9	36.7	14.5	23.3
Supervised Deep Learning based Method												
MatchSum (Zhong et al., 2020)	46.2	16.5	41.9	-	-	-	41.2	14.9	36.8	44.2	20.6	40.4
PEGASUS (Zhang et al., 2020)	47.5	18.7	24.9	57.3	40.2	45.8	45.1	19.6	27.4	44.2	21.5	41.1

Table 2: Comparisons between our OTExtSum and the state-of-the-art methods across different categories. The highest scores are **bold**, and the second highest ones are <u>underlined</u>.

(Kornilova and Eidelman, 2019) is a dataset for law document summarization, which contains long state bill documents. *PubMed* (Cohan et al., 2018) is a scientific article dataset that uses the abstract section as the ground-truth summary and the long body section as the document. Table 1 shows an overview of the four datasets. The dataset details are in Appendix A.

While *CNN/DM* contains shorter documents and summaries, the other three datasets are more challenging because they have more extended documents and summaries, thus have a higher chance to extract sentences containing redundant contents or having limited relevance to the document.

4.2 Implementation Details

In terms of the pre-trained token embedding model, we compare the static embedding model Word2Vec and the contextual embedding models BERT and GPT2. The details of hyperparameter settings and software used are in Appendix B and C.

Our OTExtSum is compared against LEAD (See et al., 2017), ORACLE (Nallapati et al., 2017), the state-of-the-art non-learning based methods and the recent unsupervised learning-based methods. LEAD and ORACLE are standard baselines in the summarisation task. LEAD baseline extracts the first several sentences of a document as a summary. ORACLE baseline greedily extracts the sentences that maximise the ROUGE-L score based on the reference summary. We compare with the results of strong non-learning-based methods, including LSA (Gong and Liu, 2001), TextRank (Mihalcea and Tarau, 2004), and LexRank (Erkan and Radev, 2004).

Their results on MultiNews, BillSum, PubMed, and CNN/DM are from (Fabbri et al., 2019), (Kornilova and Eidelman, 2019), (Cohan et al., 2018), and (Padmakumar and He, 2021) respectively. For an informative reference, we report recent unsupervised learning-based methods, including PacSum (Zheng and Lapata, 2019), which its released model was trained on the news domain, and PMI (Padmakumar and He, 2021), of which the released models were trained on the news and science domains. Their results on CNN/DM are from (Padmakumar and He, 2021). Their results on Multi-News, BillSum, and PubMed are evaluated on the datasets with the corresponding released models from the same domains. And we include the results of the state-of-the-art supervised learning-based methods with extractive approach MatchSum from (Zhong et al., 2020), and those with abstractive approach PEGASUS from (Zhang et al., 2020).

4.3 Quantitative Analysis

The commonly used ROUGE metric (Lin, 2004) is also adopted for our quantitative analysis. It evaluates the content consistency between the generated summary and the reference summary. In detail, ROUGE-n scores measure the number of overlapping n-grams between the generated summary and the reference summary. A ROUGE-L score considers the longest common subsequence between the generated summary and the reference summary.

Performance Overview. The experimental results of OTExtSum on the four datasets are listed in Table 2 in terms of ROUGE-1, ROUGE-2 and ROUGE-L F-scores. We observed that the BS strat-

egy could generally achieve better optimisation results than the BIP strategy. It is in line with our design understanding that beam search can better reach the global optimum. Whereas, the two strategies achieve similar results in CNN/DM, which could be because CNN/DM has fewer document sentences and lower budget, thus fewer possible solutions and easier to find the optimum.

OTExtSum outperforms the state-of-the-art non-learning based methods and is comparable to the learning-based methods. Note that the state-of-the-art methods usually optimise at the sentence level, whilst OTExtSum is based on the summary level OT evaluation, by which the quality of the resulting summaries is improved.

We observed that OTExtSum obtains significantly better ROUGE scores than the baseline methods on Multi-News, BillSum and PubMed, while the improvement is not that significant on CNN/DM . When the summary is more extended, such as these three more challenging datasets, the summary sentences are more likely to have redundant content. That is, even summary-level optimisation is more difficult to achieve, our OTExtSum demonstrates higher improvements.

