Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset[*]

Piji Li[†]   **Lidong Bing**[‡]   **Wai Lam**[†]

[†]Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong
[‡]AI Lab, Tencent Inc., Shenzhen, China
[†]`{pjli, wlam}@se.cuhk.edu.hk`, [‡]`lyndonbing@tencent.com`

Abstract

We investigate the problem of reader-aware multi-document summarization (RA-MDS) and introduce a new dataset for this problem. To tackle RA-MDS, we extend a variational auto-encodes (VAEs) based MDS framework by jointly considering news documents and reader comments. To conduct evaluation for summarization performance, we prepare a new dataset. We describe the methods for data collection, aspect annotation, and summary writing as well as scrutinizing by experts. Experimental results show that reader comments can improve the summarization performance, which also demonstrates the usefulness of the proposed dataset. The annotated dataset for RA-MDS is available online[1].

1 Introduction

The goal of multi-document summarization (MDS) is to automatically generate a brief, well-organized summary for a topic which describes an event with a set of documents from different sources. (Goldstein et al., 2000; Erkan and Radev, 2004; Wan et al., 2007; Nenkova and McKeown, 2012; Min et al., 2012; Bing et al., 2015; Li et al., 2017). In the typical setting of MDS, the input is a set of news documents about the same topic. The output summary is a piece of short text document containing several sentences, generated only based on the input original documents.

With the development of social media and mobile equipments, more and more user generated

Figure 1: Reader comments of the news "The most important announcements from Google's big developers' conference (May, 2017)".

content is available. Figure 1 is a snapshot of reader comments under the news report "The most important announcements from Google's big developers' conference"[2]. The content of the original news report talks about some new products based on AI techniques. The news report generally conveys an enthusiastic tone. However, while some readers share similar enthusiasms, some others express their worries about new products and technologies and these comments can also reflect their interests which may not be very salient in the original news reports. Unfortunately, existing MDS approaches cannot handle this issue. We investigate this problem known as reader-aware multi-document summarization (RA-MDS). Under the RA-MDS setting, one should jointly consider news documents and reader comments when generating the summaries.

One challenge of the RA-MDS problem is how to conduct salience estimation by jointly considering the focus of news reports and the reader interests revealed by comments. Meanwhile, the model should be insensitive to the availability of diverse aspects of reader comments. Another challenge is that reader comments are very noisy, not fully grammatical and often expressed in infor-

[1]`http://www1.se.cuhk.edu.hk/~textmine/dataset/ra-mds/`

[2]https://goo.gl/DdU0vL

mal expressions. Some previous works explore the effect of comments or social contexts in single document summarization such as blog summarization (Hu et al., 2008; Yang et al., 2011). However, the problem setting of RA-MDS is more challenging because the considered comments are about an event which is described by multiple documents spanning a time period. Another challenge is that reader comments are very diverse and noisy. Recently, Li et al. (2015) employed a sparse coding based framework for RA-MDS jointly considering news documents and reader comments via an unsupervised data reconstruction strategy. However, they only used the bag-of-words method to represent texts, which cannot capture the complex relationship between documents and comments.

Recently, Li et al. (2017) proposed a sentence salience estimation framework known as *VAE-Sum* based on a neural generative model called Variational Auto-Encoders (VAEs) (Kingma and Welling, 2014; Rezende et al., 2014). During our investigation, we find that the Gaussian based VAEs have a strong ability to capture the salience information and filter the noise from texts. Intuitively, if we feed both the news sentences and the comment sentences into the VAEs, commonly existed latent aspect information from both of them will be enhanced and become salient. Inspired by this consideration, to address the sentence salience estimation problem for RA-MDS by jointly considering news documents and reader comments, we extend the VAESum framework by training the news sentence latent model and the comment sentence latent model simultaneously by sharing the neural parameters. After estimating the sentence salience, we employ a phrase based compressive unified optimization framework to generate a final summary.

There is a lack of high-quality dataset suitable for RA-MDS. Existing datasets from DUC[3] and TAC[4] are not appropriate. Therefore, we introduce a new dataset for RA-MDS. We employed some experts to conduct the tasks of data collection, aspect annotation, and summary writing as well as scrutinizing. To our best knowledge, this is the first dataset for RA-MDS.

