

Abstractive Multi-Document Summarization based on Semantic Link Network

Wei Li, Hai Zhuge*, *Senior Member, IEEE*

Abstract — The key to realize advanced document summarization is semantic representation of documents. This paper investigates the role of Semantic Link Network in representing and understanding documents for multi-document summarization. It proposes a novel abstractive multi-document summarization framework by first transforming documents into a Semantic Link Network of concepts and events and then transforming the Semantic Link Network into the summary of the documents based on the selection of important concepts and events while keeping semantics coherence. Experiments on benchmark datasets show that the proposed summarization approach significantly outperforms relevant state-of-the-art baselines and the Semantic Link Network plays an important role in representing and understanding documents.

Index Terms— Abstractive Summarization, Information Extraction, Multi-Document Summarization, Semantic Link Network

1 INTRODUCTION

THE general aim of multi-document summarization is to produce a concise and coherent summary representing the core information of a given set of documents. The summarization is called query-focused multi-document summarization if the summary is generated for answering a query [1]. Existing approaches can be categorized into extractive approaches and abstractive approaches in general. Extractive approaches generate summary by ranking sentences within documents while abstractive approaches are expected to generate new set of sentences as summary, which requires semantic representation of documents.

Semantic Link Network (SLN) is a semantics-rich self-organized semantic model for semantically organizing resources to support semantics-based services [2], [3], [4]. It consists of semantic nodes representing various resources, semantic links representing various relations between semantic nodes, and a semantic space that specifies semantics on nodes and links. SLN provides a semantics layer for implementing various advanced information services like abstractive document summarization [12][29].

Concepts and events are basic semantic representation units in a class of documents like news. Events indicate actions related to concepts, usually in form of “who did what to whom when and where”, representing the general concerns of readers on the documents like news. As an instance of Semantic Link Network, Semantic Link Network of concepts and events is suitable for representing the semantics of news documents.

SLN inspires an SLN-based multi-document summarization approach, which transforms the input documents

into a Semantic Link Network of concepts and events, summarizing the SLN to get a condensed and coherent SLN, and then transforming the condensed SLN into a concise text summary, as illustrated in Fig. 1.

The SLN-based summarization approach has the following main advantages: 1) Core information of documents can be selected more effectively by summarizing the SLN according to semantic features like semantic richness of semantic nodes and links [2][4][61]. 2) Co-referent events and co-referent concepts can be easily detected (e.g. events evoked by “attacked” and “assaulted”, and concepts “Morris Dees” and “Morris S. Dees Jr.”) and merged to avoid redundancy, and related information from different sentences can be aggregated. 3) Concepts and events in the summary-SLN carries useful syntactic information, which helps transform them into appropriate sentences. 4) Semantic links can increase the coherence and informativeness of generated summaries. For example, the “Temporal_after” link between events “defend.v” and “assault.v” make the summary sentence more coherent and informative.

To verify the advantages of the text-SLN-text summarization approach, we conduct experiments on three benchmark DUC (Document Understanding Conference) datasets. Experiments show that our approach outperforms several extractive summarization baselines and state-of-the-art abstractive baselines. It also empirically validates that SLN of concepts and events plays an important role in representing and understanding documents as a good summarization is based on the ability of representation and understanding.

• Wei Li is with the Institute of Computing Technology in Chinese Academy of Sciences and School of Computer and Control Engineering in University of Chinese Academy of Sciences. E-mail: weili.ict.kg@gmail.com

• * Corresponding author: Hai Zhuge is a chair in computer science in Aston University, he leads the International Research Network on Cyber-Physical-Social Intelligence consisting of Guangzhou University, Key Lab of Intelligent Information Processing at the Institute of Computing Technology in Chinese Academy of Sciences, Systems Analytics Research Institute at Aston University, and University of Chinese Academy of Sciences. E-mail: h.zhuge@aston.ac.uk

2 GENERAL ARCHITECTURE

The architecture of our approach illustrated in Fig. 2 mainly consists of three parts: Semantic Link Network construction, Semantic Link Network summarization and summary text generation.

Source Sentences: (1) Southern Poverty Law Center, a nonprofit organization in Birmingham, Ala, **tracks** hate groups across the United States. (2) 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. (3) 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.