OTExtSum is a non-learning based method, and training is not required. Unlike learning-based methods, it is not limited by the training data domain and can be used for different domains. Experimental results demonstrate generalisation ability of OTExtSum over news, law, and science domains.

Effects of Token Embeddings Models. OTExtSum is dependent on a pre-trained token embedding method. Specifically, the token embedding model affects the cost matrix C and the tokenisation, thus the frequency vector, of the document. We examine how different token embedding models would affect the performance of OTExtSum by comparing static embedding model Word2Vec, and contextual embedding models BERT and GPT2.

The results on most of the datasets indicate that a more advanced contextual embedding model such as BERT and GPT2 is more effective than a static embedding model Word2Vec. It is in line with the intuitive understanding that a more representative model with adequate training samples often approximates better token embeddings and representation. Despite that, the performance of OTExtSum with Word2Vec is surprisingly competitive.

Effects on Stop-words. We investigate the impact of stop-words on the performance of OTExtSum. As shown in Table 3 in Appendix E, the effect varies slightly across the datasets, and may not much influence the ROUGE scores. It could be because text summarisation does not generally depend on stop-words. A side benefit of removing the stop-words is reducing the vocabulary size and thus the computation time of OT.

Effects on Distance Measurement. We examine how the distance measurement of the cost matrix would impact the performance of OTExtSum. As shown in Table 3 in Appendix E, cost matrix based on the cosine distance and the Euclidean distance usually yield similar ROUGE scores.

4.4 Interpretable Visualisation

OTExtSum is able to provide an interpretable visualisation of the summarisation procedure. Figure 2 in Appendix D illustrates the transport plan heatmap, which indicates the transportation of semantic contents between tokens in the document and its resulting summary. The higher the intensity, the more the semantic content of a particular document token is covered by a summary token.

4.5 Qualitative Analysis

Figure 3 , 4 , 5, and 6 in Appendix F compare the summaries produced by OTExtSum and TextRank. TextRank extracted sentences that are salient on their own yet redundant when combined to form a summary. In comparison, OTExtSum is able to compose summaries that have higher semantic coverage and less redundant content.

5 Conclusion

In this paper, we have presented OTExtSum, the first optimal transport-based optimisation method for extractive text summarisation. It aims to identify an optimal subset of sentences for producing a summary that achieves high semantic coverage of the document by minimising the Wasserstein distance between the semantic distributions of the document and the summary. It helps obtain a summary from a global perspective and provides an interpretable visualisation of extraction results. In addition, OTExtSum does not require computationally expensive training. The comprehensive experiments demonstrate the effectiveness of OTExtSum, which is generalisable over various document domains. In our future work, we will explore other OT solvers for extractive summarisation.

References

Jason Altschuler, Jonathan Weed, and Philippe Rigollet. 2017. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *International Conference on Neural Information Processing Systems(NeurIPS)*.

Liqun Chen, Shuyang Dai, Chenyang Tao, Haichao Zhang, Zhe Gan, Dinghan Shen, Yizhe Zhang, Guoyin Wang, Ruiyi Zhang, and Lawrence Carin. 2018. Adversarial text generation via feature-mover's distance. In *International Conference on Neural Information Processing Systems (NeurIPS)*.

Arman Cohan, Franck Dernoncourt, Doo Soon Kim, Trung Bui, Seokhwan Kim, Walter Chang, and Nazli Goharian. 2018. A discourse-aware attention model for abstractive summarization of long documents. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

H. P. Edmundson. 1969. New methods in automatic extracting. *Journal of the ACM*, 16:264–285.

G. Erkan and D. R. Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*.

Alexander Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir Radev. 2019. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouve, and Gabriel Peyré. 2019. Interpolating between optimal transport and mmd using sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*.

Dan Gillick and Benoit Favre. 2009. A scalable global model for summarization. In *Workshop on Integer Linear Programming for Natural Language Processing*.

Yihong Gong and Xin Liu. 2001. Generic text summarization using relevance measure and latent semantic analysis. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*.

Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *International Conference on Neural Information Processing Systems (NeurIPS)*.

Eduard Hovy and Chin-Yew Lin. 1998. Automated text summarization and the Summarist system. In *TIPSTER Text Program Phase III*.

