NUS at DUC 2006: Document Concept Lattice for Summarization

Shiren Ye and **Tat-Seng Chua**
School of Computing, National University of Singapore
Singapore, 117543
{yesr|chuats}@comp.nus.edu.sg

Abstract

Concepts composed of open-class terms after semantic equivalence discovery can be considered as simplified basic elements. We utilize frequent concept sets to construct a Document Concept Lattice, which contains hierarchical summary information of a document cluster. Based on this lattice, we further extract a set of sentences with maximal representative power and minimal redundancy for summarization. The implementation of our summarization approach via concept lattice obtains competitive performance in DUC 2006.

1 Introduction

Text summarization is the process of distilling the most important information from sources to produce an abridged version for a particular users and tasks (Mani and Maybury, 1999). It has been widely used in news and meeting summarizations, search engine hits, and so on. Document understanding (Dang, 2005) is the core of text summarization, as well as information extraction and question answer.

The quality of machine-crafted summary is generally unsatisfactory up to now. Almost all existing techniques and systems utilize statistics to discover the salient sentences for an extractive summary. Their principles are far from *true* document understanding, and cannot compress and condense the contents existing in the documents comprehensively and accurately. Meanwhile, the given queries cannot express the user's intention without ambiguity, and there are few agreements and overlappings between diverse summaries, even though crafted by humans (Nomoto and Matsumoto, 2001).

In this paper, we explore the (partial) understanding of text through conceptual and semantic analysis, and then apply a model to document summarization. Our work is based on the concept link approach proposed by Ye et al. (2005). Specifically, we investigate the use of a Document Concept Lattice (DCL) to capture the inter-linking and relationship between concepts in a document cluster. DCL is a hierarchical summary structure based on the co-occurrence of concepts among sentences. The derived nodes high up in the DCL hierarchy contain co-occurred concepts, which tend to be stable and reliable to cover the diverse sentences having close meaning. Our experiment results on DUC 2006 corpus indicate that DCL can be used to select a set of salient sentences with maximal representative power and minimal redundancy for summarization.

The rest of this paper is organized as follows: Section 2 introduces the background of concepts and related work. After discussing the principles of Basic Element, Section 3 describes how to present the contents in text. Sections 4 and 5 address DCL and its construction, as well as the algorithm for constructing the summary based on DCL. We report the experiment results on DUC 2006 corpus in Section 6. Finally, we conclude the paper in Section 7.

2 Background

To simplify the problem of summarization, like most existing approaches, we focus on the selection of a subset of sentences that can best represent the raw text. In general, sorting (e.g., Latent Semantic Index (Ando et al., 2000)) and clustering (e.g., Centroid (Radev et al., 2004)) the sentences according to one or more sets of features are two main branches of techniques to achieve it.

As stated in Katz's G model (Katz, 1996), the selected sentences in a summary must possess two important properties: *redundancy* and *diversity*. The former relates to how repetitive the concepts (or content words) are; while the latter relates to how many distinct concepts are appearing. In this paper, we propose a lattice model based on concepts that indexes all sentences in the cluster to facilitate the selection of sentences with these two properties.

For redundancy, the core issue is the detection of the repetition among a set of sentences. We can say that a sentence is a good representative, if and only if, it stands for a set of coherent sentences, as well as for itself. In DCL, the associations between such sentences can be reflected by the repeated entities and their actions that we called Concepts. We calculate and organize all possible combinations of high-frequency concepts, which will form a hierarchy that summarize these concepts from diverse perspectives. The sentences containing the corresponding concept set will be indexed by the (multiple) nodes in DCL. The sentences with the highest representative power indexed by the significant nodes in DCL will render the most redundancy. The details on the construction of DCL are given in Section 4.

For diversity, the most typical approach is Maximal Marginal Relevance (MMR) (Carbonell and Goldstein, 1998). MMR ranks the sentences according to a combined criterion of query relevance and novelty of information, where the novelty refers to the degree of dissimilarity between the document being considered and those previously selected in the ranked list. In DCL, the diversity is reflected by the fact that all selected sentences do not share significant nodes. Since one sentence may be covered by more than one of such nodes, the total number of repetitive concept sets is usually larger than the number of selected sentences, which facilitates diversity according to our discussion in Section 5.2.