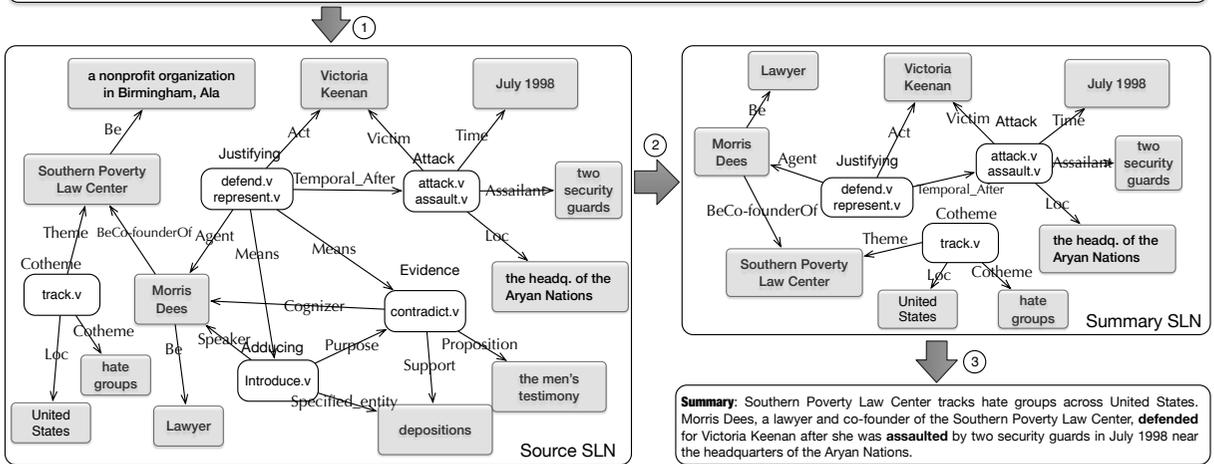


Fig. 1. An illustration of the process of our summarization approach. In the source sentences, concepts are underlined and events are in bold face. An SLN is constructed from the source sentences. Several semantic links between concepts such as “BeCo-founderOf”, and semantic links between events such as “Temporal_After”, are extracted and shown in the SLN. The semantic links between concepts connect some events without direct relations (e.g. events evoked by “tracks” in (1) and “defended” in (3)), making the SLN denser and richer. Co-referent events (e.g. events evoked by “attacked” in (2) and “assaulted” in (3)) and co-referent concepts (e.g. “Morris Dees” in (2) and “Morris S. Dees Jr.” in (3)) are both detected and merged. Then, the source-SLN is summarized into a summary-SLN, which contains several salient events and concepts, as well as the semantic links between them. Finally, a text summary is generated from the summary-SLN based on sentence generation templates.

2.1 Semantic Link Network Construction

Semantic Link Network construction contains two components: concept extraction with relation identification (Section 3.1), and event extraction with relation identification (Section 3.2-3.3). Since the existing definition of event schemas in ACE¹¹ (Automatic Context Extraction) 2005 event extraction program is rather incomplete and coarse, we redefine the event schemas that has wider coverage and are more fine-grained based on semantic frames in FrameNet corpus. New hierarchical event schemas are built by extracting all frames expressing event information and the frame-to-frame relations between them. Structured event representation is automatically extracted from documents based on the hierarchical event schemas. Various semantic links between events, such as *Cause-Effect* link, *Purpose* link and *Temporal* link, are also defined and extracted from documents automatically. Semantic links between concepts are further extracted to link the events that have no explicit connection.

2.2 Semantic Link Network Summarization

The summarization of SLN is modeled as a structure prediction problem [5] that trades off among selecting salient information, maintaining coherence, and conveying correct and complete information. The semantics richness features, such as the frequency of semantic nodes, frequency and diversity of semantic links, are used to help extract the core information of documents more effectively [2][61]. Several network-flow based coherence constraints are used to ensure the semantics coherence of the summary SLN, which are proved to be effective for improving both the informativeness and coherence of the

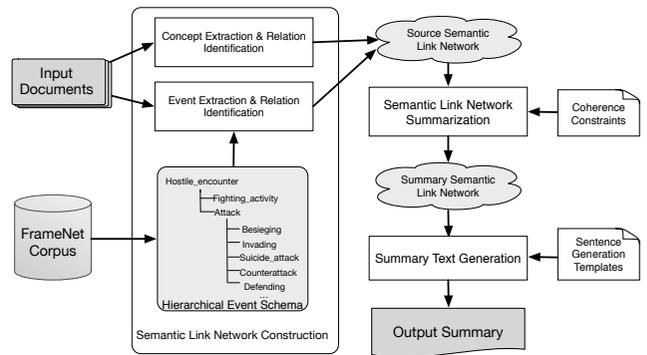


Fig. 2. The architecture of the SLN-based approach.

resulting summaries (presented in Section 4 in detail).

2.3 Summary Text Generation

Both the structures of events and concepts in SLN carry useful syntactic information, which is helpful for transforming a summary SLN into a grammatical and fluent text summary. A set of concept description templates, event description templates and sentence structuring rules are extracted from the FrameNet annotated corpus. To improve the linguistic quality of generated summaries, more candidate sentences are firstly generated based on the learned templates, and then greedily selected to compose a fluent and coherent text summary without information redundancy (presented in Appendix A in detail).

¹¹ <https://www ldc.upenn.edu/collaborations/past-projects/ace>

3 SEMANTIC LINK NETWORK CONSTRUCTION

3.1 Concept Extraction with Relation Identification

Concepts and events are two basic information units in a class of documents like news. Most concepts usually act as arguments of events. The relations between concepts also carry important relational information between events. Importing concepts and relations between them into the Semantic Link Network representation of documents connects some events without direct relations, which make the SLN denser and semantically richer. Fig. 1 shows an example that the relation between concepts “*Morris Dees*” and “*Southern Poverty Law Center*” connects the events related to them (e.g. events evoked by “*tracks*” and “*defended*”), making the summary more informative and coherent.