Anastassia Kornilova and Vladimir Eidelman. 2019. BillSum: A corpus for automatic summarization of US legislation. In *Workshop on New Frontiers in Summarization (NFiS)*.

Julian Kupiec, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. In *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International Conference on Machine Learning (ICML)*.

Logan Lebanoff, Kaiqiang Song, Franck Dernoncourt, Doo Soon Kim, Seokhwan Kim, Walter Chang, and Fei Liu. 2019. Scoring sentence singletons and pairs for abstractive summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*.

Chin-Yew Lin and Eduard Hovy. 2000. The automated acquisition of topic signatures for text summarization. In *International Conference on Computational Linguistics (COLING)*.

Ling Luo, Xiang Ao, Yan Song, Feiyang Pan, Min Yang, and Qing He. 2019. Reading like HER: Human reading inspired extractive summarization. In *Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Chris J. Maddison, Andriy Mnih, and Yee Whye Teh. 2016. The concrete distribution: A continuous relaxation of discrete random variables. In *International Conference on Learning Representations (ICLR)*.

Inderjeet Mani. 2001. *Automatic summarization*. John Benjamins Publishing.

Daniel Marcu. 1999. Discourse trees are good indicators of importance in text. In *Advances in Automatic Text Summarization*.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Ryan McDonald. 2007. A study of global inference algorithms in multi-document summarization. In *European Conference on IR Research*.

Clara Meister, Ryan Cotterell, and Tim Vieira. 2020. If beam search is the answer, what was the question? In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *AAAI Conference on Artificial Intelligence*.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Ranking sentences for extractive summarization with reinforcement learning. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Shashi Narayan, Joshua Maynez, Jakub Adamek, Daniele Pighin, Blaz Bratanic, and Ryan McDonald. 2020. Stepwise extractive summarization and planning with structured transformers. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ani Nenkova and Kathleen McKeown. 2011. Automatic summarization. *Foundations and Trends® in Information Retrieval*, 5:103–233.

Andrew Y Ng. 2004. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *International Conference on Machine Learning (ICML)*.

Chikashi Nobata and Satoshi Sekine. 2004. Crl/nyu summarization system. In *Document Understanding Conference (DUC)*.

Vishakh Padmakumar and He He. 2021. Unsupervised extractive summarization using pointwise mutual information. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Gabriel Peyré, Marco Cuturi, et al. 2019. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*.

Dragomir R. Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *Conference of the North American Chapter of the Association for Computational Linguistics - Applied Natural Language Processing Conference Workshop: Automatic Summarization (NAACL-ANLP-AutoSum)*.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. 2013. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning (ICML)*.

Kyle Swanson, Lili Yu, and Tao Lei. 2020. Rationalizing text matching: Learning sparse alignments via optimal transport. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Christoph Tillmann and Hermann Ney. 2003. Word reordering and a dynamic programming beam search algorithm for statistical machine translation. *Computational Linguistics*.

Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Discourse-aware neural extractive text summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jingjing Xu, Hao Zhou, Chun Gan, Zaixiang Zheng, and Lei Li. 2021. Vocabulary learning via optimal transport for neural machine translation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao. 2017. Recent advances in document summarization. *Knowledge and Information Systems*, 53(2):297–336.

Mikhail Yurochkin, Sebastian Claici, Edward Chien, Farzaneh Mirzazadeh, and Justin Solomon. 2019. Hierarchical optimal transport for document representation. In *International Conference on Neural Information Processing Systems (NeurIPS)*.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning (ICML)*.

Xingxing Zhang, Furu Wei, and Ming Zhou. 2019. HIBERT: Document level pre-training of hierarchical bidirectional transformers for document summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Hao Zheng and Mirella Lapata. 2019. Sentence centrality revisited for unsupervised summarization. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

Ming Zhong, Pengfei Liu, Yiran Chen, Danqing Wang, Xipeng Qiu, and Xuanjing Huang. 2020. Extractive summarization as text matching. In *Annual Meeting of the Association for Computational Linguistics (ACL)*.