3 The Concept

As proposed in (Hovy et al., 2005), Basic Element (BE), a minimal semantic unit which is broken down from reference sentences, is the ideal representation (unit) for text summarization and evaluation. Furthermore, they considered Summary Content Units in Pyramid (Passonneau et al., 2005) and various n-grams in ROUGE (Lin and Och, 2004) as different versions of BEs with clause-length and small-size, respectively. Hovy et al. also argued that the smaller basic elements tend to allow the automation in the procedure of unit identification, and facilitate the matching of diverse equivalent expression.

However, it is difficult to identify and match BEs in text, no matter what granularity are used, although some explorative work has been conducted. Following the idea of concept link described in Ye et al. (2005), we believe that concepts can be considered as BEs with a proper granularity, where minimal human efforts will be involved.

In order to better understand this, let us review the sentences from DUC 2005 corpus (Cluster *d324e*) as shown in Figure 1. Here sentences (1) to (4) mention the Falkland islands war between Britain and Argentina. Although their structures vary largely, they contain a set of stable and overlapping concepts representing key entities and their actions in the story. This is consistent with our observations that even though the authors may perceive an idea from different perspectives, and narrate it using various terms (such as synonyms, alias or words with different part-of-speech) and in diverse sentence constructs, the concepts used still belong to a finite set and are overlapping in most cases.

Actually, the primary and ideal BE is defined as (i) the head of a major syntactic constituent, or (ii) a relation between a head-BE and a single dependent (Hovy et al., 2005). Consider a pair of BEs, *Mr Douglas Hurd* and *Mr Hurd*, in sentences 6 and 8, which refer to the identical person. If they are considered as five independent concepts in these two sentences as we lack proper deep parser tool to identify and co-refer the true concepts, two concept links between the occurrences of *Mr* and *Hurd*, rather than one between the true entities, would be found. The problem of this processing strategy is that it tends to bring in some redundant concept links. However, it is equivalent to assigning higher weights to longer repeated NPs and VPs, and it is also consistent with the basic criteria proposed by Hovy et al. (2005): small size (to allow proper scoring of atomic bits of content), regularity/simplicity of definition and automatic production.

We can find that these concepts are usually expressed by open-class terms, such as nouns, verbs, adjectives and adverbs rather than terms in closed-class (e.g., preposition and conjunction, etc.). For example, in the sentences shown in Figure 1, the underlined terms are concepts. Compared with all terms in sentences, these concepts can better track the coherence between sentences. Stemming

No	*Sentence*
1	Argentine-British relations since the Falkland Islands War in 1982 have gradually improved.
2	Thirteen years after the war between Britain and Argentina over the Falkland Islands, ...
3	Argentina was still obsessed with the Falkland Islands even in 1994, 12 years after its defeat in the 74-day war with Britain.
4	Argentina and British fought over Falkland islands in 1982.
5	Commercial relations have continued to improve between UK and Argentina.
6	Mr Douglas Hurd, Britain foreign secretary, is to visit Argentina early next year.
7	Britain lifted military protection zones around the Falklands in 1990, 8 years after the Argentina-British war over the area.
8	Mr Hurd's visit to Argentina is the first by a cabinet minister since the Falklands conflict indicating improved diplomatic relations between UK and Argentina.

Figure 1: Some sentences from cluster at DUC 2005

method can definitely remove most morphological variations (such as *Argentine* and *Argentina*), but it cannot handle the synonym problem (such as *war* and *battle*). Open-domain unsupervised concept discovery is still a difficult problem (Kazi and Ravin, 2000). Therefore, we can say that this method is feasible, although it is far from the ideal solution of BE.

Based on the discussion above, concepts are considered as the terms that remain after removing closed-class words in sentences. We check WordNet for entries that match sequences of words as multi-word concepts, such as *Falkland islands* in the above example. For unknown words that do not appear in WordNet, we use them directly as separate concepts.

The details of how to compute the similarity between concepts for identifying the identical and close BEs can be found in (Ye et al., 2005).