All noun phrases extracted from documents based on dependency trees are regarded as concepts. Texts are pre-processed by Stanford CoreNLP pipeline [6]. The dependency trees are transformed into semantic graph by pronoun resolution [7]. All named entities are identified as concepts. For other nouns, we expand on “*compound*”, “*name*”, “*amod*”, “*neg*”, “*nummod*” and “*dep*” dependency labels to build the basic noun-phrase concept. We also expand on “*appos*”, “*acl*”, “*acl:relcl*”, “*nmod:of*” and “*nmod:poss*” labels for non-proper nouns, since these are relative clauses that convey important information.

Relations between concepts are usually indicated by phrases between concepts within text. In order to differentiate event triggers (usually are indicated by verbs), the valid syntactic patterns of relations are 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*”. The evaluation results on 100 random sentences show that the three patterns are able to cover over 90% relations except for the relations represented in verbs that can be recognized as events (presented in Appendix B.4 in detail). 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, all coreferent concepts need to be recognized both within each document and on all documents. The co-reference resolution within each document has been conducted during the preprocessing stage by Stanford CoreNLP pipeline, so those resolution rules can be adopted. The concept co-reference resolution across documents is processed by a hierarchical agglomerative clustering algorithm as introduced in [8]. A set of clusters is obtained and each cluster contains concept mentions refer to the same concept in the documents. For each cluster of co-referential concepts, only the most representative one is reserved, and the relations of all other concept mentions are merged into it. For example, the concept “*Morris Dees*” in Fig. 1 is merged from “*Morris Dees*” in source sentence (1) and “*Morris S.*

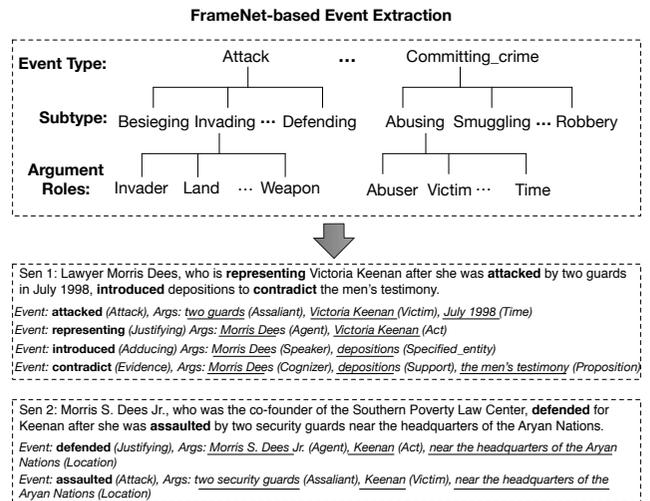


Fig. 3. FrameNet-based event extraction. The top part shows the hierarchical event schemas defined based on FrameNet. The bottom part shows the event extraction results for two example sentences based on the new event schemas.

Dees Jr.” in source sentence (2) as well as their semantic relations “*Be*” with “*Layer*” and “*BeCo-founderOf*” with “*Southern Poverty Law Center*”.

3.2 FrameNet-based Event Extraction

Event extraction is the task of extracting structured event information from text according to pre-defined event schemas. Most of existing approaches are based on ACE (Automatic Context Extraction) 2005 event extraction program, which defines 8 types of events with 33 subtypes in total. Each event is defined as a composition of an event type, an event trigger and a set of argument roles. Due to the limited number of defined event schemas, the event extraction results usually miss much salient event information, so that they are not suitable for practical applications. The FrameNet[®] corpus is a linguistic corpus containing complete definition of semantic frames and frame-to-frame relations. Frames in FrameNet share highly similar structures with event schemas in ACE [9]. This work not only redefines event schemas based on semantic frames in FrameNet [10] but also builds a hierarchy of event schemas with leveraging the frame-to-frame relations. Employing the hierarchical event schemas as an instance of the semantic space of SLN [4][60][61][62], a probabilistic event extraction algorithm is further proposed to extract structured event information from text, which consists of two parts: (1) event trigger extraction, and (2) event arguments extraction.

3.2.1 Hierarchical Event Schemas based on FrameNet

The FrameNet is a taxonomy of manually identified semantic frames. A frame is defined as a composition of a frame type, a set of Frame Elements (FEs) and a list of Lexical Units (LUs). A LU is a word or phrase that evokes the corresponding frame. FEs indicate the set of semantic roles associated with the participants of the frame. Most frames contain a set of exemplars with annotated LUs and FEs. Moreover, there is a set of labeled relations between frames. FrameNet[®] in total contains more than 1200 vari-

[®] <https://framenet.icsi.berkeley.edu/fndrupal/>

[®] We use FrameNet 1.7 version.

ous frames, 13500 LUs with 202000 manually annotated exemplars and 1800 frame-to-frame relations. Three main types of frame-to-frame relations are defined as following:

Inheritance. “Frame A inherits from frame B” indicates that A corresponds to an equal or more specific fact about B. For example, “invading” inherits from “attack”.