A Dataset Details

We followed (Zhong et al., 2020) to set B for CNN/DM, PubMed and Multi-News, and used the average number of sentences in the summaries to set B for BillSum since this is a common practice in the literatures (Narayan et al., 2018). These datasets were obtained from a source, namely HuggingFace Datasets [3].

Since OTExtSum does not require training, for a fair comparison, all experimental results are reported on the test splits of the four datasets only.

B Hyperparameter Details

For the hyperparameter settings of the BIP strategy, the number of iteration T was set to 200, α was set to 1, and it used the SGD optimiser (Sutskever et al., 2013) with learning rate 0.1. For the BS strategy, the beam width K was set to 5 [4].

C Software and Hardware Used

We obtained the pre-trained Word2vec (Google News 300 dimension) from GENSIM [5], and the contextual embedding models BERT (base version) and GPT2 from HuggingFace [6]. To compute the Wasserstein distances, we adopted GENSIM, the POT [7] and GeomLoss (Feydy et al., 2019) libraries. List of stop-words was from NLTK library [8]. Our experiments were run on a GeForce GTX 1080 GPU card. We obtain our ROUGE scores by using the pyrouge package [9].

---

[3] https://huggingface.co/docs/datasets/
[4] We chose the beam width in line with a common practice in the literature (Meister et al., 2020)
[5] https://radimrehurek.com/gensim/index.html
[6] https://huggingface.co
[7] https://pythonot.github.io
[8] https://www.nltk.org
[9] https://pypi.org/project/pyrouge/

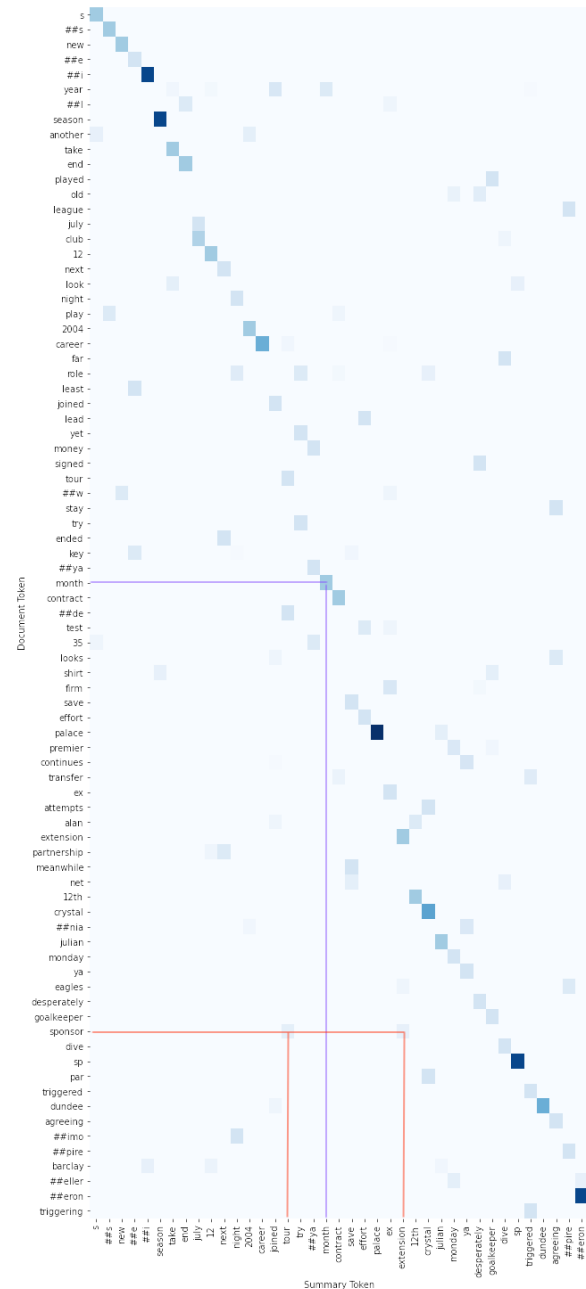## D Example of Interpretable Visualisation

Figure 2: Interpretable visualisation of the OT plan from a source document to a resulting summary on the CNN/DM dataset. The higher the intensity, the more the semantic content of a particular document token is covered by a summary token. Purple line highlights the transportation from the document to the summary of semantic content of token "month", which appears in both the document and the summary. Red line highlights how the semantic content of token "sponsor", which appears in the document only but not the summary, are transported to token "tour" and "extension", which are semantically closer and have lower transport cost, and thus achieve a minimum transportation cost in the OT plan.

