Abstractive News Summarization based on Event Semantic Link Network

Wei Li[1,+]**, Lei He**[2,*]**, Hai Zhuge**[3,+,*]

[+]Key Lab of Intelligent Information Processing, Institute of Computing Technology,
University of Chinese Academy of Sciences, Beijing, China
[*]Aston University, Birmingham, UK
[1]`weili.ict.kg@gmail.com`, [2]`hel2@aston.ac.uk`, [3]`zhuge@ict.ac.cn`

Abstract

This paper studies the abstractive multi-document summarization for event-oriented news texts through event information extraction and abstract representation. Fine-grained event mentions and semantic relations between them are extracted to build a unified and connected event semantic link network, an abstract representation of source texts. A network reduction algorithm is proposed to summarize the most salient and coherent event information. New sentences with good linguistic quality are automatically generated and selected through sentences over-generation and greedy-selection processes. Experimental results on DUC 2006 and DUC 2007 datasets show that our system significantly outperforms the state-of-the-art extractive and abstractive baselines under both pyramid and ROUGE evaluation metrics.

1 Introduction

Automatic summarization on news documents enables readers more easily to get general information of interesting news. Most of existing summarization methods have neglected the important event-oriented characteristics of news texts although some popular tasks such as DUC (Document Understanding Conference) and TAC (Text Analysis Conference) target at summarizing news documents. The examples below show that the core information of news texts is the atomic event mentions as shown in bolded words and their related concepts as shown in italic phrases.

- Lawyer *Morris Dees*, who is **representing** *Victoria Keenan* after she was **attacked** by *two guards* in *July 1998*, **introduced** *depositions* to **contradict** *the men's testimony*.
- *Morris S. Dees Jr.*, who was the co-founder of the Southern Poverty Law Center, **defended** for *Keenan* after she was **assaulted** by *two security guards* near *the headquarters of the Aryan Nations*.

An event usually tells us "**who** did **what** to **whom when** and **where** ..." The most important components of an event include its **actor** (who, the agent of the event), **action** (what, the core meaning of the event) and **receiver** (whom, the target of the event action). Other arguments indicate other attributes of the event, such as **time** (when) and **location** (where). The event arguments are concepts related with the event action. For event-based news summarization, extracting the most salient events and related concepts are the core tasks.

One of the most similar related work (Glavaš and Šnajder, 2014) investigated constructing event graph for multi-document summarization. The nodes in event graph denote event mentions while edges denote temporal relations between event mentions. It ranks event mentions based on the temporal relations and then generates summary by extracting sentences that contain salient event mentions. However, the problems of information overlapping and lacking of coherence cannot be overcome by extractive methods. This paper explores the issue of abstractive summarization for event-oriented news texts. The semantic relations between events like cause-effect relation are also extracted to help generate more coherent and informative summary in our system.
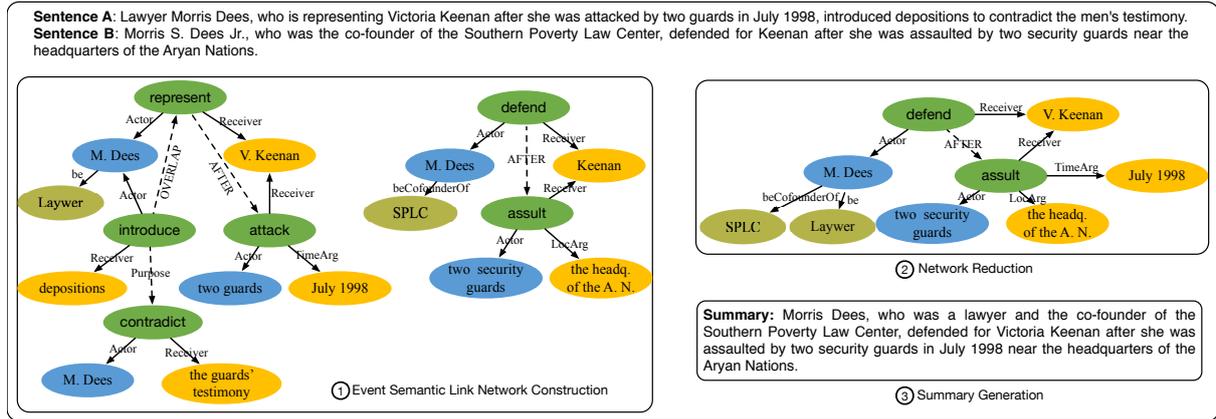

Figure 1: An example illustrating the framework of our summarization system.

Figure 1 illustrates the procedure of our system. Firstly, the semantic information of texts is represented by constructing event semantic link network (Zhuge, 2012). The semantic nodes of the network are events extracted from the source texts while semantic links are relations between events. Concept co-reference resolution and event co-reference resolution are both conducted within and cross documents to aggregate information from different places. Secondly, the event semantic link network is reduced to obtain connected and condensed summary network. A network reduction algorithm that makes use of the semantic links between event nodes is proposed to trade off among selecting salient information, maintaining coherence, and conveying correct and complete information. Finally, coherent and concise summary is automatically generated based on the summary network through sentences over-generation and greedy selection processes. The contributions of this work include:

- The abstractive summarization for event-oriented news texts is made by extracting fine-grained events and constructing event semantic link network as the abstract representation of source texts.
- An ILP-based network reduction algorithm using semantic links between events is proposed to obtain the most condensed, salient and coherent semantic information of source texts.
- Informative and concise summary is automatically generated based on the event semantic link network after reduction.

