Supervised Lazy Random Walk for Topic-Focused Multi-Document Summarization

Pan Du   Jiafeng Guo   Xueqi Cheng

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

dupan@software.ict.ac.cn, guojiafeng@ict.ac.cn ,cxq@ict.ac.cn

Abstract—Topic-focused multi-document summarization aims to produce a summary given a specific topic description and a set of related documents. It has become a crucial text processing task in many real applications that can help users consume the massive information. This paper presents a novel extractive approach based on supervised lazy random walk (SuperLazy). This approach naturally combines the rich features of sentences with the intrinsic sentence graph structure in a principled way, and thus enjoys the advantages of both the existing supervised and unsupervised approaches. Moreover, our approach can achieve the three major goals of topic-focused multi-document summarization (i.e. relevance, salience and diversity) simultaneously with a unified ranking process. Experiments on the benchmark dataset TAC2008 and TAC2009 are performed and the ROUGE evaluation results demonstrate that our approach can significantly outperform both the state-of-the-art supervised and unsupervised methods.

Keywords-supervised lazy random walk; topic-focused multi-document summarization; relevance; salience; diversity

I. INTRODUCTION

With the massive explosion of information on the Web, it has become increasingly important to provide improved mechanisms to present and explore online textual information. Topic-focused multi-document summarization [1], [2], [3], [4] has proved to be an effective and efficient way to present and help people consume online information. Given a specified topic description (e.g. a user query or a short narrative), topic-focused multi-document summarization aims to create a short summary over a set of topic-related documents, which conveys the most important aspects of the topic. Generality, there are three major goals [5] one needs to achieve simultaneously in topic-focused multi-document summarization: **Relevance**: The summary needs to either answer the need for information expressed in the topic or explain the topic. **Salience**: The summary has to neglect those trivial contents and preserve the salient information. **Diversity**: There should be less redundant information in the summary. In this way, we can cover as many aspects of the topic as possible within limited summary space.

Here we focus on extractive summarization approaches for topic-focused multi-document summarization, which extract sentences from the original documents to compose a summary [6], [7], [8], [9]. Extractive approaches can be further cast into two categories, namely supervised approaches [10], [11], [8] and unsupervised approaches [6], [12], [13]. In most supervised approaches, summarization is viewed as a two-class classification problem at the sentence level. A classification function is trained by considering rich linguistic and statistical features, including similarity between sentences, sentence location, term prominence, presence or absence of certain syntactic features and so on. However, the intrinsic structure among sentences from all the documents, which helps unveil both the local and global sentence relationships, is seldom exploited in the supervised approaches.

On the other hand, many unsupervised approaches treat summarization as graph-based ranking problem, by taking into account the global information collectively computed from the entire sentence graph. However, unsupervised approaches cannot make full use of the rich features mentioned above and often rely on heuristic rules that are difficult to generalize. Moreover, most existing approaches lacks a unified process that can automatically balance the relevance, salience and diversity in summarization. Some additional heuristic strategies or iterative process have often been applied to achieve diversity [3], [7], [12].

In this paper, we propose a novel extractive approach based on supervised lazy random walk (SuperLazy) for topic-focused multi-document summarization. This approach naturally combines the rich features of sentences with the intrinsic sentence graph structure in a principled way, and thus enjoys the advantages of both the existing supervised and unsupervised approaches. Moreover, we can achieve the three major goals of topic-focused multi-document summarization simultaneously with a unified ranking process.

Specifically, our approach leverages a process of random walk with restart [14] over the sentence graph in essentials, and introduces self-loop over sentence nodes to make the random walk "lazy" (i.e., the random walker stays at the node with some probability). Intuitively, by making the topic as the restart node, the random walk with restart over the sentence graph will obtain a topic biased PageRank over sentence nodes. Meanwhile, by introducing self-loop over sentence nodes, we can make nodes compete with each other to emphasize salient sentence nodes and more importantly, achieve diversity. Therefore, if we properly set the strengths of edges (between nodes) and self-loops, the final ranking of sentences under the lazy random walk can achieve relevance, salience and diversity in a unified way. Rather than setting the strengths of edges and self-loops in a heuristic way, we develop a learning method in a

supervised way that learns these strengths (i.e., random walk transition probabilities) so that the lazy random walk on such a weighted sentence graph is more likely to visit "positive" nodes (i.e. *summary sentences*) than "negative" nodes (i.e. *non-summary sentences*). We show that we can leverage rich sentence features to learn the strengths efficiently.