No	*Concept*	*No*	*Concept*
A	Argentine Argentina	I	island
B	British Britain UK	J	1982
C	war fought conflict military	K	year
D	diplomatic foreign	L	Mr
E	secretary minister	M	Hurd
F	area zone	N	visit
G	improve	O	relation
H	Falkland islands	P	Falklands

Table 1: Equivalent concepts (A˜ F) and repeated concepts (G˜ P) appearing in Sentences in Figure 1. Here NP *Falkland islands* is considered as a concept as we can find it in WordNet; Hurd and Falklands are also considered as two concepts since they are unknown words (missing in WordNet). In total, 16 concepts instead of 61 distinct words found in these sentences can be used to represent the contents.

4 Document Concept Lattice

Suppose that three sentences, say A, B and C, share partial (but not all) concepts. Namely, if $cpts(X)$ denotes the concepts in Sentence X, then $cpts(A) \neq cpts(B) \neq cpts(C) \neq \phi$, and $cpts(A) \cap cpts(B) \neq cpts(B) \cap cpts(C) \neq cpts(A) \cap cpts(C)$. It is difficult to split them into two clusters or sort them into a list using an ideal paradigm that can achieve reasonable interpretation in all possible concept distributions. In contrast, we investigate all possible combinations, such as $cpts(A), cpts(B), cpts(C), cpts(A) \cap cpts(B), cpts(B) \cap cpts(C), cpts(A) \cap cpts(C)$ and $cpts(A) \cap cpts(B) \cap cpts(C)$, and further organize them into a lattice. A, B and C will be indexed many times in this lattice, which can be considered as the compound of the results of diverse clustering. At the same time, when we travel all nodes in this lattice by means of various strategies or evaluation functions (i.e., representative power in Eqn. 2), we obtain different returned listings of sorting sentences.

Like Data Cube for structured data which is used to summarize measures according to their dimensions, such as the total sale in the specified regions, seasons and categories (Stumme et al., 1998), we can build a concept summary structure hidden in a document cluster as well. By denoting the set of concepts using the alphabets as listed in Table 1, the construction of a concept summary structure as shown in Figure 2 can be carried out as follows. (1) The repeated concepts extracted from the parsing result of each sentence are regarded as elements in a base node (such as node 1-8 for the 8 sentences

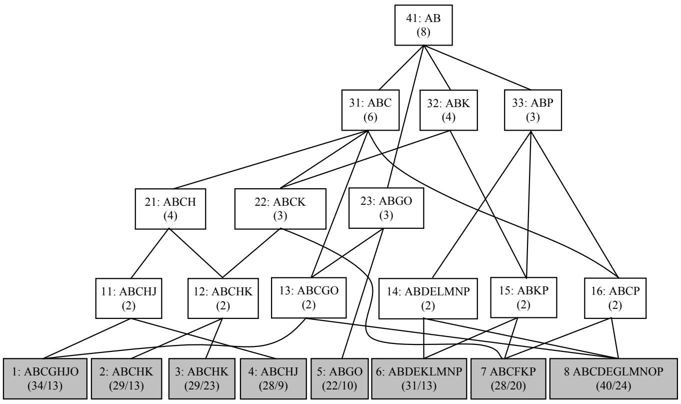

Figure 2: Document concept lattice derived from terms existing in sentences in Figure 1. Here, the terms with frequency of 1 and stop words are ignored. The nodes with gray background are base nodes, and the others are derived nodes

given in Figure 1). (2) The maximal common concepts are used to form the derived nodes. For example, the maximal common concepts for nodes 1 and 4 are *ABCHJ* in node 11 rather than *AB* in node 41 or *ABC* in node 31. (3) More derived nodes are recursively generated from the existing base nodes and derived nodes until no new derived node appears. (4) A pair of nodes with inclusion relation and least element variation are linked together to form a partially ordered relations. For instance, nodes 1 and 31 are linked via nodes 11 and 21 rather than by a direct linkage. (5) If there are more than one upper nodes existing, a node ϕ with no element are used to cover all upper nodes. (6) Finally, a new node with all existing concepts are introduced to derive all base nodes.

The resulting structure is a lattice called **Document Concept Lattice (DCL)** when we define the partial ordering relation $\preceq$ between node i and j iff $element(i) \subseteq element(j)$. A formal description and construction algorithm of a lattice can be found in Wang et al. (1999).