2   Event Semantic Link Network Construction

As shown in Figure 1, the first procedure of our system is to extract events and construct event semantic link network (ESLN). Within ESLN, semantic nodes are event mentions consisting of event actions and arguments. The action indicates the central meaning of an event, while the arguments render the attributes of an event (Ahn, 2006). In this work, each event is represented as a tree with the event action as its root node. The children of the root node are event arguments, including actor, receiver, time and location. The collapsed form of an event tree can be denoted as *e=Action (Actor, Receiver, TimeArg, LocArg)*. We use semantic relations between events as semantic links (subsection 2.3). ESLN provides an event-based abstract representation for news documents, which is a directed and connected graph.

The ESLN is constructed by: (1) extracting concepts from documents; (2) identifying event actions and extracting event arguments; (3) predicting the semantic links between event mentions.

2.1   Concept Extraction

All noun phrases extracted from documents are defined as concepts. To enrich the semantics of a concept, we model it as an object which consists of its core noun phrase and attributes. The attributes of a concept reflect the relationships between this concept and other concepts. A concept a implied by its core noun phrase is denoted by $a(\xrightarrow{r_1} c_1, \xrightarrow{r_2} c_2, \cdots, \xrightarrow{r_n} c_n)$ where c_i indicates another concept and r_i indicates a specific relation between concept a and concept c_i. Concept c_i is defined as an attribute of concept a.

Lexical features	word, lemma of the token and its surrounding tokens (five tokens to the left and right)
POS-tag features	part-of-speech tags of the token and its surrounding tokens (five to the left and right)
Syntactic features	the set of dependency relations of the token
Modifier features	modal modifiers, auxiliary verbs and negations.
Word vectors	100-dimensional GloVe word vector (Pennington et al., 2014)

Table 1: The features for the event identification model

Position features	the set of features that measure the distance between event actions (number of tokens) and their relative position (same sentence, adjacent sentences, adjacent event mentions)
Lexical features	word, lemma, stem, and pos-tag of both event actions as well as features indicating whether the word forms are the same, the semantic similarity between actions words, the word and lemma of each token between the action words
Syntactic features	syntactic path between the actions (dependency labels on the syntactic path between the actions), features indicating whether one action syntactically dominates the other, features indicating whether one is a predicate of an adverbial clause governed by the other event, and the set of dependency relations of both actions
Modifier features	the set of features that describe the modal, auxiliary, negation, and determination modifiers of both event actions
Word vectors	100-dimensional GloVe word vector of both event action words
Discourse features	the discourse relations between event mentions. We use the document-level discourse analysis method (Joty et al., 2013) to extract the discourse relations between event mentions.

Table 2: The features for the event relation prediction model

We extract concepts and their attributes based on dependency trees. Texts are preprocessed by Stanford CoreNLP pipeline (Manning et al., 2014). The dependency trees are transformed into semantic graph by pronoun resolution (Schuster et al., 2015). All named entities are identified as concepts. For other nouns, we expand on "`compound`", "`name`", "`amod`", "`neg`", "`nummod`" and "`dep`" dependency edges to build the basic noun-phrase concept. We also expand on "`appos`", "`acl`", "`acl:relcl`", "`nmod:of`" and "`nmod:poss`" edges for non-proper nouns, since these are relative clauses that convey important information.

To extract the attributes of a concept, we extract the relations between the concept and other related concepts. In order to differentiate with event actions, the valid syntactic patterns of relations between the head concept and its attributes is restricted as "`be`", "`be-NP-prep`" and "`be-AP-prep`" where NP indicates noun phrase and AP indicates adjective phrase, such as *"Morris Dees* **is** *a lawyer"* and *"Morris Dees* **is the co-founder of** *Southern Poverty Law Center"*. Several syntactic rules, which use the dependency labels (including "`nsubj`", "`appos`", "`nmod:of`" and "`nmod:poss`") between head tokens of concepts, are designed to detect those specific relations between concepts.

To aggregate information across documents, we need to recognize all concept co-references across documents. The co-reference resolution within single document has been conducted during the preprocessing stage by Stanford CoreNLP pipeline, so those resolution rules can be adopted. We formulate the co-references detection in a hierarchical agglomerative clustering framework similar to (Shen et al., 2013). A set of clusters are obtained and each cluster contains mentions refer to the same concept in the documents. For each cluster of co-referential concepts, we only reserve the most representative one and merge the attributes of all other mentions. For example, the concept *"Morris Dees"* in Figure 1.

2.2 Event Identification

The procedure of event identification consists of two steps: event action identification and event arguments extraction. The first step is formulated as a supervised classification task with features as shown in Table 1.