SubFrame. “Frame A is a subFrame of frame B” indicates that B is a complex frame referring to sequences of separate states and transitions, and A is one of them. For example, “Committing_crime” is a subFrame of “Crime_scenario”.

Using. “Frame A using frame B” indicates that a part of the scene evoked by A refers to B. It is a directional relation, where B is usually more abstract than A (e.g., “Abusing” using “Cause_harm”).

Many frames in FrameNet actually express certain types of events in ACE, such as frames “Attack” and “Invading”. The structure of frames in FrameNet is also similar to event schemas in ACE. The LUs and FEs of a frame play similar roles as the event triggers and argument roles of events in ACE, respectively.

The above observations motivate us to construct new event schemas by extracting all frames expressing event information from FrameNet. The frame-to-frame relations *Inheritance*, *Subframe* and *Using* are also used to build a hierarchy of event schemas. The frame types are directly used as event types and their FEs are directly used as argument roles, as shown in Fig. 3. For each frame in FrameNet, there are core FEs and non-core FEs. Core FEs are conceptually or syntactically necessary to the central meaning of the frame (analogous to the core arguments ARG0-ARG5 in PropBank). In contrast, the non-core FEs loosely correspond to syntactic adjuncts and carry broadly applicable information such as time and place. In our event schema definition, we reserve all core FEs and only some of the common non-core FEs such as *Time* and *Place* as argument roles.

The advantages of our hierarchical event schemas⁴ include the following aspects:

- (1) It contains 655 in total well-defined event schemas. Based on our event schemas, more complete event information can be extracted from text. For example, in Fig. 3, four events are extracted from the first example sentence and two events are extracted from the second example sentence. However, only one event (“Attack”) can be extracted from both sentences by the ACE evaluation.
- (2) The event types in our event schemas are more fine-grained than that in ACE. For example, all kinds of violent acts, such as street fights and wars, are treated as a single type “Attack”, however, there are richer definitions in our event schemas (e.g. “Attack” has five subtypes in total in our event schemas).
- (3) The large manually annotated dataset in FrameNet, including exemplars and full text annotation, can be directly used as training and evaluation data for the event extraction model.

3.2.2 Event Trigger Extraction

The task of event trigger extraction needs to identify all event triggers and classify their event types. For example, the first example sentence in Fig. 3 contains four event triggers: “representing”, “attacked”, “introduced” and “contradict”. They evoke events “Justifying”, “Attack”, “Ad-ducing” and “Evidence” respectively. The processes of trigger identification and classification are performed in a unified manner, which have been proved to be superior to be handled separately [11]. Specifically, a log-linear classification model is used to classify each trigger candidate into one of candidate event types (including a “None” class to differentiate non-event triggers).

In our event schemas, both verbs and nouns can evoke events. Instead of treating all words as trigger candidates, we build a valid event lexical unit (LU) set by extracting all LUs that evoke event frames from FrameNet, including annotations in exemplars and full texts. Then, the event LU set is expanded by using synsets in WordNet and including all their morphological variants. When given a sentence, a word or phrase is considered as a trigger candidate if it is contained by the event LU set. For each trigger candidate, its candidate event types are restricted to those events that include the trigger candidate as a LU. The simple approach is able to cover 90.94% of the gold-standard event triggers in the test set. The trigger candidate extraction approach not only reduces the label bias of the training data (most are non-event triggers) but also largely reduces the number of candidate event types for each trigger candidate (from 656 to about 10 on average).

For a given sentence $X = \langle x_1, \dots, x_n \rangle$ with candidate event triggers $T = \langle t_1, \dots, t_m \rangle$, t_i denotes the i th trigger word and t_i^l denotes its lemma. Let \mathcal{L} be the event LU set, \mathcal{L}_f be the subset of event LU set which evokes a particular event type f . Let \mathcal{L}^l and \mathcal{L}_f^l be the lemmatized versions of \mathcal{L} and \mathcal{L}_f . The set of candidate event types for t_i is defined as $\mathcal{F}_i = \{f | \forall f, t_i^l \in \mathcal{L}_f^l\} \cup \{None\}$ (“None” indicates a non-event class). We seek a list of event types $F = \langle f_1, \dots, f_m \rangle$ for all target trigger words $T = \langle t_1, \dots, t_m \rangle$.

For each trigger candidate t_i , we aim to find the best event type from candidate event types \mathcal{F}_i :

$$f_i = \operatorname{argmax}_{f \in \mathcal{F}_i} p_\theta(f | t_i, X) \quad (1)$$

A log-linear model is used to model the event type classification problem as following:

$$p_\theta(f | t_i, X) = \frac{\exp(\theta^T g(f, t_i, X))}{\sum_{f' \in \mathcal{F}_i} \exp(\theta^T g(f', t_i, X))} \quad (2)$$

where g indicates a feature vector for event type classification (shown in Table D.1 in Appendix D) and θ denotes the corresponding feature weights.