As a document concept model, DCL has the following characteristics:

1. All high-frequency concepts existing in the identical node suggests that they have strong cohesion. For example, *British* and *Argentine*, the core agents in the cluster, are the highest-frequency concepts that usually co-occur. Particularly, the node with a number of high-frequency concepts will reflect an important event or scenario in the cluster. A striking example of this is node 21, whose concepts *ABCH* describe a concrete event of the war between Britain and Argentine.
2. The number in parentheses in the derived nodes gives the frequency of occurrences of their concept set. It indicates the significance of the corresponding set of concepts if we follow the assumption that repetitive concepts are more salient.
3. If a set of high-frequency concepts are found adjacent to each other (such as *the Falkland Islands*), they will most likely be noun phrases or verb phrases tied to particular entities or actions. In cases where the concepts appear in isolation, such as *Falklands* in sentence 7, they might be the alias or informal expressions of the canonical concept names.
4. The node having a set of high-frequency elements denotes a set of concepts with intensive relations, where many sentences support such relations. The sum of value of concepts could be considered as an indicator of its significance.
5. The derived nodes in higher layers in DCL contain fewer elements, but the frequency of these elements are higher. They cover more base nodes, and thus imply more general description about the topic. When we zoom down through DCL, we can find more details.
6. We are able to add equivalent terms (i.e., synonymy) as annotation for concepts. For each derived node, a representative sentence (e.g., with maximal RP in Eqn. 2), can be considered as the typical instantiation result of the existing concept set. Consequently, we can say that the contents of DCL are abundant, which will facilitate a visualization approach to interpreting document skeleton.

However, the computation cost of constructing a complete DCL is very expensive since almost every possible concept combinations need to be examined. Fortunately, we are only interested in nodes with high-frequency concepts, as other nodes with low-frequency concepts tend to be about trivial information that can be ignored. For instance, we can simplify nodes 11, 12 and 21 as one node, where the statistical information relating to the low-frequency concepts K and J is ignored. Here we can use a variation of association rules (Agrawal and Srikant, 1999) to mine the nodes with high-frequency concepts. Meanwhile, the concept sets with smaller number of concepts whose frequency is equal to a larger concept set will be ignored. For instance, concept sets for *A* and *B* will be excluded since they appear 8 times as the larger concept set *AB*. Namely, just the largest frequent sets are taken into account. Finally, it is convenient to build a concise version of DCL based on the frequent concept sets according their relations of set inclusion. The computation cost of generating the sets of frequent concepts is linear to the number of concepts, and the construction of the concise DCL is $O(n^2)$, where n is the number of frequent sets, making it feasible to construct a DCL for a document cluster.

5 Use of Document Concept Lattice for Summarization

5.1 Sentence Representative Power

Intuitively, we can evaluate the significance of every sentence tied to a base node according to the significance of the derived nodes that cover it. We assume that the derived nodes will represent the reoccurrence concepts in other sentences. Therefore, the sum of its concept frequencies can reflect its significance. In Figure 2, the number in parentheses in the derived nodes denotes the frequency of existing concepts; while the two numbers in parentheses in the base nodes are: (a) the sum of concept frequencies; and (b) the number of words in the sentence respectively. We further employ the inverse document frequency (IDF[1]) to weight the frequency to obtain the **significance** of sentence s as follows:

$$Sig(s) = \mathbf{\Sigma}\{Freq(c)\log(N/df_c)\}, \qquad (1)$$

where N is the number of documents in corpus, and df_c is the number of documents containing concept c. Similarly, the significance of a derived node is also computed using the same formula.

As previously discussed, for summarization task based on sentence-selection, we attempt to obtain a set of sentences which could: (i) contain more concepts within a short word length; and (ii) have minimal redundancy among the selected sentences. The first condition can be measured by their ratio of Sig and word number that we called the **representative power**:

$$RP(s) = Sig(s)/word\_num(s). \qquad (2)$$

In the example given in Figure 2, RP of node 4 will exceed other base nodes, for its RP is $35/9$ if we ignore IDF. The second condition can be linked to the constraint that the corresponding base nodes should be covered by as many high-frequency derived nodes as possible.