The arguments of an event are concepts related to the event action. Since we have extracted all concepts from documents, the argument extraction is to judge the argument type of each concept. We define in total fifteen dependency patterns using Semgrex expressions (Chambers et al., 2007). These patterns mainly capture the subject-predicate-object constructions, subject-predicate constructions, passive constructions, prepositional constructions and clausal constructions.

Since important events are usually mentioned many times in the documents. For example, in Figure 1 *"Victoria Keenan* was **attacked** by *two guards* in *July 1998"* and *"Keenan* was **assaulted** by *two security guards"* refer to the same event. To determine whether two event mentions are co-referential, both the event actions and event arguments are compared. We use WordNet-based similarity method (Pedersen et

al., 2004) to judge the semantic similarity between event actions. Two event mentions are identical only when the similarity between event actions is above a threshold (set as 0.8 after tuning) and corresponding event arguments are identical or co-referential. For all identical event mentions, we just reserve the most representative one and merge the relations and arguments of other mentions.

2.3 Event Relation Prediction

We leverage the sentence structures and discourse features in documents to infer the relations between events in order to construct an informative event semantic link network. Through analyzing large numbers of news texts, we find following types of semantic relations between events are very common:

- *Temporal link*. It indicates the temporal relations between two events, which consists of directed asymmetric links (BEFORE and AFTER) and symmetric links (OVERLAP). For symmetric links, we add two directed links in opposite directions between two event nodes;
- *Cause-effect link*, denoted by ce as in $e \xrightarrow{ce} e'$, for which the predecessor event e is a cause of its successor event e' and the successor event e' is an effect of its predecessor event e.
- *Purpose link*, denoted by pur as in $e \xrightarrow{pur} e'$, for which the successor event e' is the purpose of its predecessor event e. Event e' is to be realized through event e.
- *Means link*, denoted by mea as in $e \xrightarrow{mea} e'$, for which the event e' is a method or instrument which tends to make realization of event e more likely.
- *Condition link*, denoted by con as in $e \xrightarrow{con} e'$, for which the predecessor event e is a condition of its successor event e'. Realization of e' depends on realization of event e.
- *Sequential link*, denoted by seq as in $e \xrightarrow{seq} e'$, for which the event e' is a successor of event e. It usually describes a number of event actions with succession relationships.
- *Attribution link*, denoted by $attri$ as in $e \xrightarrow{attri} e'$, for which event e' is an attribution of event e, indicating its specific contents.

For predicting the semantic links between each pair of event nodes, we use an L2-regularized maximum entropy classifier with features as shown in Table 2.

In order to make the event semantic link network denser and more informative, we add "*Common Argument*" links between event nodes that share the same concept as argument. For example, "*Morris Dees defended for Keenan*" and "*Morris Dees contradicted the men's testimony*" both use concept "*Morris Dees*" as actor argument. After expanding the semantic links between events, we get a unified, connected and informative ESLN to represent the abstract information of source texts.

3 Summarization

The constructed ESLN is an abstract representation of source documents. We summarize the documents by summarizing the network and generate summary based on the reduced network. For event-based summarization, the summary network must contain the most salient events and concepts information. We model the summarization of ESLN as a structured prediction problem (Collins, 2002) that trades off among selecting salient information, maintaining coherence, and conveying correct and complete information.

Let E and C denote all the event nodes and concepts in ESLN, where each node $e \in E$ represents a unique event and each concept $c \in C$ is an argument of an event. To obtain the most salient and condensed summary network, we seek to maximize the summation of saliency scores of the selected events and concepts. For summary network which contains event set E' and concept set C', its saliency score is:

$$\sum_{e \in E'} \theta^T f(e) + \sum_{c \in C'} \psi^T g(c) \tag{1}$$

where $f(e)$ and $g(c)$ represent the features of event e and concept c respectively (described in Table 3). θ and ψ are vectors of feature weights for events and concepts respectively.

The network reduction problem is decoded as an integer linear programming (ILP) by incorporating some priori knowledge as constraints (§3.1). Features weights are estimated by using structured pre-

Concept Features	Concept Type	one binary feature indicates whether it is named entity
	Concept Freq.	one binary feature for each frequency threshold t=0/1/2/5/10
	Concept Head	word, lemma, pos, depth in the dependency tree (binarized by 5 depth thresholds)
	Concept Pos.	average and foremost position of sentences containing the concept (binarized by 5 thresholds)
	Concept Span	average and longest word span of concept (binarized by 5 thresholds)
Event Features	Action Word	word, lemma, pos of event action and depth in the dependency tree of sentence (binarized by 5 thresholds)
	Action Freq.	binary feature for each frequency threshold t=0/1/2/5/10, average and foremost position of sentences containing the concept
	Actor Arg.	all concept features of actor argument
	Receiver Arg.	all concept features of receiver argument. If don't contain receiver argument, all set as 0
	Time Argument	one binary feature indicates whether it contains time argument
	Location Arg.	one binary feature indicates whether it contains location argument
	Semantic Links	total number of links from and to the event node in event graph (binarized by 5 thresholds))

Table 3: Event and concept features (all binaries)

diction algorithm (§3.2). After obtaining the summary network, concise and coherent summary can be generated through sentences over-generation and greedy selection (§3.3).