A set of experiments is conducted based on the benchmark dataset of TAC2008 and TAC2009 to evaluate our approach. The ROUGE evaluation results show that our approach can significantly outperform both the state-of-the-art supervised and unsupervised baseline methods.

The rest of the paper is organized as follows: Section 2 discusses some background and related work. Section 3 describes our approach in detail. Section 4 demonstrates our evaluation results and the conclusion is made in Section 5.

II. RELATED WORKS

A variety of approaches have been proposed for topic-focused multi-document summarization. Most of them can be categorized as either abstractive [15] or extractive [4].

Extractive summarization involves assigning scores to some units (e.g. sentences, paragraphs) of the documents and extracting the sentences with highest scores, which can be further categorized as supervised and unsupervised approaches. Supervised extractive approaches usually treat the summarization task as a two-class classification problem at the sentence level. For example, the Support Vector Machine (SVM) [16] based methods focus on constructing a decision boundary between summary sentences and non-summary sentences [10], [11]. Besides, Maximum Entropy [17], Hidden Markov Models [9], Conditional Random Field [8], and Regression models [4] have also been adopted to leverage the rich sentence features for summarization.

Some unsupervised extractive approaches take summarization as a metric-based sentence ranking problem, e.g. MMR [18] and information-theoretical models [6]. Recently, many unsupervised graph-based approaches have emerged which leverage the sentence graph for summarization. For example, Erkan and Radev [7] proposed LexRank, a variant of PageRank, for generic text summarization. Topic-sensitive LexRank [12] has been applied to the task of query-focused summarization. Mihalcea and Tarau [19] proposed another random-walk-based model TextRank, which was applied to single-document summarization. Zha [20] proposed a mutual reinforcement principle for sentence extraction using HITS. Wan et al. [3] applied a manifold-ranking algorithm to query-focused summarization.

To capture diversity in ranking process, DivRank [13] leveraged a vertex-reinforced random walk, which can simulate the "rich-get-richer" phenomenon to achieve salience and diversity simultaneously for summarization. Absorbing random walk [21] was also applied to achieve a diverse rank for summarization. Du et al. [5] introduced sink points into

manifold ranking to select topic-related, diverse and salient sentences for topic-focused summarization.

III. OUR APPROACH

We leverage random walk with restarts over sentences graph to address both the criteria of relevance and salience simultaneously. Meanwhile, we introduce self-loops over sentence nodes to make the random walk "lazy". A node with a heavy self-loop is likely to "absorb" the visits of its neighbors with light self-loops. Therefore, we can make the nodes "compete" with each other on "absorbing" visits to emphasize salient sentence nodes to achieve diversity in a unified process if we can set the strengths of edges and self-loops properly. Inspired by [22], we aim to leverage the rich sentence features (e.g. features that consider the passages) to help capture both the strengths of edges and self-loops in a principled way. We achieve this by developing a learning method that use rich features to learn the strength function in a supervised way.

A. SuperLazy Summarizer

Given a topic description q and a collection of sentences S from its related documents, an undirected sentence graph $G(V, E)$ is constructed, where V denotes all the sentences and E consists of edges between sentences and self-loops. Note here the topic description is also viewed as a pseudo sentence. The basic idea is that we will apply a lazy random walk originated from q over the sentence graph G, and select the top ranked nodes according to their stationary distribution $\boldsymbol{p}$ as summary sentences. Suppose $M = \{m_1, \cdots, m_j\}$ denotes the desired summary sentences in S, and $O = \{o_1, \cdots, o_k\}$ denotes other non-summary sentences. We aim to bias the lazy random walk so that it will visit nodes in M more often than nodes in O, i.e., $p_m > p_o$, for each $m \in M$ and $o \in O$. We achieve this by learning a function that will assign each edge and self-loop a strength so that the random walk will be more likely to visit nodes in M.