We discriminatively train the event type classification model by maximizing the following log-likelihood for training datasets $\langle X^{(j)}, T^{(j)}, F^{(j)} \rangle$:

$$\max_\theta \sum_{j=1}^d \sum_{i=1}^{m_j} \log p_\theta(f_i^{(j)} | t_i^{(j)}, X^{(j)}) \quad (3)$$

where d denotes the total number of annotated sentences in training dataset and m_j indicates the total number of target triggers in sentence $X^{(j)}$. $t_i^{(j)}$ and $f_i^{(j)}$ denote the i -

⁴ <https://github.com/weili-ict/EventSchemasBasedOnFrameNet>

th trigger word and its corresponding event type in sentence $X^{(j)}$ respectively.

3.2.3 Event Arguments Extraction

Event arguments are concepts which play different roles in an event. Event argument extraction is to identify concepts that act as arguments, and classify their argument roles. For example, in the first sentence in Fig. 3, “Victoria Keenan” plays the role of “Victim” and “two guards” plays the role of “Assailant” in event “Attack”. This work treats all concepts extracted from a sentence as argument candidates, and then utilizes a unified log-linear classification model to identify event arguments and classify their argument roles.

Given a sentence $X = \langle x_1, \dots, x_n \rangle$, the set of target event triggers are denoted as $T = \langle t_1, \dots, t_m \rangle$ and the corresponding event types are denoted as $F = \langle f_1, \dots, f_m \rangle$. For each event type f_i , let $\mathcal{R}_{f_i} = \{r_1, \dots, r_{|\mathcal{R}_{f_i}|}\}$ denote its argument roles. All concepts extracted from the sentence (as introduced in Section 3.1) are treated as argument candidates S for filling any role $r \in \mathcal{R}_{f_i}$. They are able to cover 98.4% of the gold-standard argument mentions in the test set.

For each event trigger t_i with event type f_i and argument roles $\mathcal{R}_{f_i} = \{r_1, \dots, r_{|\mathcal{R}_{f_i}|}\}$, let A_i be the mapping of concepts in S to argument roles in $\mathcal{R}_{f_i} = \mathcal{R}_{f_i} \cup \{None\}$ (“None” indicates a non-argument role, used to differentiate non-argument concepts). The argument role of each concept $s \in S$ is represented as following:

$$A_i(s) = \operatorname{argmax}_{r_k \in \mathcal{R}_{f_i}} p_{\vartheta}(r_k | s, f_i, t_i, X) \quad (4)$$

where $A_i(s)$ denotes the argument role of concept s .

We use a conditional log-linear model on candidate concepts for each role of each evoked frame:

$$p_{\vartheta}(A_i(s) = r_k | s, f_i, t_i, X) = \frac{\exp(\vartheta^T h(s, r_k, f_i, t_i, X))}{\sum_{r \in \mathcal{R}_{f_i}} \exp(\vartheta^T h(s, r, f_i, t_i, X))} \quad (5)$$

where h indicates a feature vector for argument role identification (shown in Table D.2 in Appendix D) and ϑ denotes the corresponding feature weights.

We train the argument identification model by maximizing the log-likelihood of training dataset $\langle X^{(j)}, T^{(j)}, F^{(j)}, A^{(j)} \rangle$ as following:

$$\max_{\vartheta} \sum_{j=1}^d \sum_{i=1}^{m_j} \sum_{s \in S} \log p_{\vartheta}(A_i^{(j)}(s) | f_i^{(j)}, t_i^{(j)}, \mathcal{R}_{f_i}^{(j)}, X^{(j)}) \quad (6)$$

where $A_i^{(j)}(s)$ indicates the annotated argument role of concept s for event trigger $t_i^{(j)}$ with event type $f_i^{(j)}$.

3.3 Event Relation Extraction

The sentence structures and discourse features in documents are leveraged to infer the relations between events in order to construct an informative Semantic Link Network. The common types of semantic links between events are defined as following:

- (1) **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. Clue words, such as “before”, “after”, “when” and “during”, can help detect

the corresponding temporal links.

- (2) **Cause-effect link**, denoted as ce or $e - ce \rightarrow e'$, for which the predecessor event is a cause of its successor and the successor is an effect of its predecessor. Common clue words include “because”, “since”, “so”, “lead to”, “therefore”, “thus”, “hence” etc.
- (3) **Purpose link**, denoted as pur or $e - pur \rightarrow e'$, for which the successor event e' is the purpose of its predecessor event e . Event e' is to be realized through event e . Common clue words include “to”, “in order to”, “so that” etc.
- (4) **Means link**, denoted as mea or $e - mea \rightarrow e'$, where event e' is an approach or instrument that tends to realize event e more likely. Common clue words include “by”, “through” etc.
- (5) **Condition link**, denoted as con or $e - con \rightarrow e'$, where the predecessor event e is a condition of its successor event e' . Realization of e' depends on realization of e . Common clue words include “if”, “only if”, “only when” etc.
- (6) **Sequential link**, denoted by seq or $e - seq \rightarrow e'$, where event e' is a successor of event e . It usually describes a list of event actions with succession relations. Common clue words include “and”, “also”, “or”, “then” etc.
- (7) **Attribution link**, denoted as $attri$ or $e - attri \rightarrow e'$, where event e' is an attribution of e , indicating its specific contents. Common clue words include “think”, “say”, “suggest”, “believe”, “hope”, “suppose”, etc.