3.1 Network Reduction

Let M and N be the total number of event nodes and concepts in source ESLN. We use e_i and c_j to represent the i-th event and j-th concept respectively. Let u_i and v_j be binary variables. u_i is set to 1 iff event e_i is selected and v_j is set to 1 iff concept c_j is selected.

The ILP maximization objective can be transformed into Equation 2, which contains two parts: the first part tends to select more important events; and the second part tends to select more concepts to increase information diversity and reduce redundancy in the summary.

$$\sum_{i=1}^{M} u_i \theta^T f(e_i) + \sum_{j=1}^{N} v_j \psi^T g(c_j) \tag{2}$$

To ensure the summary network could generate coherent summary and convey complete and correct information, the following groups of constraints are required:

Complete facts. To guarantee the selected event node convey complete fact, the following constraints are introduced:

$$\forall i, if c_j \in Arguments(e_i), v_j \geq u_i \tag{3}$$

$$\forall j, \sum_{i \in c_j.relatedEvents} u_i + \sum_{k \in c_j.attributes} v_k \geq v_j \tag{4}$$

$$\forall i, k, \text{if } e_i \xrightarrow{Attribution} e_k, u_i \leq u_k \tag{5}$$

Equation 3 ensures that if an event was selected, the arguments of the event should all be selected. Equation 4 guarantees that if a concept was selected, at least one event that it related to or an attribute that it has should be selected. These two constraints ensure the selected event or concept convey complete information. If event e_k is an attribution of event e_i, then e_k describes specific contents of event e_i. Equation 5 guarantees that if event e_i is selected, its attribution e_k must be selected.

Coherence. In order to generate coherent summary, the reduced summary network should be connected. Flow-based constraints have previously been used (Thadani and McKeown, 2013; Liu et al., 2015) to ensure the connectivity of subgraph. For each pair of event nodes e_i and e_k, the binary variable $l_{i,k}$ indicates the semantic link between them. Only if both e_i and e_k are selected and there is a link between them, $l_{i,k}$ can be set to 1, otherwise 0, which can be formulated as following:

$$\begin{aligned} \forall i, k, l_{i,k} \leq e_i, l_{i,k} \leq e_k \\ \text{if there is no link from } e_i \text{ to } e_k, l_{i,k} = 0 \end{aligned} \tag{6}$$

A set of single-commodity flow variables $f_{i,k}$ that each takes a non-negative integral value and represents the flow from event node e_i to e_k, are used to enforce the connectivity of summary network. We set a dummy "ROOT" node which is connected with only one selected event node in the ESLN (Equation 7), denoted as e_0. The root node sends up to M units of flows to the selected event nodes (Equation 8). Each selected node consumes one unit of flow (Equation 9). Flow can only be sent over a link if and only if the link variable l is 1 (Equation 10).

CONCEPT DESCRIPTION RULES

For concept $a(\xrightarrow{r_1} c_1, \xrightarrow{r_2} c_2, \cdots, \xrightarrow{r_n} c_n)$, the description of concept a can be:

1.**Appositive modifier** "$a, c_1, c_2...$", e.g. *"Morris Dees, civil rights lawyer, co-founder of Souther Poverty Low Center, ... "*

2.**Attributive clause** "a who/which/that $r_1 c_1 ...$ ", e.g. *"Morris Dees who was the co-founder of Southern Poverty Law Center and a civil rights lawyer ... "*

3.**Appositive modifier mixed with attributive clause**, e.g. *"Civil rights lawyer Morris Dees who was the co-founder of Southern Poverty Law Center ... "*

SENTENCE STRUCTURING RULES

if $e_1 \xrightarrow{after/before/overlap} e_2$, then generate "$e_1$ after/before/when e_2"; if $e_1 \xrightarrow{ce} e_2$, then generate "Because $e_1, e_2,$ " and "e_2 because e_1";

if $e_1 \xrightarrow{pur} e_2$, then generate "$e_1$ in order to e_2" and "e_1 so that e_2"; if $e_1 \xrightarrow{mea} e_2$, then generate "$e_1$ by e_2" and "e_1 by the way that e_2"

if $e_1 \xrightarrow{attri} e_2$, then generate "$e_1 e_2$", "$e_1$ about/on/in/with/at e_2" and "e_1 that e_2"; if $e_1 \xrightarrow{seq} e_2$, then generate "$e_1, e_2$", "$e_1$ and e_2";

Table 4: The set of concept description and sentence structure rules.

$$\forall i \geq 1, l_{0,i} \leq u_{i;}, \sum_{i=1}^{M} l_{0,i} = 1 \tag{7}$$

$$\sum_{i=1}^{M} f_{0,i} - \sum_{i=1}^{M} u_i = 0 \tag{8}$$

$$\forall k \geq 1, \sum_{i} f_{i,k} - \sum_{p} f_{k,p} - u_k = 0 \tag{9}$$

$$\forall i \geq 0, k \geq 1, M \cdot l_{i,k} - f_{i,k} \geq 0 \tag{10}$$

Length Constraint. To control the summary compression rate, the total number of selected events is limited less than L:

$$\sum_{i=1}^{M} u_i \leq L \tag{11}$$

where parameter L is set to control the graph size after reduction.

