Neurocomputing 253 (2017) 6–14

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# Multi-factors based sentence ordering for cross-document fusion from multimodal content

Lin Yue [a,b,c,d], Zhenkun Shi [c,f], Jiayu Han [c,f], Sen Wang [e], Weitong Chen [d], Wanli Zuo [c,f,*]

[a] School of Computer Science and Information Technology, Northeast Normal University, Changchun 130024, China
[b] School of Environment, Northeast Normal University, Changchun 130024, China
[c] Key Laboratory of Symbolic Computation and Knowledge Engineering, Ministry of Education, Jilin University, Changchun, 130012, P.R. China
[d] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
[e] School of Information and Communication Technology, Griffith University, Australia
[f] College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

ARTICLE INFO

*Article history:*
Received 29 April 2016
Revised 10 September 2016
Accepted 7 December 2016
Available online 8 March 2017

*Keywords:*
Sentence ordering
Document fusion
Continuous Hopfield neural network(CHNN)

ABSTRACT

Organizing a coherent structure of the sentences extracted from multiple documents, guarantees the fluency and readability of the fused document. In this paper, sentence ordering problem is treated as a combinatorial optimization problem and solved with continuous Hopfield neural network (CHNN). We unify the existing factors by considering the most frequent orders temporal information, and topical relevance between local themes during overall ordering process. Specifically, ordering algorithm traverses all the local themes and locates a shortest path as the final sentence ordering. We show the results with data from Document Understanding Conferences (DUC) 2002–2005, and demonstrate the effectiveness of the developed approach compared with Random Ordering (RO), Chronological Ordering (CO), Majority Ordering (MO), and Precedence Relation Ordering (PRO).



## 1. Introduction

A system of interlinked hypertext documents accessed via the Internet constitutes the World Wide Web. With a web browser, one can view web pages that may contain videos, images, text, and other form of multimedia; and it allows us to navigate different content via hyperlinks. As the online-corpus is gigantic in its volume, Web search engines often return more results than actual needs. Navigation through all returned Web documents to obtain targeted information is infeasible and tedious burden, thus the automatic document summarization has been proposed for salient information retrieval [1,2,3,4] and high-efficiency knowledge acquisition, which aims to produce a shortest description containing the most important information within all documents. Document fusion as a relevant research, aims to produce a shortest description with all information contained in the document sets, but without repetition [5,6]. The significant difference is that, the former is like the intersection of document set, and the latter is the union of document set.

Both automatic document summarization and document fusion tackles the information overload problem in heterogeneous Web resources by providing a condensed and comprehensive version of a set of documents. Several key sub-tasks are involved in the research areas, such as redundancy removal [7], topic detection [8,9], sentence retraction [10], objects merging in document set [11,12], ordering sentence from different sources for keeping the logical and grammatical structure correct. Among all these extra tasks, sentence ordering is mandatory to compose sentences extracted from multiple documents into a coherent structure, which guarantees the fluency and readability of the results. The correct order of these sentences is helpful for understanding of the input articles. Moreover, the problem of information ordering is not limited to the areas mentioned hereinbefore, and concerns natural language generation (NLG) [13] applications such as in discourse planning and sentence aggregation [14,15], which are important components of NLG. Besides, a brief, well-organized, fluent answer to a need for information at the specified level of granularity is also applicable in real-world question answering system, which is a classical application in social search. While it is trivial to order sentence from one single document, usually the extracted sentences are arranged as same order as in the original documents. The problem of sentence ordering for summarization or document fusion has received relatively little attention. The case we focus on is how to arrange the sentences extracted from different documents under a particular topic. It is a very important, but also potentially a very challenge task. As the sentences from source articles for ordering are

* Corresponding author at: College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China.
*E-mail address:* yuel563@163.com (W. Zuo).

http://dx.doi.org/10.1016/j.neucom.2016.12.084
0925-2312/

written by different groups, from different viewpoints, or have different writing style, etc. There is not only the problem of subjective factors to deal with, but also the problem such as detecting rhetorical relations existing between sentences. It is difficult and unsolved in sentence ordering task. However, inferring a coherent ordering of extracted sentences with rhetorical structure analysis is not yet achievable.

Existing work on sentence ordering can be classified into two types: temporal information processing [16–18], and natural order learning in original corpora [19-23]. Among these methods, temporal information processing is the bottleneck technology which affects the quality of the ordering algorithm, and the fact is that, not all the corpora have temporal information. Although newspaper articles are assigned time stamp of its publication date, they are just publication date without exact time as hour and minute. As a result, chronological ordering based method could not achieve steady and high-quality ordering effects. For natural order learning based method, the factor of temporal information is ignored because temporal information processing is very difficult, experimental results show a significant improvement over existing sentence ordering strategies, but more complicated strategies are usually involved. To further explore ways of sentence ordering, we propose an improved ordering method, which not only examines most frequent orders in original document (i.e., majority based method) and orders sentence by publication date (i.e., temporal information based method), but also considers the topical relevance between local themes during overall ordering process. Specifically, we refine ordering problem based on majority ordering and chronological ordering with continuous Hopfield neural network (CHNN) [24], which demands ordering algorithm traversing all the local themes once and searching a shortest path. Hopfield neural network were widely used on combinatorial optimization problems, which transforms the objective function in optimization problems into energy functions in neural network, and maps variables of practical problems into the state of network. It is definitely a meaningful attempt to apply CHNN on our problem.

This paper is organized as follows: In Section 2, we introduce related work preliminaries. In Section 3, we propose our framework: continuous Hopfield neural network based ordering (CHNNO). In Section 4, metrics for semi-automatic evaluation and subjective grading are described. Next, in Section 5, experiments and corresponding results are shown. In Section 6, the conclusion and future work to the proposed work are presented.

## 2. Related. work

Existing work on sentence ordering can be classified into two types: temporal information processing [16-18], and natural order learning in original corpora [19-23,25-27]. Paper [16] first proposed ordering method based on publication date of the sentences contained in descriptions of novel events. On this basis, paper [17] improved the strategy for ordering information that combines constraints from chronological order of events and topical relatedness. Then, paper [18] proposed a method to improve chronological ordering by resolving precedent information of arranging sentences. The first type usually assumes sentences to be semantically independent and only make use of the time feature in original documents, which usually leads to unsatisfactory results. Although it is an effective heuristic ordering sentences according to the publication date, as the trends of an event changes over time. However, such temporal features may be not available in all cases as temporal inference in documents is still a problem, i.e., temporal features such as *yesterday* or *tomorrow* are just relative concept to describing temporal information. The second type usually involves more complicated processing.