Our contributions are as follows: (1) We investigate the RA-MDS problem and introduce a new dataset for the problem of RA-MDS. To our best

knowledge, it is the first dataset for RA-MDS. (2) To tackle the RA-MDS, we extend a VAEs-based MDS framework by jointly considering news documents and reader comments. (3) Experimental results show that reader comments can improve the summarization performance, which also demonstrates the usefulness of the dataset.

2 Framework

2.1 Overview

As shown in Figure 2, our reader-aware news sentence salience framework has three main components: (1) latent semantic modeling; (2) comment weight estimation; (3) joint reconstruction. Consider a dataset X_d and X_c consisting of n_d news sentences and n_c comment sentences respectively from all the documents in a topic (event), represented by bag-of-words vectors. Our proposed news sentence salience estimation framework is extended from VAESum (Li et al., 2017), which can jointly consider news documents and reader comments. One extension is that, in order to absorb more useful information and filter the noisy data from comments, we design a weight estimation mechanism which can assign a real value ρ_i for a comment sentence $\mathbf{x}_c^i$. The comment weight $\boldsymbol{\rho} \in \mathbb{R}^{n_c}$ is integrated into the VAEs based sentence modeling and data reconstruction component to handle comments.

2.2 Reader-Aware Salience Estimation

Variational Autoencoders (VAEs) (Kingma and Welling, 2014; Rezende et al., 2014) is a generative model based on neural networks which can be used to conduct latent semantic modeling. Li et al. (2017) employ VAEs to map the news sentences into a latent semantic space, which is helpful in improving the MDS performance. Similarly, we also employ VAEs to conduct the semantic modeling for news sentences and comment sentences. Assume that both the prior and posterior of the latent variables are Gaussian, i.e., $p_\theta(\mathbf{z}) = \mathcal{N}(0, \mathbf{I})$ and $q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mathbf{I})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ denote the variational mean and standard deviation respectively, which can be calculated with a multilayer perceptron (MLP). VAEs can be divided into two phases, namely, encoding (inference), and decoding (generation). All the operations are de-
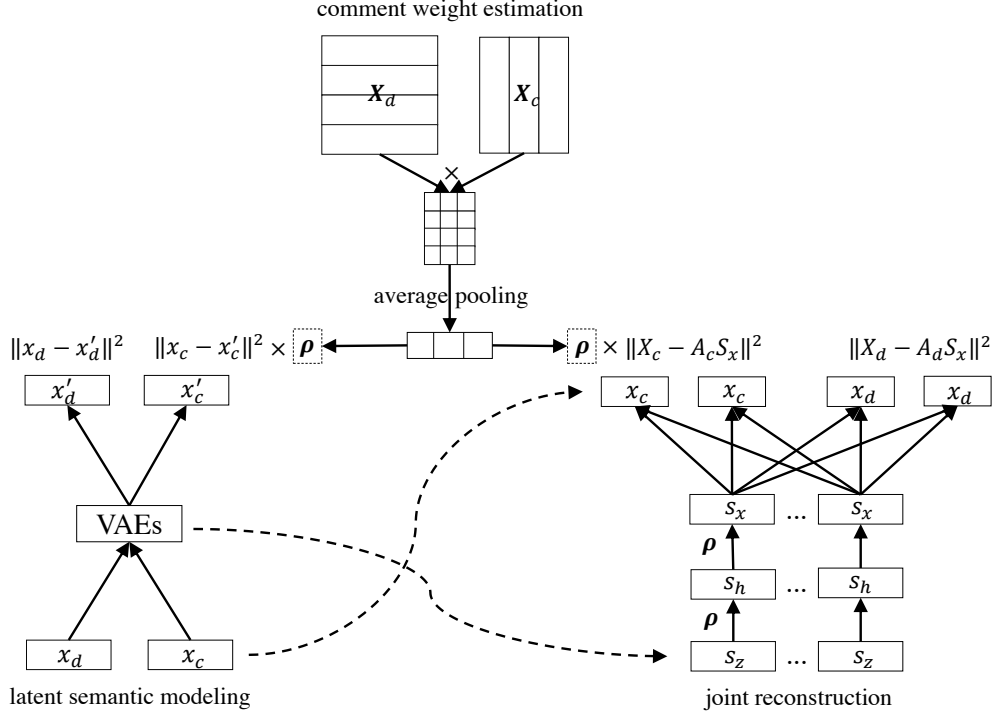
---