5.2 The Algorithm

The sentence-selection optimization is a kind of NP-hard knapsack problem if we explore all possible combinations in the search space. Here we utilize DCL to limit the size of the search space. We investigate only the sub-space covered by the derived nodes with high significance: base nodes with largest RP which is covered by a set of non-overlapping[2] derived nodes Ω are first selected to satisfy the first condition. At the same time, we select those base nodes that do not share any derived node in Ω to satisfy the second condition. The concrete operations are: (i) select a set of non-overlapping derived nodes with high significance as Ω; (ii) select a base node (marked by P) with maximal RP which are covered by the node with maximal significance in Ω; (iii) remove all derived nodes that cover P from Ω; (iv) repeat steps (ii) and (iii) until the desired length of summary is obtained.

A pivotal issue in summarization is how to choose the proper non-overlapping derived nodes that produces the summary with the required length. If we chose nodes high up in DCL for Ω (namely, nodes that cover more base nodes), we may not have enough sentences for the summary because all relevant base nodes cannot share the derived nodes in Ω. On the other hand, if we chose nodes low in DCL for Ω (i.e., nodes that cover fewer base nodes), the algorithm will tend to generate sentences from the sets of locally maximal concepts and exclude sentences

[1]We use a large-score balanced corpus available at http://elib.cs.berkeley.edu/docfreq to get the term document frequency. to retrieve the concept distribution.

[2]derived nodes A and B are non-overlapping iff neither $A \preceq B$ nor $B \preceq A$.

with higher RP. This is because in this case, there will be many eligible base nodes to choose and some base nodes with high RP may not have the opportunity to be scanned and selected.

Algorithm 1 DCLSummarizer

Input: DCL, n_0, n_{step}
Output: summary sum.
1: sort all derived nodes by their significance
2: **for** $i \leftarrow n_0, |DCL|$ **step** n_{step} **do**
3: $\quad \Omega \leftarrow \{$ top i nodes in DCL $\}$
4: $\quad \Omega \leftarrow \{M | M \in \Omega, M$ has no successor in $\Omega\}$
5: $\quad sum \leftarrow \phi$
6: $\quad$ **repeat**
7: $\quad\quad$ add sentence with maximal RP whose base node N_{cur} is covered by any node in Ω into sum
8: $\quad\quad$ remove all nodes in Ω that cover N_{cur}
9: $\quad$ **until** sum's length is OK or $\Omega = \phi$
10: $\quad$ **if** sum's length is OK **then**
11: $\quad\quad$ return sum
12: $\quad$ **end if**
13: **end for**

We utilize a generate-and-test approach to obtain a proper Ω: First select the top n_0 derived nodes to generate an Ω with fewer derived nodes, and then form a summary. If its length cannot satisfy the requirement, we will select the top $n_0 + n_{step}$ derived nodes and then repeat this procedure until a desired summary is output. The algorithm is given in the Algorithm 1.

In algorithm *DCLSummarizer*, Steps 3-4 will generate the set of non-overlapping derived nodes Ω. These nodes might have various depth from the top node. Steps 6-9 repeatedly add the salient sentence with maximal RP in sum. We may obtain a sum whose length cannot reach the users requirement when the algorithm does not pick up enough salient sentences with minimal redundancy. This problem can be corrected by loosening the criterion of Ω selection. Namely, as more significant nodes join Ω in Step 3 as i increases, there are more candidate sentences covered by these nodes for summarization.

6 System Overview and Experiments

Unlike many other existing approaches, we have not used heuristics such as sentence position, length, centroid, title overlap and even cue phrases, although it has been reported that combining these heuristics did have large contribution to system performance (Erkan and Radev, 2004).

The input given to our summarization system is composed of a cluster of relevant documents and a topic. At the preprocessing phase, our system ignores the document boundaries in the document cluster. It takes all the documents as a single document which it delimits into sentences for further analysis. The topic is treated similarly: only its sentences boundaries, if any, are detected. No other features of the topic are collected. The system, $DCLSummarizer$, summarizes the cluster according to the following workflow:

1. After a *tokenizer* delimits numbers, words, and punctuations under the given format, a *sentence delimiter* detects and annotates sentence boundaries.
2. A shallow parser named *NLProcessor* (available at http://www.infogistics.com/ textanalysis.html by *Infogistic*) outputs the part-of-speech of all words. We thus obtain all open-class terms.
3. We use the method described in Section 3 to detect semantic equivalence for all existing concepts. Here the threshold for semantic equivalence is set to 0.35.
4. We build a DCL as described in Section 4, where the minimal frequency of concepts is set to 3.
5. We utilize the algorithm *DCLSummarizer* (Section 5) to generate the desired summary.