Paper [19] proposed an approach to information ordering that is particularly suited for text-to-text generation, where a model learns constraints on the sentence order from a corpus of domain-specific texts and an algorithm yields the most likely order among several alternatives. Paper [20] proposed a sentence ordering algorithm using a semi-supervised sentence classification and historical ordering strategy, where the classification is based on the manifold structure underlying sentences, addressing the problem of limited labeled data, and the historical ordering helps to ensure topic continuity and avoid topic bias. Paper [21] considered the problem of modeling the content structure of texts within a specific domain, utilizing a novel adaptation of algorithms for Hidden Markov Models. Paper [22] presented a bottom-up approach to arranging sentences extracted for multi-document summarization, where four criteria, chronology, topical-closeness, precedence, and succession are defined, and the criteria are integrated into a criterion by a supervised learning approach, then repeatedly concatenating two textual segments into one segment based on the criterion until obtain the overall segment with all sentences arranged. Then Paper [23] further modeled the problem of sentence ordering as a one of learning the optimal combination of preference experts that determine the ordering between two given sentences, where five preference experts: chronology, probabilistic, topical-closeness, precedence, and succession are defined to capture the preference of a sentence against another sentence. Recently, more natural order learning based methods have been proposed [25-27]. The general ideas of two sentence ordering methods that related to our method are introduced in the following section, which are the basis of our ordering method.

### 2.1. Chronological ordering

Given a collection of texts from a particular theme, chronological ordering (CO) arranges sentences basing on the publication date, which is applicable for processing corpus of newspaper articles containing similar theme and temporal information. These articles usually describe different phases of the same event; hence common information is involved in these documents. Based on the baseline of organizing news articles, the content in newspaper articles is organized by a series of corresponding context, subsequent descriptions and comment on the novel events. For this reason, it is effective heuristic to order sentences with their publication date.

This kind of method can boil down to a question of local theme ordering, specifically, the position of every sentence selected for generating summarization or fusion result called summarization sentence or fusion sentence, is determined by the position of local theme it belong to. Therefore, time tagging for every local theme is obliged here. Assume each newspaper article in original corpora is tagged with date, hour and minute information of publication time, and there are not two articles with same time stamp, then we have ordering procedure and the outline of chronological ordering algorithm as shown in Fig. 1.

### 2.2. Majority ordering

Another strategy is majority order (MO), in which each summary sentence or fusion sentence is mapped to a theme, i.e., a set of similar sentences in the documents, and the order of these sentences determines that for summary sentences or fusion sentences. To do that a notation of precedence relation will be defined:

**Definition 1** (*Precedence Relation*)**.** Consider local theme $A$ and $B$ (say segment $A$ and $B$ here), if local theme $A$ precedes local theme $B$, then summary sentences or fusion sentences in local theme $A$ precedes summary sentences or fusion sentences in local theme $B$,

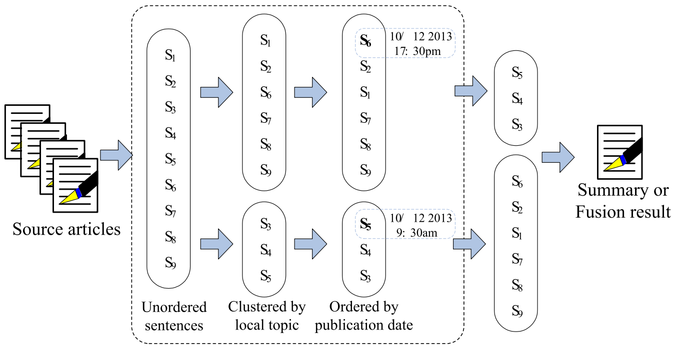



**Fig. 1.** The outline of chronological ordering algorithm.

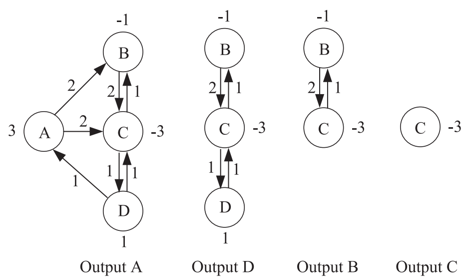



**Fig. 2.** The process of majority ordering algorithm.

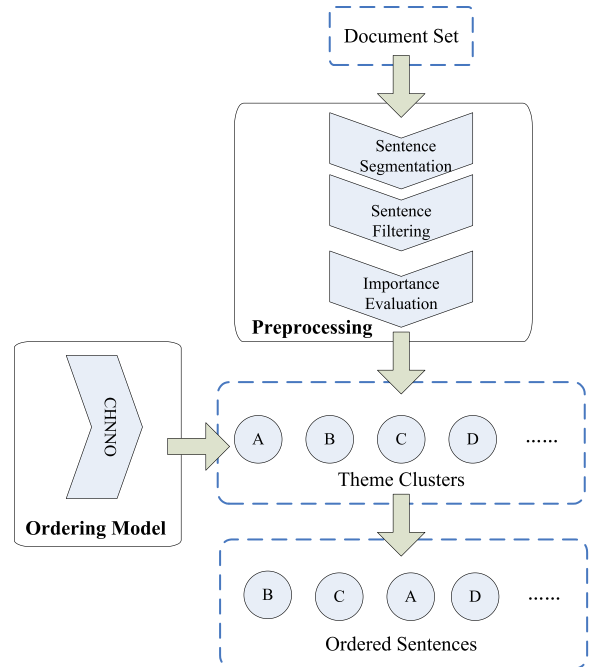



**Fig. 3.** An overview of sentence ordering process with CHNNO.

which is denoted as:

$$A \succ B \tag{1}$$

Where we assume there is a transitive relation among the sequence of ordered sentence of different local themes (If $A \succ B$ and $B \succ C$, then $A \succ C$), then the final summarization or fusion results is a linear distribution ($A, B, \ldots, N$). However, this ideal assumption was rejected due to lack of realistic evidence. Take three relations existing between local theme $A$, $B$ and $C$ as an example, if $A \succ B$, $B \succ C$ and $C \succ A$, we will get contradictory conclusions of ($A$, $B$, $C$) and ($C$, $A$). The reason for contradictory conclusions is that the order of sequences of local themes are not transitive, so high-quality ordering results cannot be achieved by aforementioned method. To address this issue, local majority principle is proposed [28].

**Definition 2** (*Majority Principle*)**.** Consider local theme $A$ and $B$ (say segment $A$ and $B$ here), the number of documents that satisfy $A \succ B$ in document set is denoted as:

$$A \propto B \quad \text{or} \quad \alpha_{AB} = A \propto B \tag{2}$$

Where if $(A \propto B) > (B \propto A)$, then segment $A$ precedes segment $B$, otherwise, segment $B$ precedes segment $A$.