We assume that each edge/self-loop has a corresponding feature vector $\boldsymbol{\phi}$ that describes the relationships between nodes (e.g. cosine similarity between sentences) and the node's attributes (e.g. sentence length). For each edge/self-loop in G, we compute its strength using $f_{\boldsymbol{w}}(\boldsymbol{\phi})$, which is a non-negative function with respect to $\boldsymbol{\phi}$ and parameterized by $\boldsymbol{w}$. Here we assume the strength function as

$$f_{\boldsymbol{w}}(\boldsymbol{\phi}) = exp(\boldsymbol{w} \cdot \boldsymbol{\phi}).$$

Function $f_{\boldsymbol{w}}$ models the random walk transition probability $\mathcal{P}$ through computing the strengths of edges and self-loops. It is exactly the function $f_{\boldsymbol{w}}(\boldsymbol{\phi})$ (i.e. the parameters $\boldsymbol{w}$) that we need to learn in the training phase of our approach. The transition matrix $\mathcal{P}$ of our lazy random walk hence can be

represented as follows

$$\mathcal{P}_{uv} = \begin{cases} (1-\alpha)\frac{exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv})}{\sum_{v'} exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv'})} + \alpha\mathbf{1}(v=q) & (u,v)\in E, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where α is the restart probability (i.e., with probability α the random walk jumps back to topic node q).

Now we formalize the learning problem of our approach. Given a training corpus $C = \{(q^{(i)}, S^{(i)})|i = 1,\ldots,N\}$, where $q^{(i)}$ denotes the i-th topic description and $S^{(i)}$ denotes the collection of sentences from $q^{(i)}$'s related documents. We also denote the human labeled summary sentences from $S^{(i)}$ for $q^{(i)}$ as $M^{(i)} = \{m_1, \cdots, m_j\}$, and other non-summary sentences as $O^{(i)} = \{o_1, \cdots, o_k\}$. Sentences in M can be considered as positive training examples, while those in O as negative ones. The problem of finding parameters $\boldsymbol{w}$ of edge strength function $f_{\boldsymbol{w}}(\boldsymbol{\phi})$ can be formalized as the following optimization problem:

$$\min_{\boldsymbol{w}} \mathcal{L}(\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda \sum_{i=1}^{N} \sum_{\substack{o\in O^{(i)} \\ m\in M^{(i)}}} l(p_o - p_m) \quad (2)$$

where $\boldsymbol{p}$ is the stationary distribution under the lazy random walk process, $p_u \in \boldsymbol{p}$ is the probability of node u, and λ is the regularization parameter that trades-off between the model complexity and the fit of the model (i.e., how much the rank constraints $p_m > p_o$ can be violated). Note the distribution $\boldsymbol{p}$ depends on the transition probability, which is further decided by edge strength $f_{\boldsymbol{w}}(\boldsymbol{\phi})$ parameterized with $\boldsymbol{w}$. Moreover, $l(\cdot)$ is a loss function that assigns a non-negative penalty according to the difference of the scores $p_o - p_m$. If $p_o - p_m < 0$ then $l(\cdot) = 0$ since the constraint $p_m > p_o$ is not violated, while for $p_o - p_m > 0$, we have $l(\cdot) > 0$. There might be different ways to select the form of the loss function. Here we use Wilcoxon-Mann-Whitney (WMW) loss $l(x) = (1 + e^{-x/b})^{-1}$ [22] as it shows consistent good performance.