3.2 Feature Weights Estimation

We learn feature weights θ and ψ by training on a set of source ESLN paired with gold summary network. The source ESLN is constructed from source texts whereas the gold summary network is constructed from reference summaries and then mapped to the source ESLN by texts similarity method (Pilehvar et al., 2013). We formulate our estimation problem as follows:

$$-score(G^*) + max_G(score(G) + cost(G; G^*)) \tag{12}$$

where G^* denotes the gold summary network. $score()$ is defined in Equation 1. $cost(G; G^*)$ penalizes each event or concept in G but not in G^*, which can be easily incorporated into the linear objective in Equation 2. We optimize our objective using AdaGrad (Duchi et al., 2011) with l_2 regularization ($\lambda = 0.01$), with an initial step size 0.1. The ILP model is solved using Gurobi 6.5.2.

3.3 Summary Generation

Since each event node is structured as *e=Action (Actor, Receiver, TimeArg, LocArg)*, we can generate complete sentence efficiently for it using SimpleNLG (Gatt and Reiter, 2009). However, through experiments we find that low linguistic quality is the biggest problem with the generated sentences, which include syntax error, monotone sentence structure and repetition of the same noun phrases. To improve the linguistic quality of summary, we first over-generate large numbers of summary sentences and then use a greedy algorithm to select sentences with the best linguistic quality and no information overlapping.

Sentence over-generation. To generate a complete and informative sentence, both the description of concepts and the organization of sentence structures need to be settled in following ways:

- Each concept with several attributes can be described in different ways using concept description rules in Table 4.
- For each event node, we use SimpleNLG (Gatt and Reiter, 2009) tool to generate several different sentences, among which the description of concepts or the orders of arguments are different.

	Concepts	Events	Event Relations	After Expanding
avg#/topic	5206	1538	1089	10383

Table 5: The average number of concepts, events, event relations and relations after expanding of each topic in annotated DUC2007 (including 45 topics, each topic has 25 documents).

	precision	recall	F1-score
Concept Extraction	N/A	0.7928	N/A
Event Action identification	0.8532	0.8468	0.8499
Event Mention extraction	0.7272	0.7067	0.7168
Event Relations prediction	0.5894	0.6222	0.6054

Table 6: The performance of concepts extraction, events identification and event relations prediction.

- If two events share semantic links with each other, we merge them to generate one unified sentence by using corresponding sentence structuring rules in Table 4. Note that, when two events share the same actor concept, only one is reserved.
- For any two events that share the same actor, we merge them to generate one sentence using conjunction word "and" to connect event actions and arguments. Only one actor is kept as the subject.

Greedy selection. After the above step, we get large numbers of candidate summary sentences. Some of them would have information overlapping with each other if generated from the same event node. To improve the linguistic quality of summary, we iteratively select a sentence with the highest linguistic quality and delete sentences that have information overlapping with it from the candidate sentences set. The linguistic quality of sentence $s = \{w_1, w_2, \ldots, w_L\}$ is defined similarly as (Banerjee et al., 2015):

$$LQ(s) = 1 \Big/ \left(1 - \left(log_2 \prod_{t=1}^{L} P(w_t|w_{t-1}w_{t-2})\right) \Big/ L\right) \tag{13}$$

where L is the total number of words in sentence s; w_0 and w_{-1} both represent the beginning of sentence s. The 3-gram model $P(w_t|w_{t-1}, w_{t-2})$ is trained on the English Gigaword corpus (http://www.keithv.com/software/giga/).

The coherence constraints guarantee the selected summary network to be connected and have a flow from the ROOT node to selected nodes. The selected sentences are ordered based on the direction of flows to obtain a coherent summary.

4 Evaluation Results

4.1 Dataset and Experimental Settings

To evaluate the performance of our system, we use two datasets that have been widely used in multi-document summarization shared tasks: DUC 2006 and DUC 2007. Each task has a gold standard dataset consisting of document clusters and reference summaries. DUC 2007 was manually annotated by using annotation tool brat (http://brat.nlplab.org) to extract gold events and gold relations between events, which are used for training the event identification model and event relations prediction model. Table 5 shows the details of the annotated dataset.

The annotated dataset was split into training set (25 topics), development set (5 topics) and test set (15 topics). After training and tuning, the performance of our system is evaluated on the test set as shown in Table 6. An event mention is correctly extracted only if both the event action and event arguments are correct. Table 6 only shows the recall of concept extraction, because we extract all kinds of concepts, whereas only event arguments are annotated in the annotated dataset. The feature weights θ and ψ of event nodes and concepts are also estimated on the training set.

To evaluate the performance of our summarization model, we use both ROUGE (Lin and Hovy, 2003) and Pyramid (Nenkova and Passonneau, 2004) evaluation metrics.