Given two definitions above, each local theme could be identified as a node in directed graph, in which a directed edge indicates precedence relation between two local themes, and weight on specific edge captures the frequency of precedence relation from one local theme to another. Then, ordering problem by precedence relation here is equal to a shortest path problem.

The problem is that, for each local theme in all documents with a relatively fixed position, the arrangement showed good readability; while for each local theme in all documents without a relatively fixed position, the result showed bad readability. This problem derives from the process itself of majority ordering algorithm as shown in Fig. 2. Specifically, every time the algorithm selects a node (a local theme) with biggest weight and meantime deletes this node and related edges in directed graph. Next step is to recalculate the weight of the nodes in current directed graph, which is only related to the rest of nodes without considering the previous nodes. That is to say, the overall connection among the nodes is ignored during this process. To address this issue, an improved ordering algorithm combining majority and chronological relation is proposed in next section.

## 3. Ordering. refinement by continuous hopfield neural network

We propose an improved ordering method, which not only examines most frequent orders in original document (majority based method) and orders sentence by publication date (temporal information based method), but also the topical relevance between local themes is considered during overall ordering process, which is based on the aforementioned two algorithms in Section 2. Here, we refine ordering problem based on majority ordering and chronological ordering with Hopfield neural network [24], which demands ordering algorithm traversing all the local themes once and searching a shortest path. The general idea of sentence ordering method proposed in this paper contains two parts: preprocessing and ordering model as shown in Fig. 3.

### 3.1. Preprocessing

**Sentence segmentation and sentence filtering**. At this step, we select out the candidate sentences from a document set for document fusion task. Detecting sentence boundary in our work is based on punctuation. That is, all the textual content ended up with question marks, exclamation marks, a full stop or suspension points is treated as a sentence. Among these sentences, long sentences and short sentences will be removed out, and that the sentence with modest length is the candidate sentences we need. Moreover, the sentence ended up with question marks will also be eliminated, as declarative content is needed for document fusion task.

**Importance evaluation**. In vector space model, candidate sentences are represented as $k$-dimensional vectors, where $k$ is the number of the words. We take *TF-IDF* to calculate the weight of each term, where TF is the absolute frequency of the term; IDF is used to weigh the frequency of the terms in each document with the factor that discounts its importance when it appears in

$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline A & 1 & 0 & 0 & 0 \\ B & 0 & 0 & 1 & 0 \\ C & 0 & 1 & 0 & 0 \\ D & 0 & 0 & 0 & 1 \end{array}$$

**Fig. 4.** The permutation matrix.

many documents. The similarity between two sentences can be achieved by calculating the similarity between two sentence vectors as shown below:

$$sim(\vec{d_i}, \vec{d_j}) = \frac{\vec{d_i} \cdot \vec{d_j}}{\left|\vec{d_i}\right|\left|\vec{d_j}\right|} \tag{3}$$

For further refining the candidate sentences, we introduce the another concept below:

**Definition 3** (*Theme Center Vector*)**.** Theme center vector is used to describe the theme of a whole document set, which is arithmetic mean value of the candidate sentences.

We can perform sentence clustering by calculating the similarity between each two sentence vectors with above-mentioned formula. After this step, we will get many theme clusters, and each cluster consists of a few sentences. For each cluster, repeated sentences with high similarity value will be removed out. Besides, we compare the similarity between each cluster and theme center, and the cluster that is far from the theme center will be removed out. Finally, the rest of the theme clusters and "distance" among these theme clusters will be processed by continuous Hopfield neural network.

### 3.2. Ordering model

Hopfield neural network were widely used on combinatorial optimization problems, which transforms the objective function in optimization problems into energy functions in neural network, and maps variables for the problems into the state of network. It is definitely a meaningful attempt to apply CHNN on our problem. The general procedures are:

**Problem analyzing:** make the output of the network mapping the solution of practical problem.

**Constructing energy function:** make the minimum value of energy function mapping the optimal solution of the problem.

**Designing network structure:** design parameters with energy function and network stability condition, and get dynamic equation.

We propose an improved ordering method, which demands ordering algorithm traversing all the local themes once and searching a shortest path. The sentences in each theme cluster are ordered with chronological information; and CHNN will be used to find out a path from theme clusters. Each local theme could be identified as a node in directed graph, in which a directed edge indicates relation between two local themes and weight on specific edge captures the distance from one local theme to another. In this paper, we use "distance" to denote the "weight" between every two theme clusters, which are the inputs to the Hopfield network.

The permutation matrix of $N \times N$ dimension ($N$ is the number of local themes) is shown in Fig. 4, which illustrates the final result of network evolution.

Where *A, B, C* and *D* (each line) denotes the local themes, and 1, 2, 3 and 4 (each column) denotes the path ordering. In Fig. 4 example the path ordering is $A \rightarrow C \rightarrow B \rightarrow D$. Note that: for each line and each column, there is only one 1 and the rest of elements are 0; all the 1 in permutation matrix is equal to $N$.

The permutation matrix of $N \times N$ dimension is identified as a continuous Hopfield neural network (*CHNN*) with $N \times N$ neurons, where the state of each neuron corresponds to the value of each element in permutation matrix when the whole network gets stable. Moreover, the distance among each local theme is treated as the constraint condition to determine the strength of the connections $w_{ij}$ among the neurons. The result of this neural network evolution is anticipated to be the optimal solution, that is to say, the permutation matrix above is the optimal ordering with the shortest path. Then, energy function that is used to describe this constraint will be introduced below:

**Definition 3** (*Energy Function*)**.**

$$E = \frac{A}{2}\sum_{x=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N}V_{xi}V_{xj} + \frac{B}{2}\sum_{i=1}^{N}\sum_{x=1}^{N}\sum_{y=x}^{N}V_{xi}V_{yi} + \frac{C}{2}\left(\sum_{x=1}^{N}\sum_{i=1}^{N}V_{xi} - N\right)^2 + \frac{D}{2}\sum_{x=1}^{N}\sum_{y=1}^{N}\sum_{i=1}^{N}d_{xy}V_{xi}\left(V_{y,i+1} + V_{y,i-1}\right) \tag{4}$$

Where *A, B, C* and *D* denotes the weight; $d_{xy}$ denotes the distance from local theme $x$ to local theme $y$. The previous three items ($\frac{A}{2}\sum_{x=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N}V_{xi}V_{xj}$, $\frac{B}{2}\sum_{i=1}^{N}\sum_{x=1}^{N}\sum_{y=x}^{N}V_{xi}V_{yi}$ and $\frac{C}{2}(\sum_{x=1}^{N}\sum_{i=1}^{N}V_{xi} - N)^2$) in above equation are constraint items; the last one ($\frac{D}{2}\sum_{x=1}^{N}\sum_{y=1}^{N}\sum_{i=1}^{N}d_{xy}V_{xi}(V_{y,i+1} + V_{y,i-1})$) is optimization objective.