For predicting semantic links between each pair of event nodes, we use an L2-regularized maximum entropy classifier with features including event type features, event argument features, syntactic features, discourse features, etc (shown in Table D.3 in Appendix D).

Some important events are usually mentioned several times in different places with different expressions. For example, in Fig. 1 “Victoria Keenan was **attacked (Attack)** by two guards” from example sentence (2) and “Keenan was **assaulted (Attack)** by two security guards” from example sentence (3) refer to the same event. To aggregate information from different places, events that are co-referential and indicate the same fact need to be identified. Two event mentions are co-referent only if they have the same event type and the corresponding core arguments are identical or co-referential. For all co-referential event mentions, only the most representative one is reserved and the relations and arguments of other event mentions are incorporated into it.

After predicting semantic links and resolving co-reference between events, a unified, connected and informative SLN is obtained as the abstract semantic representation of source documents.

4 SEMANTIC LINK NETWORK SUMMARIZATION

The summary of SLN must contains the most salient events and concepts information, and remains semantics coherent. We model the summarization of SLN as a struc-

ture prediction problem [5].

Let E and C denote all event nodes and concept nodes in SLN respectively, where each node $e \in E$ denotes a unique event and each node $c \in C$ denotes a unique concept. To obtain the most salient and condensed summary SLN, we seek to maximize the saliency scores of the selected events and concepts. For a summary SLN which contains event node set E' and concept node set C' , its saliency score is defined as following:

$$\sum_{e \in E'} \theta^T f(e) + \sum_{c \in C'} \psi^T g(c) \quad (7)$$

where $f(e)$ and $g(c)$ represent the features of event e and concept c respectively. A set of semantic features, such as link frequency (total number of links from and to an event/concept node in SLN), link diversity⁸ (total types of links from and to an event/concept node in SLN [61]), topic information (whether an event trigger/concept occurs in the topic description), are used to extract the core information of documents. The details are described in Table D.4 in Appendix D. θ and ψ are feature weight vectors estimated by structured perceptron algorithm from a set of source-SLNs paired with gold summary-SLNs (presented in Appendix C in detail).

The SLN summarization problem is decoded as an integer linear programming (ILP) by incorporating some a-priori knowledge as constraints (presented in Section 4.1).

4.1 ILP-based SLN Summarization

Let $|E|$ and $|C|$ denote the total number of event nodes and concept nodes in a source SLN. e_i and c_j represent the i -th event and the j -th concept respectively. Let u_i and v_j be binary variables. u_i is set as 1 if and only if event e_i is selected and v_j is set as 1 if and only if concept c_j is selected.

The objective function of the SLN summarization process can be represented by Equation 8, which contains two parts: the first part tends to select more important events to increase information saliency of the summary; and the second part tends to select more important concepts to increase information saliency as well as information diversity of the summary.

$$\sum_{i=1}^{|E|} u_i \theta^T f(e_i) + \sum_{j=1}^{|C|} v_j \psi^T g(c_j) \quad (8)$$

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

Complete facts. To guarantee the selected events and concepts convey complete fact, the following constraints are introduced:

$$\forall i, \text{ if } c_j \in \text{Arguments}(e_i), v_j \geq u_i \quad (9)$$

$$\forall j, \sum_{i \in c_j \text{ relatedEvents}} u_i + \sum_{k \in c_j \text{ relatedConcepts}} v_k \geq v_j \quad (10)$$

$$\forall i, k, \text{ if } e_i \text{ Attribution } e_k, u_i \leq u_k \quad (11)$$

Equation 9 ensures that if an event was selected, the arguments of the event should all be selected. Equation 10 guarantees that if a concept was selected, at least one event whose arguments include the concept or another semantically linked concept, should be selected. These two constraints ensure the selected events or concepts to convey complete facts. If event e_k is an attribution of event e_i , then e_k describes specific content of event e_i . Equation 11 guarantees that if event e_i is selected, its attribution e_k must be selected.

Coherence. In order to generate semantics coherent summary, the semantic nodes in the summary-SLN should be connected. The role of connectivity of SLN in summarization has been verified on some semantic links such as is-part-of link and cause-effect link [12]. In this work, a set of network flow-based constraints are used to ensure the connectivity of the summary-SLN [13].