E Ablation Studies

Method	Multi-News			BillSum			PubMed			CNN/DM		
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
Euc. \wo s.w.	43.1	13.9	22.5	37.5	**19.7**	32.6	39.8	13.6	32.2	**34.5**	**12.8**	**27.8**
Cos. \wo s.w.	43.1	13.9	22.5	**39.0**	19.5	**33.6**	39.8	13.6	32.3	**34.4**	**12.4**	**27.7**
Euc. \w s.w.	43.4	**14.4**	**23.4**	36.9	19.6	32.2	**40.6**	**13.8**	**33.0**	34.1	12.1	27.1
Cos. \w s.w.	**43.9**	14.2	23.1	38.1	19.6	33.0	**40.6**	13.6	32.9	34.1	12.1	27.1

Table 3: Ablation studies of OTExtSum based on the BS optimisation strategy and pre-trained BERT tokeniser. Euc. denotes the Euclidean distance and Cos. denotes the cosine distance. s.w. denotes stop-words.

F Generation Samples

Below are the generation samples of OTExtSum and TextRank. In general, OTExtSum based summary contains less redundant content and provides higher semantic coverage with the same number of extracted sentences.

Source Document:

WASHINGTON – (1) Sen. John McCain (R-Ariz.) said Wednesday night he will no longer support Richard Mourdock, unless the Indiana Senate candidate apologizes for his recent comments on rape. (...) "If he apologizes and says he misspoke and he was wrong and he asks the people to forgive him, then obviously I'd be the first. (...)McCain's call for an apology comes as Mitt Romney faces mounting pressure from Democrats to withdraw his endorsement of Mourdock and have a TV ad he starred in for the Indiana state treasurer removed from the airwaves.(...) "I think it depends on what he does," McCain told CNN's Anderson Cooper, when asked if he still counts himself among Mourdock's supporters. (...) I think that even when life begins in that horrible situation of rape, that it is something that God intended to happen." (2) Romney's campaign confirmed Wednesday that he continues to support Mourdock and has not asked that the ad be pulled. Romney campaign spokeswoman Andrea Saul did seek to distance the GOP nominee from the views Mourdock expressed." Gov. Romney disagrees with Richard Mourdock, and Mr. Mourdock's comments do not reflect Gov. Romney's views," Saul said in an emailed statement. "We disagree on the policy regarding exceptions for rape and incest but still support him." Mourdock declined to apologize during a press conference on Wednesday, saying instead that he regretted his words had been misinterpreted. UPDATE: Oct. 25 – McCain spokesman Brian Rogers on Thursday clarified (3) McCain's earlier comments that he would withdraw support for Mourdock if he did not apologize, saying that the Arizona senator still supports the Indiana Senate candidate. "Senator McCain was traveling yesterday in Florida and did not have an opportunity to see Mr. Mourdock's full press conference before he taped his CNN interview," Rogers said, referring to the Wednesday conference where Mourdock said he regretted that he had been misinterpreted. "Senator McCain is glad that Mr. Mourdock apologized to the people of Indiana and clarified his previous statement. Senator McCain hopes the people of Indiana will elect Mr. Mourdock to the U.S. Senate." President Obama on Wednesday criticized Senate candidate Richard Mourdock's controversial comments on rape, saying that the Indiana Republican was wrong when he called pregnancy resulting from rape "something God intended to happen." "I don't know how these guys come up with these ideas," Obama said during a taping of "The Tonight Show" in Burbank, Calif. "Let me make a very simple proposition. Rape is rape. It is a crime." "These various distinctions about rape don't make too much sense to me, don't make any sense to me," he said. (Also on POLITICO: Exclusive interview: Mourdock braces for fallout) Obama also laid out a contrast between his position on women's issues and Mitt Romney's, saying that "Roe vs. Wade is probably hanging in the balance" depending on the outcome of the presidential race and the winner's eventual Supreme Court appointments. Obama added that Mourdock's remarks are evidence of why government shouldn't make decisions on women's health. "This is exactly why you don't want a bunch of politicians, mostly male, making decisions about women's health care decisions. Women are capable of making these decisions in consultation with their partners, with their doctors," he said to loud applause. "And for politicians to want to intrude in this stuff, oftentimes without any information, is a huge problem. And this is obviously part of what's stake at this election." Mourdock's comment, made Tuesday, has mobilized the Obama campaign and Democrats to call on Mitt Romney to distance himself from the Senate candidate. (Also on POLITICO: GOP splits over Mourdock comment) (4) The Romney campaign distanced itself from Mourdock's statement on rape, but still supports the candidate and has not asked him to stop using an ad featuring Romney. "Gov. Romney disagrees with Richard Mourdock, and Mr. Mourdock's comments do not reflect Gov. Romney's views. We disagree on the policy regarding exceptions for rape and incest but still support him," press secretary Andrea Saul said. (...) Sen. John McCain (R-Ariz.) said his support for the Indiana Republican "depends on what he does." "If he apologizes and says he misspoke and he was wrong and asks the people to forgive him, I would be the first" to support him, McCain said on CNN's "Anderson Cooper 360." Earlier Wednesday, the Obama campaign released a web video juxtaposing clips from Romney's pro-Mourdock ad with Mourdock's remarks on rape. The video ends with message, "Mitt Romney: extremely conservative to this day." Read more about: Richard Mourdock