**Definition 4** (*Improved Energy Function*)**.**

$$E = \frac{A}{2}\sum_{x=1}^{N}\left(\sum_{i=1}^{N}V_{xi} - 1\right)^2 + \frac{A}{2}\sum_{i=1}^{N}\left(\sum_{x=1}^{N}V_{xi} - 1\right)^2 + \frac{D}{2}\sum_{x=1}^{N}\sum_{y=1}^{N}\sum_{i=1}^{N}V_{xi}d_{xy}V_{y,i+1} \tag{5}$$

**Definition 5** (*Dynamic Equation*)**.**

$$\frac{dU_{xi}}{dt} = -\frac{\partial E}{\partial V_{xi}} = -A\left(\sum_{i=1}^{N}V_{xi} - 1\right) - A\left(\sum_{y=1}^{N}V_{yi} - 1\right) - D\sum_{y=1}^{N}d_{xy}V_{y,i+1} \tag{6}$$

Corresponding algorithm procedure which uses continuous Hopfield neural network to order sentences is given below:

Step 1: Set initial value and weight, i.e., $t=0$, $A=1.5$, $D=1.0$, $U_0=0.02$;
Step 2: Read in the distance between all the local themes, i.e., $d_{xy}(x, y = 1, 2, ..., N)$;
Step 3: Initialize $U_{xi}(t)$ ($U_{xi}(t) = U_0' + \delta_{xi}$ $(x, i = 1, 2, ..., N)$)for the neural network, where $U_0' = \frac{1}{2}U_0\ln(N-1)$, $N$ is the number of local themes, and $\delta_{xi}$ is a random value between the interval $(-1, +1)$);
Step 4: Calculate $\frac{dU_{xi}}{dt}$ by using dynamic equation mentioned in Definition 5;
Step 5: Calculate $U_{xi}(t+1)$ with the first-order Euler method ($U_{xi}(t+1) = U_{xi}(t) + \frac{dU_{xi}}{dt}\Delta T$);
Step 6: Calculate $V_{xi}(t)$ with sigmoid function ($V_{xi}(t) = \frac{1}{2}(1 + \tanh(\frac{U_{xi}(t)}{U_0}))$);
Step 7: Calculate energy function $E$;
Step 8: Check the route validity, go back to Step 4 if iteration is not over;
Step 9: Output the number of iterations, optimal path, and energy function, the length of path and energy changes.

## 4. Evaluation. measures

### 4.1. Metrics for semi-automatic evaluation

Assessing the quality of sentence ordering generated by an algorithm is a non-trivial task. Three semi-automatic evaluation measures that have been used in previous work [23] are employed in this paper, which compare a sentence ordering produced by an algorithm against the ordering produced by human annotator. They are evaluation measures of rank correlation coefficients such as Spearman's rank correlation and Kendall's rank correlation, and evaluation measure of assessing continuity of pairwise sentences, which is called Average Continuity.

#### 4.1.1. Spearman's rank correlation

Spearman's rank correlation denoted by the Greek letter $\rho$ or as $r_{s,}$ is a nonparametric measure of statistical dependence between two orderings. It assesses how well the relationship between two orderings (say $\pi$ and $\sigma$) can be described using a monotonic function and is defined as follows.

$$\rho = 1 - \frac{6\sum_{i=1}^{N} (\pi(i) - \sigma(i))^2}{N(N^2 - 1)} \tag{7}$$

Where $\{S_{1,} S_{2,\ldots,} S_N\}$ is the set of $N$ sentences to be ranked, $\pi \in S_N$ and $\sigma \in S_N$ respectively denote the ordering produced by the algorithm and by standard ordering; while $\pi(i)$ and $\sigma(i)$ denote the position of sentence $S_i$ in $\pi$ and $\sigma$. Spearman's rank correlation, $\rho$, falls into range interval [−1,1].

#### 4.1.2. Kendall's rank correlation

Kendall rank correlation coefficient, commonly referred to as Kendall's ($\tau$) coefficient, is a statistic used to measure the association between two measured quantities. It is a non-parametric hypothesis test for statistical dependence based on the tau coefficient and defined as follows.

$$\tau = \frac{4C(\pi, \sigma)}{N(N-1)} - 1 \tag{8}$$

Where $\{s_1, s_2, ..., s_N\}$ is the set of $N$ sentences to be ranked, $C(\pi, \sigma)$ is the number of concordant pairs between $\pi$ and $\sigma$. Kendall rank correlation coefficient, $\tau$, falls into range interval [−1,1]. If the agreement between the two rankings is perfect (i.e., the two rankings are identical) the coefficient has value 1. If the disagreement between the two rankings is perfect (i.e., one ranking is the reverse of the other) the coefficient has value −1.

#### 4.1.3. Average continuity

In the fields of computational linguistics and probability, an $n$-gram is a contiguous sequence of $n$ items from a given sequence of text or speech, which is widely used in statistical natural language processing. Here, the quality of a sentence ordering can be estimated by the number of continuous sentences that are also reproduced in reference sentence ordering. This is equivalent to measuring a precision of continuous sentences in an ordering against the reference ordering. The precision of $n$ sentences in an ordering to be evaluated as follows.

$$P_n = \frac{m}{N - n + 1} \tag{9}$$

Where $N$ is the number of sentences in the reference ordering; $n$ is the length of continuous sentences on which we are evaluating; $m$ is the number of continuous sentences that appear in both the evaluation and reference orderings. The Average Continuity ($AC$) is defined as the logarithmic average of $P_n$ ($n$ from 2 to $k$).

$$AC = \exp\left(\frac{1}{k-1}\sum_{n=2}^{k} \log(P_n + \alpha)\right) \tag{10}$$

Where $k$ is a parameter to control the range of the logarithmic average; and $\alpha$ is a small value in case if $P_n$ is zero. We set $k = 4$ (i.e., more than five continuous sentences are not included for evaluation) and $\alpha = 0.001$. Average Continuity becomes 0 when evaluation and reference orderings share no continuous sentences and 1 when the two orderings are identical.