[3]http://duc.nist.gov/
[4]http://tac.nist.gov/

Figure 2: Our proposed framework. **Left**: Latent semantic modeling via variation auto-encoders for news sentence $\mathbf{x}_d$ and comment sentence $\mathbf{x}_c$. **Middle**: Comment sentence weight estimation. **Right**: Salience estimation by a joint data reconstruction method. $\mathbf{A}_d$ is a news reconstruction coefficient matrix which contains the news sentence salience information.

picted as follows:

$$
\begin{aligned}
h_{enc} &= relu(W_{xh}x + b_{xh}) \\
\mu &= W_{h\mu}h_{enc} + b_{h\mu} \\
\log(\sigma^2) &= W_{h\sigma}h_{enc} + b_{h\sigma} \\
\varepsilon &\sim \mathcal{N}(0, \mathbf{I}), \quad z = \mu + \sigma \otimes \varepsilon \\
h_{dec} &= relu(W_{zh}z + b_{zh}) \\
x' &= sigmoid(W_{hx}h_{dec} + b_{hx})
\end{aligned}
\tag{1}
$$

Based on the reparameterization trick in Equation 1, we can get the analytical representation of the variational lower bound $\mathcal{L}(\theta, \varphi; \mathbf{x})$:

$$
\log p(x|z) = \sum_{i=1}^{|V|} x_i \log x_i' + (1 - x_i) \cdot \log(1 - x_i')
$$

$$
-D_{KL}[q_\varphi(z|x)\|p_\theta(z)] = \frac{1}{2}\sum_{i=1}^{K}\left(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2\right)
$$

where $\mathbf{x}$ denotes a general sentence, and it can be a news sentence $\mathbf{x}_d$ or a comment sentnece $\mathbf{x}_c$.

By feeding both the news documents and the reader comments into VAEs, we equip the model a ability of capturing the information from them jointly. However, there is a large amount of noisy information hidden in the comments. Hence we design a weighted combination mechanism for fusing news and comments in the VAEs. Precisely, we split the variational lower bound $\mathcal{L}(\theta, \varphi; \mathbf{x})$

into two parts and fuse them using the comment weight $\boldsymbol{\rho}$:

$$
\mathcal{L}(\theta, \varphi; \mathbf{x}) = \mathcal{L}(\theta, \varphi; \mathbf{x}_d) + \boldsymbol{\rho} \times \mathcal{L}(\theta, \varphi; \mathbf{x}_c) \tag{2}
$$

The calculation of $\boldsymbol{\rho}$ will be discussed later.

The news sentence salience estimation is conducted by an unsupervised data reconstruction framework. Assume that $\mathbf{S}_z = \{\mathbf{s}_z^1, \mathbf{s}_z^2, \cdots, \mathbf{s}_z^m\}$ are m latent aspect vectors used for reconstructing all the latent semantic vectors $\mathbf{Z} = \{\mathbf{z}^1, \mathbf{z}^2, \cdots, \mathbf{z}^n\}$. Thereafter, the variational-decoding progress of VAEs can map the latent aspect vector $\mathbf{S}_z$ to $\mathbf{S}_h$, and then produce m new aspect term vectors $\mathbf{S}_x$:

$$
\begin{aligned}
s_h &= relu(W_{zh}s_z + b_{zh}) \\
s_x &= sigmoid(W_{hx}s_h + b_{hx})
\end{aligned}
\tag{3}
$$

VAESum (Li et al., 2017) employs an alignment mechanism (Bahdanau et al., 2015; Luong et al., 2015) to recall the lost detailed information from the input sentence. Inspired this idea, we design a jointly weighted alignment mechanism by considering the news sentence and the comment sentence simultaneously. For each decoder hidden state s_h^i, we align it with each news encoder hidden state h_d^j

by an alignment vector $a^d \in \mathbb{R}^{n_d}$. We also align it with each comments encoder hidden state h_c^j by an alignment vector $a^c \in \mathbb{R}^{n_c}$. In order to filter the noisy information from the comments, we again employ the comment weight ρ to adjust the alignment vector of comments:

$$\tilde{a}^c = a^c \times \rho \tag{4}$$

The news-based context vector c_d^i and the comment-based context vector c_c^i can be obtained by linearly blending the input hidden states respectively. Then the output hidden state can be updated based on the context vectors:

$$\tilde{s}_h^i = \tanh(W_{dh}^h c_d^i + W_{ch}^h c_c^i + W_{hh}^a s_h^i) \tag{5}$$

Then we can generate the updated output aspect vectors based on $\tilde{s}_h^i$. We add a similar alignment mechanism into the output layer.