The corpus of DUC 2006 consists of 50 document clusters from *Financial Times of London* and *Los Angeles Times*. The average document number and size per cluster is about 31.9 and $144k$ respectively. The aim of the test is to benchmark our system against other DUC participants and human generated summaries.

As shown in Figure 3, our system achieves relatively good results with respect to ROUGE measures (Lin and Och, 2004). In particular, the scores rank us among the top three systems with respect to the suggested ROUGE-2 and ROUGE-SU4 measures. The average recall under ROUGE-2 and

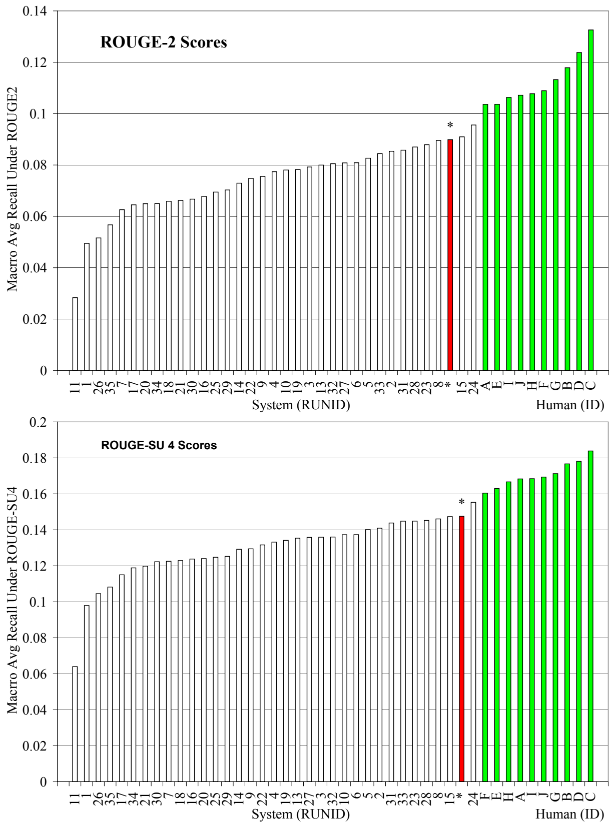

Figure 3: ROUGE Scores of system and human summaries. The column marked by "*" are the scores of our system. The columns with light background are scores of peer systems in DUC 2006, and the columns with dark background are scores of human summaries.

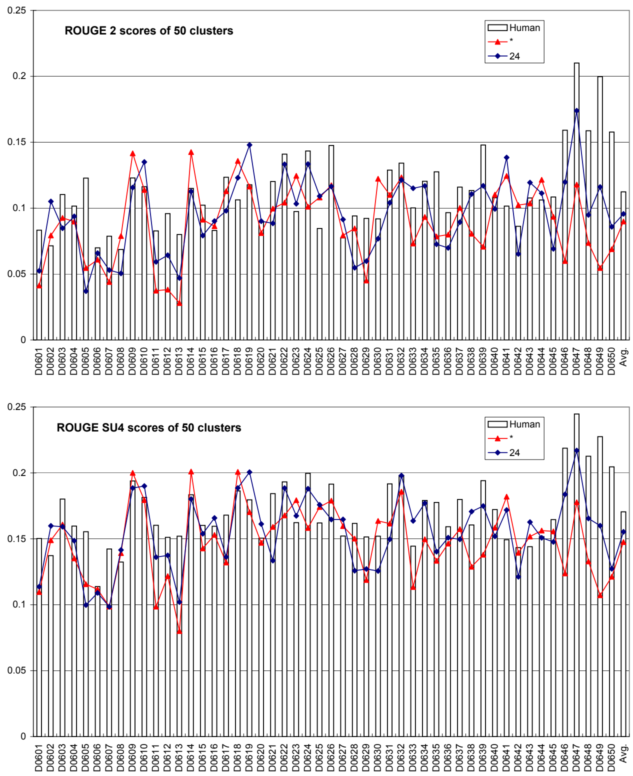

Figure 4: ROUGE Scores in 50 clusters. The columns labeled by "Human" are the average scores of 10 human summaries (Since every person craft only about 20 clusters, all persons are considered as a virtual one with average scores), and the lines labeled "24" and "*" are the ROUGE scores of best system and our system in DUC 2006, respectively.