### 4.2. Subjective grading

Intrinsic evaluation where evaluation is done by human on accessing the quality of the ordered sentences itself is also involved in our work. Four levels of Perfect, Acceptable, Poor, and Unacceptable are used in subjective grading process, where the distribution of these subjective grading made by a number of judges to ordering algorithms on four datasets and corresponding details of grading process will be given in Section 5. The readability of the sentence sequence is assessed using five linguistic quality questions which measures qualities of the ordered sentences that do not involve comparison with a standard ordering. The linguistic qualities measured are Grammaticality, Non-redundancy, Referential clarity, Focus and Structure and Coherence, which is used to assess the readability of the summaries task [29]. As the main task in our paper is to properly order the candidate summary sentences, Q1, Q4 and Q5 should be badly considered during the subjective grading process. The assessing process will not take into account Q2 and Q3 which involve other processing methods such as redundancy removal and coreference resolution or anaphora resolution.

**Q1: Grammaticality.** The ordered sentences should have no obviously ungrammatical sentences that make the text difficult to read.

**Q2: Non-redundancy.** There should be no unnecessary repetition in the ordered sentences.

**Q3: Referential clarity.** It should be easy to identify who or what the pronouns and noun phrases in the summary are referring to.

**Q4: Focus.** The final sequence should have a focus, in which sentences should contain information that is related to the rest sentences.

**Q5: Structure and Coherence.** The sequence should be well-structured and well-organized. The result should not just be a heap of related information, but should build from sentence to sentence to a coherent body of information about a topic.

## 5. Experiments. and results

### 5.1. Data set

Document Understanding Conferences (DUC) [30] is sponsored by the Advanced Research and Development Activity (ARDA), the conference series is run by the National Institute of Standards and Technology (NIST) to further progress in summarization and enable researchers to participate in large-scale experiments. In 2008, DUC became a Summarization track in the Text Analysis Conference (TAC) [31]. A TAC cycle consists of a set tracks, areas of focus in which particular NLP tasks are defined. From then on, TAC's main tasks are other corresponding research works within the Natural Language Processing (Details shown in Table 1).

We tested our work using data from DUC 2002–2005, where generic multi-document summarization has been one of the fundamental tasks in DUC 2002 and DUC 2004 (i.e. task 2 in DUC 2002 and task 2 in DUC 2004). In DUC 2002, 59 document sets of approximately 10 documents each were provided and generic summaries of each document set with lengths of 100words or less were required to be created. In DUC 2004, 50 TDT (Topic Detection and Tracking) document clusters were provided and a short

**Table 1**
The DUC tasks for the particular years.

| Year | Task |
|---|---|
| DUC 2001–2004 | Single-document Summaries and Multiple Document Summaries. |
| DUC 2005–2007 | User-oriented, Question-focused Summarization task. |
| TAC 2008–2013 | Question Answering, Recognizing Textual Entailment, Summarization and Knowledge Base Population. |

**Table 2**
The datasets for generic multi-document summarization.

| | DUC 2002 | DUC 2004 |
|---|---|---|
| Task | Task 2 | Task 2 |
| Number of clusters | 59 | 50 |
| Data source | TREC-9 | TDT-2 |

**Table 3**
The datasets for topic-focused multi-document summarization.

| | DUC 2003 | DUC 2005 |
|---|---|---|
| Task | Task 3 | The only task |
| Number of clusters | 30 | 50 |
| Data source | TREC | TREC |

summary with lengths of 665 bytes or less was required to be created. Table 2 gives a short introduction of the datasets used in the experiment. Specially, the chronological information in our experiment comes from news videos related to corresponding textual news, which guarantees the correctness of chronological information.

Topic-focused multi-document summarization has been evaluated on task 3 of DUC 2003 and the only task of DUC 2005, each task having a gold standard dataset consisting of document clusters and reference summaries. The task 3 of DUC 2003 is to produce summaries focused by viewpoints; the task of DUC 2005 is to produce summaries focused by *DUC Topics*. Table 3 gives a short introduction of the datasets.

### 5.2. Experimental results

For comparison purpose, we tested five classical sentence ordering methods based on the dataset introduced in previous section, which are Random Ordering method, Chronological Ordering method, Majority Ordering method, Precedence Relation Ordering method and Continuous Hopfield Neural Network Ordering method proposed in our paper (see Table 4).

Among these comparison methods, Precedence Relation Ordering (PRO) has got significant improvement compared with Random ordering (RO), Probalistic Ordering (PO) [19], Chronological Ordering (CO) and another two related works of themselves on 3rd Text Summarization Challenge (TSC-3) corpus [32], where the multiple document summarization task is organized by NTCIR project [33]. TSC-3 dataset contains multi-document summaries for 30 news events which are selected by the organizers of TSC task. For each topic, a set of Japanese newspaper articles are selected from Mainichi Shinbun and Yomiuri Shinbun, two popular Japanese newspapers, which is with their annotated publication date and not revised or modified once an article is published. Although the author has pointed out that there are no fundamental differences between Japanese and English for experiments, it is difficult for us to compare our work with these experimental results. Therefore, in our work, English corpus is used, with which corresponding methods are re-tested.

Tables 5–8 show the performance of different sentence ordering methods on four DUC datasets, in which the mean value of three metrics of Spearman's rank correlation, Kendall's rank correlation and Average Continuity are used to this rating task. From the four tables, it is obvious that RO is the worst ordering method with all metrics. As just mentioned, RO is the lowest standard, in which sentences are arranged randomly. This method is treated as the lower-baseline and used to indicate the performance of other methods that we would obtain. The performance of CO and MO is better than RO, but not as good as CHNNO and PRO. As the basis of ordering methods, CO and MO failed to show significant differences on all three metrics. PRO, as the highest standard except standard sentences arrangement, improved chronological ordering by resolving antecedent sentences of arranged sentences combining the refinement algorithm with topical segmentation, has achieved steady performance comparing with RO, CO and MO. However, CHNNO proposed in this paper is superior to PRO on some metrics, which reveals the improvement of our idea. Specifically, on DUC 2002, CHNNO is superior on the metrics of Spearman's rank correlation and Average Continuity with value 0.801 and 0.772; on DUC 2004, CHNNO is superior on the metric of Kendall's rank correlation with value 0.797; on DUC 2003, CHNNO is superior on the metrics of Kendall's rank correlation and Average Continuity with value 0.643 and 0.676; on DUC 2005, CHNNO is superior on the metrics of Kendall's rank correlation and Spearman's rank correlation with value 0.705 and 0.752. Although CHNNO is not superior on all three metrics over all datasets we have tested, it still suggests that the Hopfield Neural Network has better conformity on combinatorial optimization problems.