Reference Summary:

President Obama slammed Senate candidate Richard Mourdock's rape comments on the Tonight Show last night. "I don't know how these guys come up with these ideas," Obama said. "Let me make a very simple proposition. Rape is rape. It is a crime." He noted "Roe vs. Wade is probably hanging in the balance" of the election, Politico reports. Mourdock's comments that pregnancies from rape are "something God intended" are "exactly why you don't want a bunch of politicians, mostly male, making decisions about women's health care decisions," Obama said. "Women are capable of making these decisions in consultation with their partners, with their doctors." Meanwhile, asked by Anderson Cooper whether he still supports Mourdock, John McCain said it "depends on what he does," the Huffington Post reports. "If he apologizes and says he misspoke and he was wrong and he asks the people to forgive him, then obviously I'd be the first" to support him. Mourdock has in fact issued an apology. Mitt Romney, for his part, still supports Mourdock, though his campaign says he "disagrees" with the Indiana Republican's comments: "We disagree on the policy regarding exceptions for rape and incest but still support him." The campaign says it hasn't called on Mourdock to withdraw ads featuring Romney.

Figure 3: A sample summary comparison on the **Multi-News** dataset. OTExtSum based summary sentences are highlighted in yellow colour. TextRank based summary sentences are underlined in red colour. TextRank extracted redundant contents, specifically the part (1) is duplicated with the part (3), and the part (2) is duplicated with the part (4). The summary generated by OTExtSum has ROUGE-1 F-Score of 65.21 and Semantic Coverage Score of 0.93, while the summary generated by TextRank has ROUGE-1 F-Score of 44.87 and Semantic Coverage Score of 0.89. Semantic Coverage Score of the ground-truth summary is 0.89.

Source Document:

SECTION 1. SHORT TITLE; TABLE OF CONTENTS. (a) Short Title.–This Act may be cited as the "The Water Recycling and Riverside-Corona Feeder Act of 2006". (...) ① **Inland Empire and Cucamonga Valley recycling projects.** TITLE II–PROJECTS IN RIVERSIDE AND SAN BERNARDINO COUNTIES Sec. 201. Planning, design, and construction of the Riverside-Corona Feeder. Sec. 202. Project authorizations. TITLE I–THE INLAND EMPIRE REGIONAL WATER RECYCLING INITIATIVE SEC. 102. SHORT TITLE. This title may be cited as the "The Inland Empire Regional Water Recycling Initiative". SEC. 103. INLAND EMPIRE AND CUCAMONGA VALLEY RECYCLING PROJECTS. (a) Recycling Projects.–The Reclamation Wastewater and Groundwater Study and Facilities Act (Public Law 102-575, Title XVI; 43 U.S.C. 390h et seq.) is amended by adding at the end the following: "SEC. 1637. ② **INLAND EMPIRE REGIONAL WATER RECYCLING PROJECT.** "(...) ③ construction of **the Inland Empire regional water recycling project** (...) (d) Authorization of Appropriations.–There is authorized to be appropriated to carry out this section $20,000,000. "SEC. 1638. ④ **CUCAMONGA VALLEY WATER RECYCLING PROJECT.** (...) ⑤ Cucamonga Valley Water Recycling Project. (...) plan, design, and construct a water supply project, the Riverside-Corona Feeder, (...) the cost to plan, design, and construct the project described (...) (a) In General.–The Reclamation Wastewater and Groundwater Study and Facilities Act (Public Law 102-575, title XVI; 43 U.S.C. 390h et seq.) (...) (b) Cost Sharing.–The Federal share of the cost of the project described in subsection (a) shall not exceed 25 percent of the total cost of the project. (...)