In order to unify the representation of event semantic nodes and concept semantic nodes, we use n_i to denote whether the i -th semantic node (either an event node or a concept node) in the source SLN is selected. For each pair of semantic nodes n_i and n_k , the binary variable $l_{i,k}$ indicates the semantic link between them. Specifically, the argument concepts of an event are all linked with the event by argument role links. Only if both n_i and n_k are selected and there is a link between them, $l_{i,k}$ is set as 1, otherwise 0, which can be formulated as following:

$$\begin{aligned} \forall i, k, l_{i,k} \leq n_i, l_{i,k} \leq n_k \\ \text{if there is no link from } n_i \text{ to } n_k, l_{i,k} = 0 \end{aligned} \quad (12)$$

A set of single-commodity network-flow variables $f_{i,k}$ that each takes a non-negative integral value and represents the flow from the semantic node n_i to n_k , were used to enforce the connectivity of the summary SLN. We set a dummy "ROOT" node which is connected with only one selected semantic node in the source SLN (Equation 13), denoting as n_0 . The root node sends up to $(|E| + |C|)$ units of flow to the selected semantic nodes (Equation 14). Each selected node consumes one unit of flow (Equation 15). Flow can only be sent over a link if and only if the link variable l is 1 (Equation 16).

$$\forall i \geq 1, l_{0,i} \leq n_i, \sum_{i=1}^{|E|+|C|} l_{0,i} = 1 \quad (13)$$

$$\sum_{i=1}^{|E|+|C|} f_{0,i} - \sum_{i=1}^{|E|+|C|} n_i = 0 \quad (14)$$

$$\forall k \geq 1, \sum_i f_{i,k} - \sum_p f_{k,p} - n_k = 0 \quad (15)$$

$$\forall i \geq 0, k \geq 1, (|E| + |C|) \cdot l_{i,k} - f_{i,k} \geq 0 \quad (16)$$

⁸The argument roles are also viewed as semantic links between concepts and events.

TABLE 1
SUMMARY OF THE DATASETS

	DUC2005	DUC2006	DUC2007
Number of topics	50	50	45
#Docs per topic	32	25	25
Data Source	TREC	AQUAINT	AQUAINT
Summary Length	250 words	250 words	250 words

TABLE 2
COMPARISON OF ROUGE RESULTS ON DUC 2006

System	ROUGE-1	ROUGE-2	ROUGE-SU4
Centroid	0.36198* [0.34857-0.37480]	0.07150* [0.06456-0.07813]	0.12104* [0.11382-0.12779]
TextRank	0.37976 [0.36689-0.39276]	0.07983* [0.07225-0.08789]	0.13175* [0.12412-0.13945]
LexPageRank	0.37331* [0.36070-0.38524]	0.07151* [0.06432-0.07918]	0.12553* [0.11851-0.13287]
NIST Baseline	0.31691* [0.29614-0.33412]	0.05234* [0.04453-0.06058]	0.10244* [0.09405-0.11042]
AverageDUC	0.37791	0.07478	0.13098
MedianDUC	0.37868* [0.36796-0.38981]	0.07167* [0.06522-0.07809]	0.12656* [0.12069-0.13249]
MultiMR	0.39349 [0.38148-0.40648]	0.08511* [0.07693-0.09365]	0.13998* [0.13223-0.14780]
Submodular1	0.39260 [0.38140-0.40471]	0.08788* [0.08012-0.09632]	0.14137 [0.13450-0.14916]
Submodular2	0.40263 [0.39144-0.41520]	0.09134* [0.08219-0.10075]	0.14407 [0.13672-0.15214]
RA-MDS [23]	0.391	0.081	0.136
ILPSumm	0.38536 [0.37479-0.39591]	0.08027* [0.07285-0.08751]	0.13659* [0.12980-0.14327]
PSM	0.38584 [0.37363-0.39855]	0.08000* [0.07342-0.08743]	0.13290* [0.12607-0.14011]
Our System	0.39017 [0.37693-0.40205]	0.11033 [0.10162-0.11865]	0.14844 [0.14070-0.15613]

(*indicates that the improvement of our system over the baseline is statistically significant.)

The above constraints ensure the connectivity of the summary SLN, which also facilitates the process of salient information selection (shown in Table 7 and Table 8).

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

$$\sum_{i=1}^{|E|} u_i + \sum_{j=1}^{|C|} v_j \leq L \quad (17)$$

where L is a hyper-parameter tuned to control the length of the generated summary.

5 EVALUATION

5.1 Dataset and Experimental Settings

To evaluate the performance of our summarization system, we conducted a series of experiments based on the benchmark DUC 2005, DUC 2006 and DUC 2007 datasets. They contain 50, 50 and 45 topics, respectively. Each topic contains a topic description and a document set with 32, 25 and 25 news articles respectively. The length of the reference summaries for each topic is limited by 250 words. Table 1 gives a short summary of the three datasets. In the experiments, evaluations are primarily conducted on DUC 2006 and DUC 2007, and DUC 2005 is used to enrich the training examples. Specifically, we use DUC 2005 and DUC 2007 as training data to evaluate on