In order to show the agreement between human grading, these human are allocated to arrange sentences extracted for summary or fusion result independently, and then correlation between manually two groups of human-arranged orders are measured on four datasets by three metrics, Spearman's Rank Correlation, Kendall's Rank Correlation and Average Continuity (see Tables 9–12), which is deemed as a strict double-blind accessing process. Specially, they read the source articles before ordering sentences to gain background knowledge on each topic in all datasets. From this process, we achieved manually arranged order. On DUC 2002 dataset, the mean correlation values 0.786 for Spearman's Rank Correlation, 0.564 for Kendall's Rank Correlation and 0.399 for Average Continuity indicate a strong agreement between human grading orders. The similar trends of agreement on another three data sets can be seen in Tables 10–12.

We have indirectly proved the objectivity in subjective grading process above. Then, the distribution of the subjective grading on 5 different ordering algorithms made by a number of judges on four datasets is shown in Fig. 5. Each of these manual evaluations is based on four point scales:

1. Unacceptable
2. Poor
3. Acceptable
4. Perfect

During subjective grading, intrinsic evaluation where evaluation is done by human on accessing the quality of the ordered sentences itself is involved. Four levels of Perfect, Acceptable, Poor, and Unacceptable are used in this process, where the distribution of these subjective grading made by a number of judges to ordering algorithms on four datasets. The readability of the sentence sequence is assessed using five linguistic quality questions which measures qualities of the ordered sentences that do not involve comparison with a standard ordering. The linguistic qualities measured are Grammaticality, Non-redundancy, Referential clarity, Focus and Structure & Coherence, which is used to assess the readability of the summaries task. As the main task in our paper is to properly order the candidate summary sentences, Grammaticality, Focus and Structure & Coherence are badly considered. Fig. 5 shows the distribution of the subjective grading

**Table 4**
The details of comparision methods.

| Method | Description |
|---|---|
| Random Ordering (RO) | Lowest standard, in which sentences are arranged randomly. This method is treated as the lower-baseline and used to indicate the performance that we would obtain. |
| Chronological Ordering (CO) | Sentences are arranged in temporal order of the publication date, where sentences belonging to document published earlier will arranged before sentences belonging to document published later, and for the sentences with similar published date, they are ordered in same order as in original document. In the latter case, if the sentences with same publication date are not belonged to same document, a random order would be taken. |
| Majority Ordering (MO) | Sentences are arranged with the method introduced in Section 2. |
| Continuous Hopfield Neural Network Ordering (CHNNO) | Sentences are arranged with the method proposed in our paper as described in Section 3. |
| Precedence Relation Ordering (PRO) [18] | Highest standard except standard sentences arrangement, where sentences are arranged by using presupposed information, which improved chronological ordering by resolving antecedent sentences of arranged sentences combining the refinement algorithm with topical segmentation. |

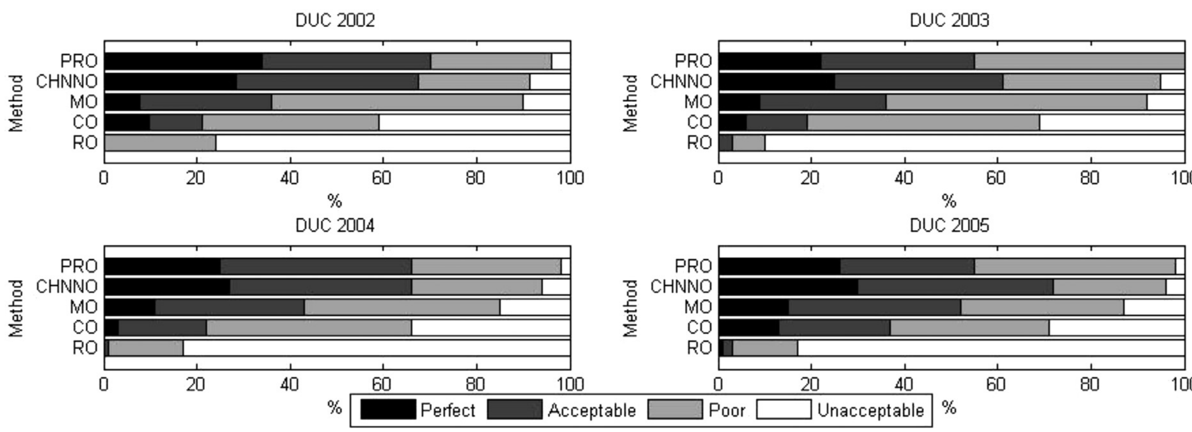



**Fig. 5.** The distribution of the subjective grading by testing Random Ordering method (RO), Chronological Ordering method (CO), Majority Ordering method (MO), Precedence Relation Ordering (PRO) and Continuous Hopfield Neural Network Ordering method (CHNNO)on DUC 2002, DUC 2002, DUC 2004 and DUC 2005 datasets.

**Table 5**
Performance comparison of sentence ordering methods on DUC 2002 dataset.

| Method | Kendall | Spearman | Average Continuity |
|---|---|---|---|
| Random Ordering | 0.115 | 0.169 | 0.052 |
| Chronological Ordering | 0.402 | 0.398 | 0.187 |
| Majority Ordering | 0.424 | 0.407 | 0.282 |
| Continuous Hopfield Neural Network Ordering | 0.728 | 0.801 | 0.722 |
| Precedence Relation Ordering | 0.746 | 0.763 | 0.658 |

**Table 6**
Performance comparison of sentence ordering methods on DUC 2004 dataset.

| Method | Kendall | Spearman | Average continuity |
|---|---|---|---|
| Random Ordering | 0.039 | 0.032 | 0.018 |
| Chronological Ordering | 0.397 | 0.343 | 0.230 |
| Majority Ordering | 0.499 | 0.591 | 0.273 |
| Continuous Hopfield Neural Network Ordering | 0.797 | 0.711 | 0.543 |
| Precedence Relation Ordering | 0.726 | 0.774 | 0.606 |

**Table 7**
Performance comparison of sentence ordering methods on DUC 2003 dataset.

| Method | Kendall | Spearman | Average continuity |
|---|---|---|---|
| Random Ordering | −0.158 | −0.258 | 0.034 |
| Chronological Ordering | 0.487 | 0.432 | 0.392 |
| Majority Ordering | 0.597 | 0.541 | 0.332 |
| Continuous Hopfield Neural Network Ordering | 0.643 | 0.680 | 0.676 |
| Precedence Relation Ordering | 0.622 | 0.701 | 0.562 |