Learning
Our goal now is to minimize Eqn. (2) with respect to the parameter vector $\boldsymbol{w}$. Here we perform gradient based optimization method to minimize the loss and find the optimal parameters $\boldsymbol{w}$. First, we derive the gradient of $\mathcal{L}(\boldsymbol{w})$ with respect to $\boldsymbol{w}$:

$$\frac{\partial \mathcal{L}(\boldsymbol{w})}{\partial \boldsymbol{w}} = \boldsymbol{w} + \lambda \sum_{i=1}^{N} \sum_{\substack{o\in O^{(i)} \\ m\in M^{(i)}}} \frac{\partial l(\delta_{om})}{\partial \delta_{om}} \left(\frac{\partial p_o}{\partial \boldsymbol{w}} - \frac{\partial p_m}{\partial \boldsymbol{w}} \right) \quad (3)$$

where $\delta_{om} = p_o - p_m$, and the derivative of WMW loss $l(\cdot)$ can be easily obtained as $\frac{\partial l(x)}{\partial x} = [b (2 + e^{-x/b} + e^{x/b})]^{-1}$. Now, only $\frac{\partial p_u}{\partial \boldsymbol{w}}$ is still left unknown in Eqn. (3).

As aforementioned, the vector $\boldsymbol{p}$ depends on the transition probability, which is further decided by edge strength $f_{\boldsymbol{w}}(\boldsymbol{\phi})$

parameterized with $\boldsymbol{w}$. In other words, we can establish the connection between vector $\boldsymbol{p}$ and the parameters $\boldsymbol{w}$ through the random walk transition matrix $\mathcal{P}$. Since the vector $\boldsymbol{p}$ is the stationary distribution of our lazy random walk, the equation $\boldsymbol{p}^T = \boldsymbol{p}^T\mathcal{P}$ always holds. It can be further rewritten as $p_u = \sum_v p_v \mathcal{P}_{vu}$ and taking the derivative now gives

$$\frac{\partial p_u}{\partial \boldsymbol{w}} = \sum_v \mathcal{P}_{vu}\frac{\partial p_v}{\partial \boldsymbol{w}} + p_v \frac{\partial \mathcal{P}_{vu}}{\partial \boldsymbol{w}}. \quad (4)$$

According to Eqn. (1), the derivative of $\mathcal{P}$ with respect to $\boldsymbol{w}$ in Eqn. (4) can be obtained as

$$\frac{\partial \mathcal{P}_{uv}}{\partial \boldsymbol{w}} = (1-\alpha)exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv})\left(\frac{\boldsymbol{\phi}_{uv}}{\sum_{v'} exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv'})} - \frac{\sum_{v'}\boldsymbol{\phi}_{uv'}exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv'})}{(\sum_{v'} exp(\boldsymbol{w}\cdot\boldsymbol{\phi}_{uv'}))^2} \right). \quad (5)$$

Therefore, we can compute the gradient $\frac{\partial p_u}{\partial \boldsymbol{w}}$ in an iterative way [23] like power-method. Up to now, all the components needed to evaluate the derivative of $\mathcal{L}(\boldsymbol{w})$ in Eqn. (2) can be obtained. Hence, we can directly minimize $\mathcal{L}(\boldsymbol{w})$ with a standard quasi-Newton method. Specifically, we use the BFGS [24] algorithm for our learning process. Note that the formalized problem is generally not convex, and the gradient descent methods will not necessarily find the global minimum. We resolve this by trying several different starting points to get a good solution in practice.

Prediction
In prediction, we apply our proposed SuperLazy Model to generate a topic-focused multi-document summary given a new topic description q' and the sentence collection S' from its related documents. Specifically, we construct a sentence graph G' for q' and calculate strengths of edges and self-loops using the learned strength function $f_{\boldsymbol{w}}$. We then perform a lazy random walk process originating from q' over the sentence graph G'. The stationary distribution $\boldsymbol{p}^*$ of the random walk assigns each node u a probability p_u^*. Nodes are then ordered according to p_u^* and the top ranked nodes are selected as the summary sentences.

B. Feature Space

Many features have proven to be useful for topic-focused multi-document summarization. Here we discuss several features we adopt in our SuperLazy summarizer. There are two types of features according to whether they are associated with the edges between nodes or self-loops, referred as edge features and self-loop features, respectively. The feature values are either binary or normalized into the interval $[0, 1]$.