Reference Summary:

Water Recycling and Riverside-Corona Feeder Act of 2006 - Inland Empire Regional Water Recycling Initiative - Amends the Reclamation Wastewater and Groundwater Study and Facilities Act to authorize the Secretary of the Interior: (1) in cooperation with the Inland Empire Utilities Agency, to participate in the design, planning, and construction (design) of the Inland Empire regional water recycling project, California; (2) in cooperation with the Cucamonga Valley Water District, to participate in the design of the Cucamonga Valley Water District satellite recycling plants in Rancho Cucamonga to reclaim and recycle approximately two million gallons per day of domestic wastewater; (3) in cooperation with the Yucaipa Valley Water District, to participate in the design of projects to treat impaired surface water, reclaim and reuse impaired groundwater, and provide brine disposal within the Santa Ana Watershed; and (4) in cooperation with the City of Corona Water Utility, to participate in the design of, and land acquisition for, a project to reclaim and reuse wastewater, including degraded groundwaters, within and outside of the City. Limits the federal cost share of each project to 25%. Authorizes the Secretary, in cooperation with the Western Municipal Water District, to participate in a project to design the Riverside-Corona Feeder, which includes 20 groundwater wells and 28 miles of pipeline in San Bernardino and Riverside Counties, California. Limits the federal share of the project design and planning study costs.

Figure 4: A sample summary comparison on the **BillSum** dataset. OTExtSum based summary sentences are highlighted in yellow colour. TextRank based summary sentences are underlined in red colour. TextRank extracted redundant contents, specifically the part ①, ②, ③, ④, and ⑤ are duplicated. The summary generated by OTExtSum has ROUGE-1 F-Score of 44.2 and Semantic Coverage Score of 0.92, while the summary generated by TextRank has ROUGE-1 F-Score of 33.2 and Semantic Coverage Score of 0.77. Semantic Coverage Score of the ground-truth summary is 0.84.

Source Document:

to report ① **the effective treatment of radiation macular edema following ruthenium-106 plaque brachytherapy for a choroidal melanoma with a dexamethasone 0.7-mg (ozurdex) intravitreal implant .** a 65-year - old caucasian woman was suffering from radiation macular edema following ruthenium-106 plaque brachytherapy for a choroidal melanoma on her left eye . (...) ② **seven months after the development of radiation macular edema , she received a single intravitreal injection of dexamethasone 0.7 mg (ozurdex) .** four weeks following the injection , her best - corrected visual acuity improved from 0.3 to 0.5 . (...) other studies suggest rates of radiation maculopathy from plaque radiotherapy of 18% , 23% , and 42.8% . (...) recently , a sustained - release dexamethasone implant (ozurdex) proved to be effective for the treatment of macular edema secondary to a variety of underlying diseases with a potentially lower rate of adverse events . there are no cases of radiation macular edema after ruthenium (ru)-106 plaque brachytherapy for choroidal melanoma resolved by an intravitreal dexamethasone 0.7-mg implant described so far in the literature . we report a case of radiation macular edema after ru-106 brachytherapy for a choroidal melanoma , refractory to a previous treatment with intravitreal bevacizumab , and resolved with significant improvement of visual function following an intravitreal injection of dexamethasone 0.7 mg . (...) she underwent one intravitreal injection of 0.5 mg bevacizumab (avastin , genentech / roche) in the following months without functional or anatomical improvement . ③ **seven months after the development of radiation macular edema , she received a single intravitreal injection of dexamethasone 0.7 mg (ozurdex) as off - label treatment .** (...) one of the main mechanisms of the chronic macular edema is the alteration of muller cells functionality ; it has been experimentally shown that steroids , by reducing the osmotic swelling of the muller 's cells , improve their functionality and reduce the macular edema . this could indicate that ④ **dexamethasone implant (ozurdex) might be an effective treatment option** not only in retinal vein occlusion and noninfectious uveitis , but can also be considered as off - label treatment **in radiation macular edema after ru-106 plaque brachytherapy for choroidal melanoma .**