4.2 Results with ROUGE Evaluation

ROUGE-1.5.5 toolkit was used to evaluate the quality of summary on DUC 2006 and DUC 2007 (test set) dataset. We differentiate the different components of our system by including and not including the coherence constraints in ILP-based network reduction algorithm and using the manually annotated gold ESLN in our system. Our systems are compared with several baselines: Centroid (Radev et al., 2000) and

	DUC2007(test set)				DUC2006			
	ROUGE-1	ROUGE-2	ROUGE-SU4	Pyramid	ROUGE-1	ROUGE-2	ROUGE-SU4	Pyramid
Baselines								
Centroid	0.36455	0.07032	0.12401	N/A	0.35211	0.06097	0.11570	N/A
LexRank	0.37501	0.07995	0.13528	N/A	0.36275	0.06830	0.12569	N/A
DUC NIST Baseline	0.33434	0.06479	0.11360	N/A	0.32082	0.05267	0.10408	N/A
AverageDUC	0.39684	0.09495	0.14671	N/A	0.37789	0.07483	0.12943	N/A
State-of-the-arts								
MultiMR	0.42041	0.10302	0.15385	N/A	0.39706	0.08508	0.13797	N/A
RA-MDS	0.403	0.092	0.146	N/A	0.391	0.081	0.136	N/A
ILPSumm (Abstractive)	0.41052	0.10060	0.15185	0.844	0.38564	0.07993	0.13279	0.811
PSM (Abstractive)	0.41917	0.10336	0.15608	0.851	0.39287	0.08173	0.13671	0.817
Our Systems								
ESLN with Coherence	**0.42723**	**0.10729**	**0.16181**	**0.865**	**0.39887**	**0.08756**	**0.14083**	**0.825**
ESLN w/o Coherence	0.41380	0.10106	0.14947	N/A	0.38586	0.08023	0.13618	N/A
Gold-ESLN with Coherence	0.44532	0.12229	0.17267	N/A	0.41162	0.09642	0.15348	N/A

Table 7: Comparison of ROUGE scores (F-score) and Pyramid scores on DUC 2006 and 2007(test set).

LexRank (Erkan and Radev, 2004). The performance of NIST baseline and the average ROUGE scores of all the participating systems (i.e. *AveDUC*) both for DUC 2006 and DUC 2007 main tasks are also listed. According to the results in Table 7, our systems significantly outperform (paired t-test with p<0.05) all the baselines, which demonstrates that extracting event information from texts and summarizing based on structured information is much more effective than summarizing on sentence level.

In addition, we also compare our system (ESLN with coherence) with several state-of-the-art summarization methods: graph-based extractive method MultiMR (Wan and Xiao, 2009), sparse-coding-based compressive method RA-MDS (Li et al., 2015), and two most recently developed abstractive methods ILPSumm (Banerjee et al., 2015) and PSM (Bing et al., 2015). The results show that our system significantly (paired t-test with p<0.05) outperforms all the other four systems.

The results also show that our system with coherence constraints achieves better performance than the counterpart without coherence constraints. So the coherence constraints are very helpful to select more salient and coherent information. Just as expected, the system using gold ESLN achieves the best performance. Incorrect dependency parsing and co-reference resolution will reduce the accuracy of extracting event information. On the other hand, it also verifies that the method that summarize texts based on accurate event information is effective.

4.3 Results with Pyramid Evaluation

Since ROUGE metric evaluates summaries by strict string matching, we also use the pyramid evaluation metric which can measure the summary quality beyond simply string matching. It involves semantic matching of summary content units (SCUs) so as to recognize alternate realizations of the same meaning, which provides a better metric for abstractive summary evaluation. We employ the automated version of pyramid scoring (set threshold value to 0.6) in (Passonneau et al., 2013). Table 7 shows the evaluation results of our system and two abstractive baselines on both DUC 2006 and DUC 2007(test set). The results show that our system significantly (p<0.05) outperform the two baselines on both datasets, which demonstrates that our system can generate more informative summary.

4.4 Discussion

Table 8 shows a comparison of summaries generated by our system and human on DUC 2007 dataset (D0701A). The results show that our summary behaves similarly to human summary in following aspects: (1) Aggregating information from different places. For example, the description of *"Morris Dees"* includes information from several different documents, which are extracted as attributes of concept *"Morris Dees"* in our system; (2) Organizing sentences coherently. The coherence constraints in ILP-based network reduction component ensure the selected event information to be coherent. (3) Clearly pronoun reference. The adjacent sentences with the same subject in the summary are post-edited by replacing subjects of successor sentences with appropriate pronouns. Even though we incorporate the sentences over-generation and greedy-selection components in our system, some sentences in the generated summaries also have few syntax errors. Most cases are because of non-accurate event extraction caused by incorrect dependency parsing or coreference resolution.