Edge Features
Edge features are used to describe the relationships between different sentences on the sentence graph. We adopt three common features for this type, i.e., **Cosine Similarity**, **Jaccard Similarity**, and **Passage Co-occurrence**. The cosine similarity value between two sentences calculated

based on their term vector representations, which record the term frequencies. The Jaccard similarity value between two sentences which measures the ratio of overlapped terms. Passage Co-occurrence is a binary feature which indicates whether the two sentences co-occur in a same passage. It is set to be 1 if they are within one passage, 0 otherwise.

Self-loop Features

We also consider features dedicated to self-loops, which are specially important to bias the lazy random walk process so that the summary sentences would be visited more times.

The **Position** of a sentence along the sentence sequence of a document. If the sentence appears at the beginning of the document, it is set to be 1, otherwise 0. **Length** is the number of terms contained in a sentence after removing the stop-words. The **Log Likelihood** of sentence s_i being generated by the whole related document set A, $\log P(s_i|A)$. It is calculated by $\sum_{t_k} N(t_k, s_i) log p(t_k|A)$, where $N(t_k, s_i)$ is the number of occurrences of term t_k in s_i and $p(t_k|A)$ can be estimated by $N(t_k, A)/\sum_j(t_j, A)$. For sentence s_i, **Inner-document Degree** is the number of edges between sentence s_i and sentences from the same document of s_i. For sentence s_i, **Intra-document Degree** is the number of edges between sentence s_i and sentences from other documents different from that of s_i. **Local Inner-document Degree Order** is a binary feature which indicates whether the Inner-document Degree of sentence s_i is the largest among its neighbors. **Local Intra-document Degree Order** is another binary feature which indicates whether the Intra-document Degree of sentence s_i is the largest among its neighbors.

IV. EXPERIMENTS

A. Datasets and Evaluation Metric

Topic-focused multi-document summarization has been one of the main tasks in Text Analysis Conference (TAC) hold by NIST[1] for several years. In our experiments, we use datasets of TAC2008 and TAC2009. TAC2008 provided 48 topics and TAC2009 provided 44 topics. Each topic was composed of 20 relevant documents from the AQUAINT-2 collection of news articles. The documents were further equally divided into two datasets: Document Set A and Document Set B. All the documents in set A chronologically preceded the documents in set B. In TAC task, a 100-word summary was required to be generated for each set of documents. The summary of Set A should be a topic-focused multi-document summary, and that of B should be a update summary. Since here we focus on the performance comparison of topic-focused multi-document summarization (i.e. the summary for set A), we did not show the summarization results on set B.

In our evaluation, we use the ROUGE-1 (unigram-based), ROUGE-2 (bigram-based) and ROUGE-SU4 (an extended version of ROUGE-2) recall metrics, which have been shown

[1]http://www.nist.gov

to correlate well with human judgments based on comparison with a single model. They were also used as official automatic evaluation metrics for TAC2008 and TAC2009 respectively. The results are obtained with ROUGE version 1.5.5 with the settings used for TAC2008.

ROUGE [25] measures summary quality by counting the number of overlapping units such as n-gram, word sequences, and word pairs between the computer-generated summary and the ideal summaries created by humans. The n-gram recall measure, ROUGE-N, is computed as

$$ROUGE-N = \frac{\sum\limits_{S\in\{Refs\}} \sum\limits_{gram_n\in S} Cnt_{match}(gram_n)}{\sum\limits_{S\in\{Refs\}} \sum\limits_{gram_n\in S} Cnt(gram_n)},$$

where n is the length of the n-gram, $Cnt_{match}(gram_n)$ is the maximum number of n-grams co-occurring in a candidate summary and a set of reference summaries $Refs$, and $Cnt(gram_n)$ is the number of n-gram in the reference summaries.