$\mathbf{S}_z$, $\mathbf{S}_h$, and $\mathbf{S}_x$ can be used to reconstruct the space to which they belong respectively. In order to capture the information from comments, we design a joint reconstruction approach here. Let $\mathbf{A}_d \in \mathbb{R}^{n_d \times m}$ be the reconstruction coefficient matrix for news sentences, and $\mathbf{A}_c \in \mathbb{R}^{n_c \times m}$ be the reconstruction coefficient matrix for comment sentences. The optimization objective contains three reconstruction terms, jointly considering the latent semantic reconstruction and the term vector space reconstruction for news and comments respectively:

$$\mathcal{L}_A = (\|Z_d - A_d S_z\|_2^2 + \|H_d - A_d S_h\|_2^2$$
$$+ \|X_d - A_d S_x\|_2^2) + \rho \times (\|Z_c - A_c S_z\|_2^2 \tag{6}$$
$$+ \|H_c - A_c S_h\|_2^2 + \|X_c - A_c S_x\|_2^2)$$

This objective is integrated with the variational lower bound of VAEs $\mathcal{L}(\theta, \varphi; \mathbf{x})$ and optimized in a multi-task learning fashion. Then the new optimization objective is:

$$\mathcal{J} = \min_{\Theta}(-\mathcal{L}(\theta, \varphi; x) + \mathcal{L}_A) \tag{7}$$

where Θ is a set of all the parameters related to this task. We define the magnitude of each row of $\mathbf{A}_d$ as the salience scores for the corresponding news sentences.

We should note that the most important variable in our framework is the comment weight vector ρ, which appears in all the three components of our framework. The basic idea for calculating ρ is that if the comment sentence is more similar to the news content, then it contains less noisy information. For all the news sentences X_d and all the comment sentences X_c, calculate the relation matrix $R \in \mathbb{R}^{n_d \times n_c}$ by:

$$R = X_d \times X_c^T \tag{8}$$

Then we add an average pooling layer to get the coefficient value for each comment sentence:

$$\mathbf{r} = \frac{1}{n_c} \sum_{i=1}^{n_c} R[i, :] \tag{9}$$

Finally, we add a sigmoid function to adjust the coefficient value to $(0, 1)$:

$$\rho = sigmoid(\mathbf{r}) \tag{10}$$

Because we have different representations from different vector space for the sentences, therefore we can calculate the comment weight in different semantic vector space. Here we use two spaces, namely, latent semantic space obtained by VAEs, and the original bag-of-words vector space. Then we can merge the weights by a parameter λ_p:

$$\rho = \lambda_p \times \rho_z + (1 - \lambda_p) \times \rho_x \tag{11}$$

where ρ_z and ρ_x are the comment weight calculated from latent semantic space and term vector space. Actually, we can regard ρ as some gates to control the proportion of each comment sentence absorbed by the framework.

2.3 Summary Construction

In order to produce reader-aware summaries, inspired by the phrase-based model in Bing et al. (2015) and Li et al. (2015), we refine this model to consider the news sentences salience information obtained by our framework. Based on the parsed constituency tree for each input sentence, we extract the noun-phrases (NPs) and verb-phrases (VPs). The overall objective function of this optimization formulation for selecting salient NPs and VPs is formulated as an integer linear programming (ILP) problem:

$$\max\{\sum_i \alpha_i S_i - \sum_{i<j} \alpha_{ij}(S_i + S_j)R_{ij}\}, \tag{12}$$

where α_i is the selection indicator for the phrase P_i, S_i is the salience scores of P_i, α_{ij} and R_{ij} is co-occurrence indicator and the similarity a pair of phrases (P_i, P_j) respectively. The similarity is

calculated with the Jaccard Index based method. In order to obtain coherent summaries with good readability, we add some constraints into the ILP framework. For details, please refer to Woodsend and Lapata (2012), Bing et al. (2015), and Li et al. (2015). The objective function and constraints are linear. Therefore the optimization can be solved by existing ILP solvers such as simplex algorithms (Dantzig and Thapa, 2006). In the implementation, we use a package called lp_solve[5].