Summary by Our System: Morris Dees who was a crusader against in-tolerance, Keenans' attorney, the chief trial counsel, executive director and co-founder of the Southern Poverty Law Center used lawsuit to fight hate groups. He kept track of hate crime. He put East Peoria leader and won significant civil judgment against White Aryan Resistance and Ku Klux Klan in touch in Chicago with David Ostendorf. He formed a broad-based coalition and won a series of civil rights suit against other racist group in a campaign on race issue. He got an unwarranted slap in the Media Watch column and introduced photograph in the same issue. Southern Poverty Law Center, montgomery-based used civil suit and previously recorded a 20-percent increase in hate group. It battled racial bias and used civil law. It tracked hate group and won major legal fight against other white supremacist group and Ku Klux Klan. The 1973 federal lawsuit had the practical effect. The practical effect provided equal service to black...	Human-written Summary: The Southern Poverty Law Center is a non-profit research group based in Montgomery, Alabama that battles racial bias. It tracks US hate crimes and the spread of racist organizations. It covers right-wing extremists in its magazine Intelligence Report. Through its Teaching Tolerance program, it provides materials to teachers to promote interracial and intercultural understanding. It freely distributes booklets on combating hate to schools, mayors, police chiefs, and other interested groups and citizens. It advises city leaders faced with hate crimes. Morris Dees co-founded the SPLC in 1971 and is its chief trial counsel and executive director, following Julian Bond. Dees and the SPLC seek to destroy hate groups through multi-million dollar civil suits that go after assets of groups and their leaders. In six lawsuits based on hate crimes or civil rights abuses, they have never lost. They successfully sued the Ku Klux Klan and the related Invisible Empire Klan, United Klan of America and ...

Table 8: Example summary of D0701A in DUC2007 dataset by our system and the gold human summary (Only several leading sentences are displayed).

5 Related Work

Abstractive Multi-document summarization. Previous researches have shown that human write summaries through sentence aggregation and fusion (Cheung and Penn, 2013). Abstraction-based approaches that gather information across sentences boundaries have become more and more popular in recent years. Different abstractive summarization methods can be summarized into four technique routes: (1) sentence fusion based methods (Barzilay and McKeown, 2005; Filippova and Strube, 2008; Banerjee et al., 2015) first cluster sentences into several themes and then generate a new sentence for each cluster by fusing the common information of all sentences in the cluster; (2) information extraction based methods (Genest and Lapalme, 2011; Li, 2015) extract information units, such as Information Items or Basic Semantic Unit, as components for generating sentences; (3) summary revision based methods (Nenkova, 2008; Siddharthan et al., 2011) try to improve quality of summary by noun phrases rewriting and co-reference resolution; (4) pattern-based sentence generation methods (Wang and Cardie, 2013; Pighin et al., 2014; Bing et al., 2015) generate new sentences based on a set of sentence generation patterns learned from corpus or designed templates.

Recently, some works studied the use of deep learning techniques for abstractive summarization tasks, which use sequence-to-sequence generation techniques on single document or sentence summarization (Rush et al., 2015; Chopra et al., 2016). A multi-dimensional summarization methodology was proposed to transform the paradigm of traditional summarization research through multi-disciplinary fundamental exploration on semantics, dimension, knowledge, computing and cyber-physical society (Zhuge, 2016).

Event extraction. Event extraction is a traditional task in Information Extraction, which aims to recognize event mentions and arguments of predefined types (such as the ACE tasks). The works on event extraction either divide the task into separate subtasks, such as event-trigger extraction and argument extraction (Liao and Grishman, 2010; Hong et al., 2011) or model it jointly (Li et al., 2013; Li and Ji, 2014). These works mainly focus on predefined event and argument types. However, we focus on open-domain and more fine-grained event information extraction for multi-document summarization.

Abstract representations. With the development of Abstract Meaning Representation (AMR) (Banarescu et al., 2012), representing semantic information with graphs has been studied in such tasks as summarization (Liu et al., 2015) and event detection (Kai and Grishman, 2015). Although several techniques on parsing sentences to AMR (Flanigan et al., 2014; Wang et al., 2015) have been developed, the performance of AMR parsing is very limited at the present.

6 Conclusions

The approach proposed in this paper generates summary based on event information extraction and abstract representation, which achieves good performance on both DUC 2006 and DUC 2007 datasets. It generates new sentences based on structured event information and organizes sentences coherently based on semantic links. The experiment results show that the summaries generated by our system are relatively informative, coherent and compact, which demonstrates that the semantic link network based abstract representation of source texts is effective in making abstractive summarization.

References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2012. Abstract meaning representation (amr) 1.0 specification. In *EMNLP*, pages 1533–1544.

Siddhartha Banerjee, Prasenjit Mitra, and Kazunari Sugiyama. 2015. Multi-document abstractive summarization using ilp based multi-sentence compression. In *IJCAI*.

Regina Barzilay and Kathleen R McKeown. 2005. Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3):297–328.

Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca J Passonneau. 2015. Abstractive multi-document summarization via phrase selection and merging. *arXiv preprint arXiv:1506.01597*.

Nathanael Chambers, Daniel Cer, Trond Grenager, David Hall, Chloe Kiddon, Bill MacCartney, Marie-Catherine De Marneffe, Daniel Ramage, Eric Yeh, and Christopher D Manning. 2007. Learning alignments and leveraging natural logic. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 165–170. Association for Computational Linguistics.