B. Baseline Methods

We compared our proposed SuperLazy summarizer with both the state-of-the-art unsupervised and supervised approaches. Among the extensive unsupervised approaches, we choose Leading Sentence Selection (LEAD), Maximal Marginal Relevance (MMR), Personalized PageRank (PPR), Manifold Ranking (MR), DivRank (DR), and GrassHopper (GH) as baselines. We also choose Support Vector Machine (SVM) as a supervised baseline for comparison.

LEAD is a simple baseline approach which selects the leading sentences of documents to compose the summary. It can be considered as a lower bound of extractive approach for topic-focused multi-document summarization. **MMR** method [18] measures the relevance and diversity independently and provides a linear combination, called "marginal relevance", as the metric. The summary sentences are then selected iteratively according to the metric. **PPR** [26] is a widely used random walk based ranking approach, which is appropriate for identifying relevant and salient sentences for summary. **MR** approach [27] is a graph-based ranking approach using the graph regularization. It is capable of giving higher ranks to the salient sentences which are close to the topic on the manifold. An iterative sentence selection process is required to achieve diversity. **DR** [13] is also a graph-based ranking approach which uses a vertex-reinforced random walk. It focuses on achieving diversity in sentence ranking for summary. **GH** [21] conducts an absorbing random walk over the graph. It aims to capture diversity in ranking by iteratively select top ranked nodes and set it as absorbing state. **SVM** [16] is widely used as a binary classifier, which is a supervised approach that generally used to distinguish summary sentences from non-summary sentences. Additional steps are taken to remove redundant sentences.

We denote our approach based on supervised lazy random walk as **SuperLazy**. For training, we leverage the human extracted summaries as our ground truth. NIST has provided a set of manual summarization results on TAC2009, which is extracted by a team of five human "sentence-extractors" from the University of Montreal. Therefore, we used such human extracted summaries as our ground truth for training. We evaluated the performance of our approach and baseline methods over both the dataset TAC2008 and TAC2009.

For our SuperLazy approach, we set the parameter $\lambda = 1$ and $b = 0.1$ in our experiment. For the restart probability α, we set it as 0.2 since it performs best on ROUGE-2 evaluation as shown in latter experiments. For other baselines, we set their parameters (if exist) to the values as they can achieve their best performances on ROUGE-2 evaluation for summarization.

C. Evaluation Results

1) Evaluation Results on Benchmark of TAC2009: We first conducted performance comparison based on the benchmark dataset TAC2009. We split the 44 topics of dataset TAC2009 into 4 folders. Each time we select 3 folders for training/tuning the parameters and the remaining 1 folder for testing. The 4-fold cross-validation results are demonstrated in Table I. The numbers in the parentheses are the relative improvements compared with baseline method LEAD.

Table I
THE SUMMARIZATION PERFORMANCE BASED ON A 4-FOLD CROSS VALIDATION ON TAC2009

Method	ROUGE-1	ROUGE-2	ROUGE-SU4
LEAD	0.30192	0.06311	0.09869
PPR	0.36163	0.08490	0.12497
MMR	0.34291	0.07915	0.11138
GH	0.36802	0.08927	0.12469
MR	0.35499	0.08612	0.12072
DR	0.36571	0.08283	0.12306
SVM	0.35889	0.09103	0.12794
SuperLazy	**0.38210**	**0.10857**	**0.14245**