3 Data Description

In this section, we describe the preparation process of the dataset. Then we provide some properties and statistics.

3.1 Background

The definition of the terminology related to the dataset is given as follows.[6]

Topic: A topic refers to an event and it is composed of a set of news documents from different sources.

Document: A news article describing some aspects of the topic. The set of documents in the same topic typically span a period, say a few days.

Category: Each topic belongs to a category. There are 6 predefined categories: (1) Accidents and Natural Disasters, (2) Attacks (Criminal/Terrorist), (3) New Technology, (4) Health and Safety, (5) Endangered Resources, and (6) Investigations and Trials (Criminal/Legal/Other).

Aspect: Each category has a set of predefined aspects. Each aspect describes one important element of an event. For example, for the category "Accidents and Natural Disasters", the aspects are "WHAT", "WHEN", "WHERE", "WHY", "WHO_AFFECTED", "DAMAGES", and "COUNTERMEASURES".

Aspect facet: An aspect facet refers to the actual content of a particular aspect for a particular topic. Take the topic "Malaysia Airlines Disappearance" as an example, facets for the aspect "WHAT" include "missing Malaysia Airlines Flight 370", "two passengers used passports stolen in Thailand from an Austrian and an Italian." etc. Facets for the aspect "WHEN" are " Saturday morning",

"about an hour into its flight from Kuala Lumpur", etc.

Comment: A piece of text written by a reader conveying his or her altitude, emotion, or any thought on a particular news document.

3.2 Data Collection

The first step is to select topics. The selected topics should be in one of the above categories. We make use of several ways to find topics. The first way is to search the category name using Google News. The second way is to follow the related tags on Twitter. One more useful method is to scan the list of event archives on the Web, such as earthquakes happened in 2017 [7].

For some news websites, in addition to provide news articles, they offer a platform to allow readers to enter comments. Regarding the collection of news documents, for a particular topic, one consideration is that reader comments can be easily found. Another consideration is that all the news documents under a topic must be collected from different websites as far as possible. Similar to the methods used in DUC and TAC, we also capture and store the content using XML format.

Each topic is assigned to 4 experts to conduct the summary writing. The task of summary writing is divided into two phases, namely, aspect facet annotation, and summary generation. For the aspect facet annotation, the experts read and digested all the news documents and reader comments under the topic. Then for each aspect, the experts identified the related facets from news document and reader comments, and stored them in XML format. The summaries were generated based on the annotated aspect facets. When selecting facets, one consideration is those facets that are popular in both news documents and reader comments have higher priority. Next the facets that are popular in news documents have the next priority. The generated summary should cover all the aspects of a topic if such information can be found in the documents. The summary should be well-organized using complete sentences with a length restriction of 100 words.

After finishing the summary writing procedure, we employed another expert for scrutinizing the summaries. Each summary is checked from five linguistic quality perspectives: grammaticality, non-redundancy, referential clarity, focus, and co-

---

[5]http://lpsolve.sourceforge.net/5.5/

[6]In fact, for the core terminology, namely, topic, document, category, and aspect, we follow the MDS task in TAC (https://tac.nist.gov//2011/Summarization/Guided-Summ.2011.guidelines.html).

[7]https://en.wikipedia.org/wiki/Category:2017_earthquakes

herence. Finally, all the model summaries are stored in XML files.

3.3 Data Properties

The dataset contains 45 topics from those 6 pre-defined categories. Some examples of topics are "Malaysia Airlines Disappearance", "Flappy Bird", "Bitcoin Mt. Gox", etc. All the topics and categories are listed in Appendix A. Each topic contains 10 news documents and 4 model summaries. The length limit of the model summary is 100 words (slitted by space). On average, each topic contains 215 pieces of comments and 940 comment sentences. Each news document contains an average of 27 sentences, and each sentence contains an average of 25 words. 85% of non-stop model summary terms (entities, unigrams, bigrams) appeared in the news documents, and 51% of that appeared in the reader comments. The dataset contains 19k annotated aspect facets.