Jackie Chi Kit Cheung and Gerald Penn. 2013. Towards robust abstractive multi-document summarization: A caseframe analysis of centrality and domain. In *ACL (1)*, pages 1233–1242.

Sumit Chopra, Michael Auli, Alexander M Rush, and SEAS Harvard. 2016. Abstractive sentence summarization with attentive recurrent neural networks. *NAACL*.

Michael Collins. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In *EMNLP*, pages 1–8. Association for Computational Linguistics.

John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159.

Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, pages 457–479.

Katja Filippova and Michael Strube. 2008. Sentence fusion via dependency graph compression. In *EMNLP*, pages 177–185. Association for Computational Linguistics.

Jeffrey Flanigan, Sam Thomson, Jaime G Carbonell, Chris Dyer, and Noah A Smith. 2014. A discriminative graph-based parser for the abstract meaning representation.

Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics.

Pierre-Etienne Genest and Guy Lapalme. 2011. Framework for abstractive summarization using text-to-text generation. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 64–73. Association for Computational Linguistics.

Goran Glavaš and Jan Šnajder. 2014. Event graphs for information retrieval and multi-document summarization. *Expert systems with applications*, 41(15):6904–6916.

Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou, and Qiaoming Zhu. 2011. Using cross-entity inference to improve event extraction. In *ACL*, pages 1127–1136. Association for Computational Linguistics.

Shafiq R Joty, Giuseppe Carenini, Raymond T Ng, and Yashar Mehdad. 2013. Combining intra- and multi-sentential rhetorical parsing for document-level discourse analysis. In *ACL*, pages 486–496.

Xiang Li Thien Huu Nguyen Kai and Cao Ralph Grishman. 2015. Improving event detection with abstract meaning representation. *ACL-IJCNLP 2015*, page 11.

Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In *ACL (1)*, pages 402–412.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *ACL (1)*, pages 73–82.

Piji Li, Lidong Bing, Wai Lam, Hang Li, and Yi Liao. 2015. Reader-aware multi-document summarization via sparse coding. *arXiv preprint arXiv:1504.07324*.

Wei Li. 2015. Abstractive multi-document summarization with semantic information extraction. *EMNLP*.

Shasha Liao and Ralph Grishman. 2010. Using document level cross-event inference to improve event extraction. In *ACL*, pages 789–797. Association for Computational Linguistics.

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *HLT-NAACL*, pages 71–78. Association for Computational Linguistics.

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A Smith. 2015. Toward abstractive summarization using semantic representations.

Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60.

Ani Nenkova and Rebecca Passonneau. 2004. Evaluating content selection in summarization: The pyramid method.

Ani Nenkova. 2008. Entity-driven rewrite for multi-document summarization.

Rebecca J Passonneau, Emily Chen, Weiwei Guo, and Dolores Perin. 2013. Automated pyramid scoring of summaries using distributional semantics. In *ACL (2)*, pages 143–147.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. Wordnet:: Similarity: measuring the relatedness of concepts. In *Demonstration papers at HLT-NAACL 2004*, pages 38–41. Association for Computational Linguistics.

Daniele Pighin, Marco Cornolti, Enrique Alfonseca, and Katja Filippova. 2014. Modelling events through memory-based, open-ie patterns for abstractive summarization. In *ACL (1)*, pages 892–901.

Mohammad Taher Pilehvar, David Jurgens, and Roberto Navigli. 2013. Align, disambiguate and walk: A unified approach for measuring semantic similarity. In *ACL (1)*, pages 1341–1351.

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.

Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.

Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei, and Christopher D Manning. 2015. Generating semantically precise scene graphs from textual descriptions for improved image retrieval. In *Proceedings of the Fourth Workshop on Vision and Language*, pages 70–80.

Chao Shen, Fei Liu, Fuliang Weng, and Tao Li. 2013. A participant-based approach for event summarization using twitter streams. In *HLT-NAACL*, pages 1152–1162.

Advaith Siddharthan, Ani Nenkova, and Kathleen McKeown. 2011. Information status distinctions and referring expressions: An empirical study of references to people in news summaries. *Computational Linguistics*, 37(4):811–842.

Kapil Thadani and Kathleen McKeown. 2013. Sentence compression with joint structural inference. In *CoNLL*, pages 65–74.

Xiaojun Wan and Jianguo Xiao. 2009. Graph-based multi-modality learning for topic-focused multi-document summarization. In *IJCAI*, pages 1586–1591.

Lu Wang and Claire Cardie. 2013. Domain-independent abstract generation for focused meeting summarization. In *ACL (1)*, pages 1395–1405.

Chuan Wang, Nianwen Xue, Sameer Pradhan, and Sameer Pradhan. 2015. A transition-based algorithm for amr parsing. In *HLT-NAACL*, pages 366–375.

Hai Zhuge. 2012. *The Knowledge Grid: Toward Cyber-Physical Society*. World Scientific.

Hai Zhuge. 2016. Multi-dimensional summarization in cyber-physical society. *Morgan Kaufmann*.