From the results we can see that, our approach outperforms all the unsupervised baselines in terms of all the ROUGE metrics consistently. When compared with the overall best performed unsupervised approach GH, the relative improvement over the ROUGE-1, ROUGE-2 and ROUGH-SU4 scores of our approach is 3.8%, 21.6% and 14.2%, respectively. It indicates that by leveraging rich sentence features and learning the model in a supervised way, we can achieve better quality in summarization. Meanwhile, our SuperLazy approach can also outperform the supervised approach SVM in terms of all the evaluation metrics. In comparison with SVM method, The relative improvements over the ROUGE-1, ROUGE-2 and ROUGH-SU4 scores of our approach are 6.5%, 19.3% and 11.3%, respectively. This shows that by exploring the intrinsic local and global structure over sentences, we can learn a better summarization
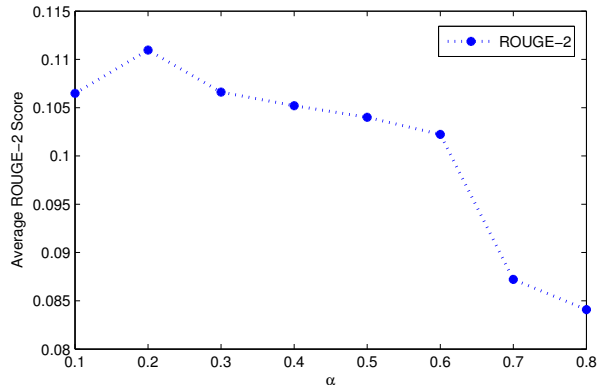

Figure 1. ROUGE-2 Score vs. parameter α.

model. We further conducted statistical tests on the results, which indicates that all these improvements are statistically significant (p-value< 0.01).

2) Evaluation Results on Benchmark of TAC2008: We conducted experiments on dataset TAC2008 to further evaluate the effectiveness of our model. Since there is no human labeled summarization results as ground truth on dataset TAC2008, we trained our model based on dataset TAC2009 and used dataset TAC2008 as the test set. In this way, we can also evaluate the generalization ability of our approach. The evaluation results are shown in Table II, where both the average ROUGE-N scores and the confidence interval (95%) are depicted. The evaluation results again demonstrate that our approach can achieve better performance than both the state-of-the-art supervised and unsupervised approaches on dataset TAC2008 in terms of all the ROUGE metrics. This is consistent with the results we obtained on dataset TAC2009. The results also indicate that our approach can generalize very well when we apply it on a different dataset.

Table II
THE SUMMARIZATION PERFORMANCE ON TAC2008

Method	ROUGE-1	ROUGE-2	ROUGE-SU4
LEAD	0.28889	0.05871	0.09300
PPR	0.36045	0.08889	0.12708
MMR	0.33666	0.07533	0.11383
GH	0.36549	0.09039	0.13017
MR	0.36312	0.09402	0.12910
DR	0.36548	0.09077	0.12897
SVM	0.36092	0.09432	0.12814
SuperLazy	**0.36739**	**0.09645**	**0.13070**

D. Parameter Tuning

We also conducted experiments to test how the restart probability α affects the proposed model. Figure 1 shows the influence of parameter α on the summarization performance. As we can see, within a broad range from 0.1 to 0.6, the SuperLazy summarizer performs quite well with the best performance achieved when $\alpha = 0.2$. However, the performance drops when α becomes increasingly large. This is a reasonable result according to α's function. Intuitively, the parameter α controls for how "far" the random walker

wanders from the topic node q before it jumps back to q and restarts. High value of α gives very short and local random walks. Therefore, it would become difficult to find the right summary sentences with such a strong restriction.

V. Conclusion

In this paper, we propose a novel extractive approach, referred as SuperLazy summarizer, for topic-focused multi-document summarization. Our approach naturally and in a principled way combines the rich features of sentences with the intrinsic sentence graph structure for producing better recommendation. It is a unique model in that it can well balance the three major facts, namely relevance, salience and diversity, in a unified process for topic-focused multi-document summarization. Experimental results on benchmark datasets demonstrate that our approach can significantly outperform the other state-of-the-ar baseline methods.

Acknowledgment

References

[1] J. M. Conroy and J. D. Schlesinger, "Classy query-based multidocument summarization," in *In DUC'2005*, 2005.

[2] A. Nenkova, L. Vanderwende, and K. McKeown, "A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization," in *SIGIR '06*. New York, NY, USA: ACM, 2006, pp. 573–580.

[3] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic-focused multi-document summarization," in *IJCAI 07'*, Hyderabad, India, January 6-12 2007, pp. 2903–2908.

[4] Y. Ouyang, W. Li, S. Li, and Q. Lu, "Applying regression models to query-focused multi-document summarization," *Information Processing & Management*, vol. In Press, Corrected Proof, pp. –, 2010.