4 Experimental Setup

4.1 Dataset and Metrics

The properties of our own dataset are depicted in Section 3.3. We use ROUGE score as our evaluation metric (Lin, 2004) with standard options[8]. F-measures of ROUGE-1, ROUGE-2 and ROUGE-SU4 are reported.

4.2 Comparative Methods

To evaluate the performance of our dataset and the proposed framework **RAVAESum** for RA-MDS, we compare our model with the following methods:

- **RA-Sparse** (Li et al., 2015): It is a framework to tackle the RA-MDS problem. A sparse-coding-based method is used to calculate the salience of the news sentences by jointly considering news documents and reader comments.

- **Lead** (Wasson, 1998) : It ranks the news sentences chronologically and extracts the leading sentences one by one until the length limit.

- **Centroid** (Radev et al., 2000): It summarizes clusters of news articles automatically grouped by a topic detection system, and then

it uses information from the centroids of the clusters to select sentences.

- **LexRank** (Erkan and Radev, 2004) and **TextRank** (Mihalcea and Tarau, 2004): Both methods are graph-based unsupervised framework for sentence salience estimation based on PageRank algorithm.

- **Concept** (Bing et al., 2015): It generates abstractive summaries using phrase-based optimization framework with concept weight as salience estimation. The concept set contains unigrams, bigrams, and entities. The weighted term-frequency is used as the concept weight.

We can see that only the method RA-Sparse can handle RA-MDS. All the other methods are only for traditional MDS without comments.

4.3 Experimental Settings

The input news sentences and comment sentences are represented as BoWs vectors with dimension $|V|$. The dictionary V is created using unigrams, bigrams and named entity terms. n_d and n_c are the number of news sentences and comment sentences respectively. For the number of latent aspects used in data reconstruction, we let $m = 5$. For the neural network framework, we set the hidden size $d_h = 500$ and the latent size $K = 100$. For the parameter λ_p used in comment weight, we let $\lambda_p = 0.2$. Adam (Kingma and Ba, 2014) is used for gradient based optimization with a learning rate 0.001. Our neural network based framework is implemented using Theano (Bastien et al., 2012) on a single GPU[9].

5 Results and Discussions

5.1 Results on Our Dataset

The results of our framework as well as the baseline methods are depicted in Table 1. It is obvious that our framework RAVAESum is the best among all the comparison methods. Specifically, it is better than RA-Sparse significantly ($p < 0.05$), which demonstrates that VAEs based latent semantic modeling and joint semantic space reconstruction can improve the MDS performance considerably. Both RAVAESum and RA-Sparse are better than the methods without considering reader comments.

---

[8]ROUGE-1.5.5.pl -n 4 -w 1.2 -m -2 4 -u -c 95 -r 1000 -f A -p 0.5 -t 0

[9]Tesla K80, 1 Kepler GK210 is used, 2496 Cuda cores, 12G GDDR5 memory.

Table 1: Summarization performance.

System	R-1	R-2	R-SU4
Lead	0.384	0.110	0.144
TextRank	0.402	0.122	0.159
LexRank	0.425	0.135	0.165
Centroid	0.402	0.141	0.171
Concept	0.422	0.149	0.177
RA-Sparse	0.442	0.157	0.188
RAVAESum	**0.443***	**0.171***	**0.196***

Table 2: Further investigation of RAVAESum.

System	R-1	R-2	R-SU4
RAVAESum-noC	0.437	0.162	0.189
RAVAESum	**0.443***	**0.171***	**0.196***

5.2 Further Investigation of Our Framework

To further investigate the effectiveness of our proposed RAVAESum framework, we adjust our framework by removing the comments related components. Then the model settings of RAVAESum-noC are similar to VAESum (Li et al., 2017). The evaluation results are shown in Table 2, which illustrate that our framework with reader comments RAVAESum is better than RAVAESum-noC significantly($p < 0.05$).