**Table 8**
Performance comparison of sentence ordering methods on DUC 2005 dataset.

| Method | Kendall | Spearman | Average continuity |
|---|---|---|---|
| Random Ordering | 0.034 | 0.124 | 0.018 |
| Chronological Ordering | 0.379 | 0.362 | 0.236 |
| Majority Ordering | 0.398 | 0.359 | 0.265 |
| Continuous Hopfield Neural Network Ordering | 0.705 | 0.752 | 0.493 |
| Precedence Relation Ordering | 0.698 | 0.734 | 0.576 |

by testing Random Ordering method (*RO*), Chronological Ordering method (*CO*), Majority Ordering method (*MO*), Precedence Relation Ordering (*PRO*) and Continuous Hopfield Neural Network Ordering method (*CHNNO*)on DUC 2002, DUC 2002, DUC 2004 and DUC 2005 datasets. From Fig. 5, we can see that most RO are rated as unacceptable (76% on DUC 2002; 90% on DUC 2003; 83% on DUC 2004; 82% 0n DUC 2005), MO and CO has gained similar proportion of Poor. PRO and CHNNO were found as the steadiest methods, because there were scarcely any orders rated as Unacceptable

**Table 9**
Correlation between two groups of human-ordered sentences on DUC 2002 dataset.

| Metric | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|
| Spearman's Rank Correlation ($\rho$) | 0.786 | 0.299 | −0.3 | 1 |
| Kendall's Rank Correlation ($\tau$) | 0.564 | 0.323 | 0.1 | 1 |
| Average continuity | 0.399 | 0.411 | 0.05 | 1 |

**Table 10**
Correlation between two groups of human-ordered sentences on DUC 2003 dataset.

| Metric | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|
| Spearman's Rank Correlation ($\rho$) | 0.664 | 0.280 | −0.5 | 1 |
| Kendall's Rank Correlation ($\tau$) | 0.526 | 0.322 | 0 | 1 |
| Average continuity | 0.508 | 0.497 | 0.002 | 1 |

**Table 11**
Correlation between two groups of human-ordered sentences on DUC 2004 dataset.

| Metric | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|
| Spearman's Rank Correlation ($\rho$) | 0.721 | 0.412 | 0.2 | 1 |
| Kendall's Rank Correlation ($\tau$) | 0.409 | 0.274 | −0.003 | 1 |
| Average continuity | 0.384 | 0.392 | 0 | 1 |

**Table 12**
Correlation between two groups of human-ordered sentences on DUC 2005 dataset.

| Metric | Mean | Std.Dev | Min | Max |
|---|---|---|---|---|
| Spearman's Rank Correlation ($\rho$) | 0.733 | 0.306 | 0 | 1 |
| Kendall's Rank Correlation ($\tau$) | 0.654 | 0.258 | −0.2 | 1 |
| Average continuity | 0.454 | 0.414 | 0 | 1 |

on DUC2003, DUC 2004 and DUC 2005. Over 57% of ordered sentences using PRO and CHNNO are rated as Perfect or Acceptable. Moreover, a larger proportion of Perfect, Acceptable gained by and CHNNO compared with the performance of other methods on all four datasets. On DUC 2002, the ratio of Perfect and Acceptable is 34% and 36% using PRO; 28.5% and 39% using CHNNO, where the performance of PRO is better than CHNNO. However, CHNNO has got higher ratio of Perfect and Acceptable on DUC 2003, DUC 2004 and DUC 2005. The gap between PRO and CHNNO was not gigantic. This fact showed that there is still a vast amount of work need to be done to pushing poor ordering to an acceptable level or a perfect level.

## 6. Conclusion and future work

Sentence ordering problem in our paper is treated as a combinatorial optimization problem and solved with continuous Hopfield neural network (CHNN). With Continuous Hopfield Neural Network Ordering (*CHNNO*), not only did we examine most frequent orders in original document (i.e., majority based method) and orders sentence by publication date (i.e., temporal information based method), but also we considered the topical relevance between local themes during overall ordering process. The ordering algorithm traversed all the local themes and searched a shortest path as the final sentence ordering. The CHNNO proposed in this paper is superior to RO, CO and MO on all metrics, which proved the effectiveness of our idea. Although CHNNO is not superior to PRO on all three metrics over all datasets we tested, it still, at some point, revealed that the Hopfield Neural Network Ordering (CHNNO) has better conformity on sentence ordering problem. During subjective grading, the distribution of the subjective grading showed that there is still a vast amount of work need to be done to pushing poor ordering to an acceptable level or a perfect level.

Our study will further focus on improvement of sentence ordering method. On one hand, the CHNNO proposed in this paper strongly relies on model parameters and initial conditions and the energy function is not unique, which all make improvement by no means exhaustive. On the other hand, numerous other methodology should be involved on sentence ordering problem. Moreover, we would also like to explore establishing a standard subjective evaluation system as another task.

## Acknowledgment

Dr. Lin YUE has been awarded a scholarship under the State Scholarship Fund to pursue her study at the University of Queensland as a joint PhD. Student, and this work has been awarded by China Scholarship Council (CSC). Besides, we really appreciate Document Understanding Conferences (DUC) providing test collection for evaluation, and the anonymous reviewers for their constructive comments which will make substantial improvements to this paper.