ROUGE-SU4 are 8.99% and 14.76%, respectively. Furthermore, we also observed that for our system, the difference between recall and precision for each document cluster is small.

Figure 4 illustrates the detailed ROUGE scores of the (average) human summary, the best system and our system in 50 clusters. Overall, we find that: (1) The variations of ROUGE scores of both humans, the best system and our system in different clusters are large, the the estimated standard deviations of 0.030, 0.029 and 0.029 in ROUGE-2 respectively. (2) The difference between humans and machines per cluster are also large. For example, the average ROUGE-2 difference between the human and our system per cluster is 0.032 although the difference of average ROUGE-2 scores of all clusters is 0.0225. (3) In some clusters (about 10 clusters), machine-crafted summaries outperform human-crafted ones in terms of ROUGE scores. However, it seems that we cannot say that the quality of the former is better than the latter in these clusters.

7 Conclusion

By extending the summarization approach based on concept link, we proposed a Document Concept Model and corresponding summarization algorithm. Here, the elements in DCL nodes consist of meaningful concepts, which are terms in open class after semantic equivalence discovery. In the procedure of constructing DCL, all sentences in documents are represented by a basket of concepts in base nodes, and frequent concept sets mined from these base nodes will form the derived nodes. All nodes in DCL are partially ordered under inclusion relation, where upper nodes have larger frequency and cover more base nodes. Therefore, a set of proper sentences with maximal representative power and minimal redundancy can be selected for summarization when they are covered by a set of non-overlapping derived nodes with maximal significance. The experiment results also argue that DCL is a competitive model as compared with technologies based on sentence-clustering and sentence-sorting.

Acknowledgements

We would like to thank Dr Min-Yen Kan for the valuable and comprehensive comments. We would also like to thank Long Qiu and Hang Cui for their helps with the evaluation.

References

Rakesh Agrawal and Ramakrishnan Srikant. 1999. Fast algorithms for mining association rules. In *VLDB*, pages 487–499.

Rie Ando, Branimir Boguraev, Roy Byrd, and Mary Neff. 2000. Multi-Document Summarization by Visualizing Topical Content. Seattle, WA, April.

Jaime Carbonell and Jade Goldstein. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336.

Hoa Trang Dang. 2005. Overview of DUC 2005. In *Document Understanding Conference*.

GüneşErkan and Dragomir R. Radev. 2004. The university of michigan at DUC 2004. In *Document Understanding Conference*.

Eduard Hovy, Chin-Yew Lin, and Liang Zhou. 2005. Evaluating DUC 2005 using basic elements. In *Document Understanding Conference*.

Slava Katz. 1996. Distribution of content words and phrases in text and language modelling. *Natural Language Engineering*, 2(1):15–60.

Zunaid Kazi and Yael Ravin. 2000. Who's who? identifying concepts and entities across multiple documents. In *HICSS*.

Chin-Yew Lin and Franz Josef Och. 2004. Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *ACL*. http://www.isi.edu/˜cyl/ROUGE/.

Inderjeet Mani and Mark T. Maybury. 1999. *Advances in Automatic Text Summarization*. The MIT Press.

Tadashi Nomoto and Yuji Matsumoto. 2001. A new approach to unsupervised text summarization. In *SIGIR*, pages 26–34, New York, NY, USA. ACM Press.

Rebecca J. Passonneau, Ani Nenkova, Kathleen McKeown, and Sergey Sigelman. 2005. Applying the pyramid method in DUC 2005. In *Document Understanding Conference*.

D. Radev, H. Jing, and M. Budzikowska. 2004. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40:919–938.

Gerd Stumme, Rudolf Wille, and Uta Wille. 1998. Conceptual knowledge discovery in databases using formal concept analysis methods. *Lecture Notes in Computer Science*, pages 450–458.

Hui Wang, Werner Dubitzky, Ivo Düntsch, and David Bell. 1999. Lattice machine approach to automated casebase design: Marrying lazy and eager learning. In *IJCAI*, pages 254–259.

Shiren Ye, Long Qiu, Tat-Seng Chua, and Min-Yen Kan. 2005. NUS at DUC 2005: Understanding documents via concept links. In *Document Understanding Conference*.