Moreover, as mentioned in VAESum (Li et al., 2017), the output aspect vectors contain the word salience information. Then we select the top-10 terms for event "Sony Virtual Reality PS4", and "'Bitcoin Mt. Gox Offlile'" for model RAVAESum (+C) and RAVAESum-noC (-C) respectively, and the results are shown in Table 3. It is obvious that the rank of the top salience terms are different. We check from the news documents and reader comments and find that some terms are enhanced by the reader comments successfully. For example, for the topic "Sony Virtual Reality PS4", many readers talked about the product of "Oculus", hence the word "oculus" is assigned a high salience by our model.

5.3 Case Study

Based on the news and comments of the topic "Sony Virtual Reality PS4", we generate two summaries with our model considering comments (RAVAESum) and ignoring comments (RAVAESum-noC) respectively. The summaries and ROUGE evaluation are given in Table 4. All

the ROUGE values of our model considering comments are better than those ignoring comments with large gaps. The sentences in ***italic bold*** of the two summaries are different. By reviewing the comments of this topic, we find that many readers talked about "Oculus", the other product with virtual reality techniques. This issue is well identified by our model and select the sentence "*Mr. Yoshida said that Sony was inspired and encouraged to do its own virtual reality project after the enthusiastic response to the efforts of Oculus VR and Valve, another game company working on the technology.*".

6 Conclusions

We investigate the problem of reader-aware multi-document summarization (RA-MDS) and introduce a new dataset. To tackle the RA-MDS, we extend a variational auto-encodes (VAEs) based MDS framework by jointly considering news documents and reader comments. The methods for data collection, aspect annotation, and summary writing and scrutinizing by experts are described. Experimental results show that reader comments can improve the summarization performance, which demonstrate the usefulness of the proposed dataset.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.

Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian Goodfellow, Arnaud Bergeron, Nicolas Bouchard, David Warde-Farley, and Yoshua Bengio. 2012. Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*.

Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca Passonneau. 2015. Abstractive multi-document summarization via phrase selection and merging. In *ACL*, pages 1587–1597.

George B Dantzig and Mukund N Thapa. 2006. *Linear programming 1: introduction*. Springer Science & Business Media.

Günes Erkan and Dragomir R Radev. 2004. Lexpagerank: Prestige in multi-document text summarization. In *EMNLP*, volume 4, pages 365–371.

Jade Goldstein, Vibhu Mittal, Jaime Carbonell, and Mark Kantrowitz. 2000. Multi-document summarization by sentence extraction. In *NAACL-ANLPWorkshop*, pages 40–48.

Table 3: Top-10 terms extracted from each topic according to the word salience values

Topic	±C	Top-10 Terms
"Sony Virtual	−C	Sony, headset, game, virtual, morpheus, reality, vr, project, playstation, Yoshida
Reality PS4"	+C	Sony, game, vr, virtual, headset, reality, morpheus, *oculus*, project, playstation
"Bitcoin Mt.	−C	bitcoin, gox, exchange, mt., currency, Gox, virtual, company, money, price
Gox Offile"	+C	bitcoin, currency, money, exchange, gox, mt., virtual, company, price, world

Table 4: Generated summaries for the topic "Sony Virtual Reality PS4".

System	R-1	R-2	R-SU4
RAVAESum-noC	0.482	0.184	0.209
A virtual reality headset that's coming to the PlayStation 4. *Today announced the development of "Project Morpheus" (Morpheus) "a virtual reality (VR) system that takes the PlayStation4 (PS4)".* Shuhei Yoshida, president of Sony Computer Entertainment, revealed a prototype of Morpheus at the Game Developers Conference in San Francisco on Tuesday. Sony showed off a prototype device V called Project Morpheus V that can be worn to create a virtual reality experience when playing games on its new PlayStation 4 console. *The camera on the Playstation 4 using sensors that track the player's head movements.*			
RAVAESum	**0.490**	**0.230**	**0.243**
Shuhei Yoshida, president of Sony Computer Entertainment, revealed a prototype of Morpheus at the Game Developers Conference in San Francisco on Tuesday. A virtual reality headset that's coming to the PlayStation 4. Sony showed off a prototype device V called Project Morpheus V that can be worn to create a virtual reality experience when playing games on its new PlayStation 4 console. *Mr. Yoshida said that Sony was inspired and encouraged to do its own virtual reality project after the enthusiastic response to the efforts of Oculus VR and Valve, another game company working on the technology.*			

Meishan Hu, Aixin Sun, and Ee-Peng Lim. 2008. Comments-oriented document summarization: Understanding documents with readers' feedback. In *SIGIR*, pages 291–298.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *ICLR*.

Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *ICLR*.

Piji Li, Lidong Bing, Wai Lam, Hang Li, and Yi Liao. 2015. Reader-aware multi-document summarization via sparse coding. In *IJCAI*, pages 1270–1276.

Piji Li, Zihao Wang, Wai Lam, Zhaochun Ren, and Lidong Bing. 2017. Salience estimation via variational auto-encoders for multi-document summarization. In *AAAI*, pages 3497–3503.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*, volume 8.

Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. In *EMNLP*, pages 1412–1421.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into texts. Association for Computational Linguistics.

Ziheng Lin Min, Yen Kan Chew, and Lim Tan. 2012. Exploiting category-specific information for multi-document summarization. *COLING*, pages 2093–2108.

Ani Nenkova and Kathleen McKeown. 2012. A survey of text summarization techniques. In *Mining Text Data*, pages 43–76. Springer.

Dragomir R Radev, Hongyan Jing, and Malgorzata Budzikowska. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic summarization*, pages 21–30. Association for Computational Linguistics.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. 2014. Stochastic backpropagation and approximate inference in deep generative models. In *ICML*, pages 1278–1286.

Xiaojun Wan, Jianwu Yang, and Jianguo Xiao. 2007. Manifold-ranking based topic-focused multi-document summarization. In *IJCAI*, volume 7, pages 2903–2908.

Mark Wasson. 1998. Using leading text for news summaries: Evaluation results and implications for commercial summarization applications. In *ACL*, pages 1364–1368.

Kristian Woodsend and Mirella Lapata. 2012. Multiple aspect summarization using integer linear programming. In *EMNLP-CNLL*, pages 233–243.

Zi Yang, Keke Cai, Jie Tang, Li Zhang, Zhong Su, and Juanzi Li. 2011. Social context summarization. In *SIGIR*, pages 255–264.

Appendices

A All The Topics

Table 5: All the topics and the corresponding categories. The 6 predefined categories are: (1) Accidents and Natural Disasters, (2) Attacks (Criminal/Terrorist), (3) New Technology, (4) Health and Safety, (5) Endangered Resources, and (6) Investigations and Trials (Criminal/Legal/Other).

Topic	Category
Boston Marathon Bomber Sister Arrested	6
iWatch	3
Facebook Offers App With Free Access in Zambia	3
441 Species Discovered in Amazon	5
Beirut attack	2
Great White Shark Choked by Sea Lion	1
Sony virtual reality PS4	3
Akademik Shokalskiy Trapping	1
Missing Oregon Woman Jennifer Huston Committed Suicide	6
Bremerton Teen Arrested Murder 6-year-old Girl	6
Apple And IBM Team Up	3
California Father Accused Killing Family	6
Los Angeles Earthquake	1
New Species of Colorful Monkey	5
Japan Whaling	5
Top Doctor Becomes Latest Ebola Victim	4
New South Wales Bushfires	1
UK David Cameron Joins Battle Against Dementia	4
UK Cameron Calls for Global Action on Superbug Threat	4
Karachi Airport Attack	2
Air Algerie Plane Crash	1
Flappy Bird	3
Moscow Subway Crash	1
Rick Perry Lawyers Dismissal of Charges	6
New York Two Missing Amish Girls Found	6
UK Contaminated Drip Poisoned Babies	4
Taiwan Police Evict Student Protesters	2
US General Killed in Afghan	5
Monarch butterflies drop	5
UN Host Summit to End Child Brides	4
Two Tornadoes in Nebraska	1
Global Warming Threatens Emperor Penguins	5
Malaysia Airlines Disappearance	1
Google Conference	3
Africa Ebola Out of Control in West Africa	4
Shut Down of Malaysia Airlines mh17	1
Sochi Terrorist Attack	2
Fire Phone	3
ISIS executes David Haines	2
UK Rotherham 1400 Child Abuse Cases	6
Rare Pangolins Asians eating Extinction	5
Kunming Station Massacre	2
Bitcoin Mt. Gox	3
UK Jimmy Savile Abused Victims in Hospital	6
ISIS in Iraq	2