## References

[1] D.R. Radev, H. Jing, M. Styś, D. Tam, Centroid-based summarization of multiple documents, Inf. Process. Manag. 40 (2004) 919–938.
[2] S. Afantenos, V. Karkaletsis, P. Stamatopoulos, Summarization from medical documents: a survey, Artif. Intell. Med. 33 (2005) 157–177.
[3] R.M. Alguliev, R.M. Aliguliyev, M.S. Hajirahimova, C.A. Mehdiyev, MCMR: maximum coverage and minimum redundant text summarization model, Exp. Syst. Appl. 38 (2011) 14514–14522.
[4] Y. Ouyang, W. Li, S. Li, Q. Lu, Applying regression models to query-focused multi-document summarization, Inf. Process. Manag. 47 (2011) 227–237.
[5] C. Monz, Document fusion for comprehensive event description, in: Proceedings of the Workshop on Human Language Technology and Knowledge Management-Volume 2001, Association for Computational Linguistics, 2001, p. 8.
[6] J. Azzopardi, C. Staff, Fusion of news reports using surface-based methods, in: 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2012, IEEE, 2012, pp. 809–814.
[7] E. Newman, W. Doran, N. Stokes, J. Carthy, J. Dunnion, Comparing redundancy removal techniques for multi-document summarization, in: Proceedings of STAIRS, 2004, pp. 223–228.
[8] R. Witte, S. Bergler, Fuzzy, clustering for topic analysis and summarization of document collections, Advances in Artificial Intelligence, Springer, 2007, pp. 476–488.
[9] L. Dai, J.l. Tang, Y.q. Xia, Subtopic-based multi-document summarization, in: 2009 International Conference on Machine Learning and Cybernetics, IEEE, 2009, pp. 3505–3510.
[10] Y.D. Xu, X.D. Zhang, G.R. Quan, Y.D. Wang, MRS for multi-document summarization by sentence extraction, Telecommun. Syst. 53 (2013) 91–98.
[11] A. Bronselaer, G. De Tré, Aspects of object merging, fuzzy information processing society (NAFIPS), in: 2010 Annual Meeting of the North American, IEEE, 2010, pp. 1–6.
[12] A. Bronselaer, D. Van Britsom, G. De Tré, A framework for multiset merging, Fuzzy Sets Syst. 191 (2012) 1–20.
[13] E. Reiter, R. Dale, Z. Feng, Building Natural Language Generation Systems, MIT Press, 2000.
[14] N. Karamanis, H.M. Manurung, Stochastic text structuring using the principle of continuity, in: Proceedings of INLG, 2002, pp. 81–88.
[15] P.A. Duboue, K.R. McKeown, Content planner construction via evolutionary algorithms and a corpus-based fitness function, in: Proceedings of INLG, 2002, pp. 89–96.
[16] K. McKeown, J. Klavans, V. Hatzivassiloglou, R. Barzilay, E. Eskin, Towards multidocument summarization by reformulation: progress and prospects, in: AAAI/IAAI, 1999, pp. 453–460.
[17] R. Barzilay, N. Elhadad, Inferring strategies for sentence ordering in multidocument news summarization, J. Artif. Intell. Res. 17 (2002) 35–55.
[18] N. Okazaki, Y. Matsuo, M. Ishizuka, Improving chronological sentence ordering by precedence relation, in: Proceedings of the 20th International Conference on Computational Linguistics, Association for Computational Linguistics, 2004, p. 750.
[19] M. Lapata, Probabilistic text structuring: experiments with sentence ordering, in: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, Volume 1, Association for Computational Linguistics, 2003, pp. 545–552.
[20] P.D. Ji, S. Pulman, Sentence ordering with manifold-based classification in multi-document summarization, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2006, pp. 526–533.

[21] R. Barzilay, L. Lee, Catching the drift: probabilistic content models, with applications to generation and summarization, arXiv preprint cs/0405039, (2004).
[22] D. Bollegala, N. Okazaki, M. Ishizuka, A bottom-up approach to sentence ordering for multi-document summarization, Inf. Process. Manag. 46 (2010) 89–109.
[23] D. Bollegala, N. Okazaki, M. Ishizuka, A preference learning approach to sentence ordering for multi-document summarization, Inf. Sci. 217 (2012) 78–95.
[24] M.T. Hagan, H.B. Demuth, M.H. Beale, O. De Jesús, Neural Network Design, PWS publishing company Boston, 1996.
[25] W. Liu, X. Luo, J. Xuan, Z. Xu, D. Jiang, Cognitive memory-inspired sentence ordering model, Knowl. Based Syst. 104 (2016) 1–13.
[26] T. Yanase, T. Miyoshi, K. Yanai, M. Sato, M. Iwayama, Y. Niwa, P. Reisert, K Inui, Learning sentence ordering for opinion generation of debate, in: NAACL HLT 2015, 2015, p. 94.
[27] W. Liu, X. Luo, D. Jiang, A spreading activation based sentence ordering model collaborated with cognitive logical structure, in: 2015 11th International Conference on Semantics, Knowledge and Grids (SKG), IEEE, 2015, pp. 119–125.
[28] R. Barzilay, N. Elhadad, K.R. McKeown, Sentence ordering in multidocument summarization, in: Proceedings of the First International Conference on Human Language Technology Research, Association for Computational Linguistics, 2001, pp. 1–7.
[29] H.T. Dang, Overview of DUC 2005, in: Proceedings of the Document Understanding Conference, 2005, pp. 1–12.
[30] http://duc.nist.gov/
[31] http://www.nist.gov/tac/
[32] http://lr-www.pi.titech.ac.jp/tsc/tsc3-en.html
[33] http://research.nii.ac.jp/ntcir/index-en.html

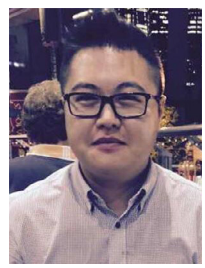


**Lin Yue** received the B.Sc. degree and M.S. degree in data mining from the College of Computer Science and Technology, Northeast Normal University, Changchun, China. She is pursuing the Ph.D. degree in Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun, China. She has been awarded a scholarship under the State Scholarship Fund to pursue her study in the University of Queensland, Australia, as a joint PhD. Student from 2014 to 2016. She has published multiple journal papers and conference papers. Her research interests include data mining, natural language processing and machine learning.

**Zhenkun Shi** received the B.Sc. degree in the College of Computer Science and Technology, Agricultural University of Hebei, Hebei, China. He is pursuing the Ph.D. degree in Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun, China. His research interests include Social Computing, Natural-Linguistic Process, and Data Mining.

**Jiayu Han** received the B.Sc. degree and M.S. degree in the College of Computer Science and Technology, Jilin University, Changchun, China. She is pursuing the Ph.D. degree in Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, Jilin University, Changchun, China. Her research interests include data mining, data fusion, recommender systems and machine learning.

**Sen Wang** received his Ph.D. in The University of Queensland (UQ), Australia in 2014. He got his M.E degree in Computer Science and B.Sc. degree in Computer Science from Jilin University, China and Liaoning Shihua University, China, respectively. He was an ARC post-doctoral research fellow in the University of Queensland. He is a lecturer in Griffith University. His research interests include signal and image processing, pattern recognition and machine learning algorithms, biomedical applications and big data analytics.



**Weitong Chen** received the B.Sc. degree in Information System form Griffith University, M.S. degree in Computer Scientist from the University of Queensland, Australia. He is pursuing his Ph.D. degree in The University of Queensland with scholarship fund in Australia. He currently is a research officer in The University of Queensland in Brisbane, Queensland, Australia. His major areas of research interests and expertise include biomedical applications and big data analytic data Mining, and social computing.

**Wanli Zuo** is a Professor in Jilin University. He received the B.Sc. degree, M.S. degree and Ph.D. degree in the College of Computer Science and Technology, Jilin University, Changchun, China. He has published more than 160 journal papers and conference papers. His research interests include data mining, information retrieval, natural language processing and machine learning, etc.