Reference Summary:

purposeto report the effective treatment of radiation macular edema following ruthenium-106 plaque brachytherapy for a choroidal melanoma with a dexamethasone 0.7-mg (ozurdex) intravitreal implant.methodsan interventional case report with optical coherence tomography (oct) scans.resultsa 65-year - old caucasian woman was suffering from radiation macular edema following ruthenium-106 plaque brachytherapy for a choroidal melanoma on her left eye . she had undergone one intravitreal injection of 0.5 mg bevacizumab (avastin , genentech / roche) in the following months without functional or anatomical improvement . seven months after the development of radiation macular edema , she received a single intravitreal injection of dexamethasone 0.7 mg (ozurdex) . four weeks following the injection , her best - corrected visual acuity improved from 0.3 to 0.5 . radiation macular edema resolved with a reduction of central retinal thickness from 498 m before ozurdex injection to 224 m after ozurdex injection , as measured by oct scan.conclusiondexamethasone 0.7 mg (ozurdex) has proven to be an effective treatment option in retinal vein occlusion and noninfectious uveitis . it can also be considered as off - label treatment in radiation macular edema following ruthenium-106 plaque brachytherapy for a choroidal melanoma .

Figure 5: A sample summary comparison on the **PubMed** dataset. OTExtSum based summary sentences are highlighted in yellow colour. TextRank based summary sentences are underlined in red colour. TextRank extracted redundant contents, specifically the part ① is duplicated with the part ④, and the part ② is duplicated with the part ③. The summary generated by OTExtSum has ROUGE-1 F-Score of 73.1 and Semantic Coverage Score of 0.92, while the summary generated by TextRank has ROUGE-1 F-Score of 66.0 and Semantic Coverage Score of 0.89. Semantic Coverage Score of the ground-truth summary is 0.91.

Source Document:

(1) **Julian Speroni will take his Crystal Palace career into a 12th season** after agreeing a new contract. The goalkeeper, who joined from Dundee in July 2004, has triggered a 12-month extension that will expire at the end of next season. Speroni, yet again, has played a lead role for Palace this season in the club's attempts to stay in the Barclays Premier League. Julian Speroni will be at Crystal Palace for another year at least after triggering a 12-month contract extension . Speroni desperately dives to try and save Yaya Toure's effort against Crystal Palace on Monday night . The 35-year-old will have a testimonial against Dundee at the end of the season. But his Eagles career looks far from over as he continues to play a key role for Alan Pardew. Meanwhile, Palace are on the look out for a new shirt sponsor after money transfer firm Neteller ended their partnership with the club. Speroni signed from Dundee in 2004 and (2) **will take his Palace career into a 12th season .**

Reference Summary:

Julian Speroni signed from Dundee in 2004 and is a Crystal Palace legend . Argentine goalkeeper triggers contract extension at Selhurst Park . Speroni will have testimonial at the end of the season against Dundee .

Figure 6: A sample summary comparison on the **CNN/DM** dataset. OTExtSum based summary sentences are highlighted in yellow colour. TextRank based summary sentences are underlined in red colour. TextRank extracted redundant contents, specifically the part (1) is duplicated with the part (2). The summary generated by OTExtSum has ROUGE-1 F-Score of 50.5 and Semantic Coverage Score of 0.89, while the summary generated by TextRank has ROUGE-1 F-Score of 35.7 and Semantic Coverage Score of 0.83. Semantic Coverage Score of the ground-truth summary is 0.80.