[5] P. Du, J. Guo, J. Zhang, and X. Cheng, "Manifold ranking with sink points for update summarization," in *Proceedings of CIKM 10'*, ser. CIKM '10. New York, NY, USA: ACM, 2010, pp. 1757–1760.

[6] T. Ma and X. Wan, "Multi-document summarization using minimum distortion," in *Proceedings of ICDM 10'*, ser. ICDM '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 354–363.

[7] G. Erkan and D. R. Radev, "Lexrank: graph-based lexical centrality as salience in text summarization," *J. Artif. Int. Res.*, vol. 22, no. 1, pp. 457–479, 2004.

[8] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields," in *IJCAI*, 2007, pp. 2862–2867.

[9] J. M. Conroy and D. P. O'leary, "Text summarization via hidden markov models," in *SIGIR '01*. New York, NY, USA: ACM, 2001, pp. 406–407.

[10] C. J. V. Rijsbergen, *Information Retrieval*, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.

[11] L. Li, K. Zhou, G.-R. Xue, H. Zha, and Y. Yu, "Enhancing diversity, coverage and balance for summarization through structure learning," in *WWW '09*. New York, NY, USA: ACM, 2009, pp. 71–80.

[12] J. Otterbacher, G. Erkan, and D. Radev, "Using random walks for question-focused sentence retrieval," in *Proceedings of EMNLP 05'*. ACL, October 2005, pp. 915–922.

[13] Q. Mei, J. Guo, and D. Radev, "Divrank: the interplay of prestige and diversity in information networks," in *Proceedings of SIGKDD 10'*, ser. KDD '10. New York, NY, USA: ACM, 2010, pp. 1009–1018.

[14] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Proceedings of ICDM*, ser. ICDM '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 613–622.

[15] K. Knight and D. Marcu, "Summarization beyond sentence extraction: a probabilistic approach to sentence compression," *Artif. Intell.*, vol. 139, pp. 91–107, July 2002.

[16] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[17] M. Osborne, "Using maximum entropy for sentence extraction," in *Proceedings of the ACL-02 Workshop on Automatic Summarization*. Morristown, NJ, USA: Association for Computational Linguistics, 2002, pp. 1–8.

[18] J. Carbonell and J. Goldstein, "The use of mmr, diversity-based reranking for reordering documents and producing summaries," in *SIGIR '98*. New York, NY, USA: ACM, 1998, pp. 335–336.

[19] R. Mihalcea and P. Tarau, "Textrank: Bringing order into texts," in *Proceedings of EMNLP 2004*, D. Lin and D. Wu, Eds. Barcelona, Spain: Association for Computational Linguistics, July 2004, pp. 404–411.

[20] H. Zha, "Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering," in *SIGIR '02*. New York, NY, USA: ACM, 2002, pp. 113–120.

[21] X. Zhu, A. Goldberg, J. Van Gael, and D. Andrzejewski, "Improving diversity in ranking using absorbing random walks," in *NAACL 07'*. Rochester, New York: Association for Computational Linguistics, April 2007, pp. 97–104.

[22] L. Backstrom and J. Leskovec, "Supervised random walks: predicting and recommending links in social networks," in *Proceedings of WSDM*, ser. WSDM '11. New York, NY, USA: ACM, 2011, pp. 635–644.

[23] A. L. ANDREW, "Iterative computation of derivatives of eigenvalues and eigenvectors," *Ima Journal of Applied Mathematics*, vol. 24, pp. 209–218, 1979.

[24] C. G. Broyden., "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, vol. 19, pp. 577–593, 1965.

[25] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, S. S. Marie-Francine Moens, Ed. Barcelona, Spain: ACL, July 2004, pp. 74–81.

[26] T. H. Haveliwala, "Topic-sensitive pagerank," in *Proceedings of WWW 02'*, ser. WWW '02. New York, NY, USA: ACM, 2002, pp. 517–526.

[27] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2004.