TABLE 3
COMPARISON OF ROUGE RESULTS ON DUC 2007

System	ROUGE-1	ROUGE-2	ROUGE-SU4
Centroid	0.39125* [0.37860-0.40392]	0.08946* [0.08208-0.09690]	0.13946* [0.13187-0.14715]
TextRank	0.40646* [0.39325-0.42011]	0.10088* [0.09229-0.10960]	0.15091* [0.14244-0.15968]
LexPageRank	0.39386* [0.38113-0.40656]	0.08336* [0.07626-0.09103]	0.13949* [0.13119-0.14786]
NIST Baseline	0.33100* [0.31484-0.34654]	0.06435* [0.05675-0.07251]	0.11118* [0.10339-0.11927]
AverageDUC	0.39684	0.09495	0.14671
MedianDUC	0.40704* [0.39568-0.41842]	0.10169* [0.09383-0.10989]	0.15455* [0.14647-0.16268]
MultiMR	0.40874* [0.39599-0.42172]	0.10264* [0.09302-0.11194]	0.15191* [0.14374-0.16038]
Submodular1	0.41751 [0.40478-0.43048]	0.10969* [0.10155-0.11804]	0.15927* [0.15141-0.16721]
Submodular2	0.42694 [0.41411-0.44069]	0.10988* [0.10109-0.11917]	0.16129* [0.15288-0.17024]
RA-MDS [23]	0.403	0.092	0.146
ILPSumm	0.40961* [0.39767-0.42184]	0.10519* [0.09670-0.11388]	0.15535* [0.14740-0.16333]
PSM	0.41069* [0.39688-0.42363]	0.10081* [0.09276-0.10911]	0.15204* [0.14439-0.16003]
Our System	0.42765 [0.41593-0.43826]	0.14177 [0.13212-0.15102]	0.17481 [0.16686-0.18225]

TABLE 4
COMPARISON OF AUTOMATIC PYRAMID RESULTS

System	Our System	ILPSumm	PSM
DUC2006	0.838	0.811*	0.817*
DUC2007	0.885	0.844*	0.851*

DUC 2006, and then use DUC 2005 and DUC 2006 as training data to evaluate on DUC 2007.

We use ROUGE-1.5.5 toolkit [14] for evaluation. Three ROUGE metrics are used: ROUGE-1 (unigram-based), ROUGE-2 (bigram-based), and ROUGE-SU4 (based on a skip bigram with a maximum skip distance of four). In order to truncate summaries longer than the length limit, we use the “-l 250” option in ROUGE toolkit. We also use the “-m” option for word stemming⁶. In addition, the Pyramid [15] is used to evaluate the performance of our system comparing with several state-of-the-art abstractive approaches.

5.2 Baselines

In the experiments, we compare our proposed approach with various baselines, including the DUC NIST baseline, the average ROUGE scores and the median scores of all DUC participating systems (i.e. *AverageDUC* and *MedianDUC*), several extractive baselines, and several state-of-the-art abstractive MDS approaches.

The extractive baselines include: **Centroid** [16], which scores sentences based on the centroid of documents and extracts sentences with the largest scores to compose a summary; **TextRank** [17], which builds a graph representation of sentences and sort sentences based on graph-ranking algorithm; **LexPageRank** [18], which computes sentence saliency scores based on the concept of eigenvector centrality in a graph representation of sentences; **Mul-**

⁶Command “-c 95 -2 4 -u -r 1000 -n 4 -w 1.2 -a -m -l 250” is used for ROUGE evaluation.

TABLE 5
EFFECTIVENESS OF OUR SLN ON DUC 2006

System	Pyramid	ROUGE-1	ROUGE-2	ROUGE-SU4
SRL	0.803*	0.38158 [0.37132-0.39105]	0.07398* [0.07062-0.08265]	0.13001* [0.12473-0.13865]
ESLN	0.825*	0.39487 [0.37539-0.40718]	0.08756* [0.07558-0.09952]	0.14083* [0.13781-0.14658]
Our System (w/oConcepts)	0.827*	0.38246 [0.36943-0.39462]	0.10460 [0.09698-0.11206]	0.14342 [0.13551-0.15113]
Our System	0.838	0.39017 [0.37693-0.40205]	0.11033 [0.10162-0.11865]	0.14844 [0.14070-0.15613]

TABLE 6
EFFECTIVENESS OF OUR SLN ON DUC 2007

System	Pyramid	ROUGE-1	ROUGE-2	ROUGE-SU4
SRL	0.829*	0.40869* [0.38693-0.42151]	0.09265* [0.08893-0.10025]	0.15017* [0.14693-0.16105]
ESLN	0.865*	0.42423 [0.40983-0.43005]	0.10897* [0.09737-0.12614]	0.16137* [0.15312-0.17251]
Our System (w/oConcepts)	0.871*	0.41375* [0.40174-0.42583]	0.12654* [0.11775-0.13588]	0.16415* [0.15592-0.17278]
Our System	0.885	0.42765 [0.41593-0.43826]	0.14177 [0.13212-0.15102]	0.17481 [0.16686-0.18225]

(*indicates that the improvement of our system over the baseline is statistically significant.)

tiMR [19], which uses the multi-modality manifold-ranking algorithm for extracting a topic-focused summary from multiple documents; and **Submodular**, which performs summarization by maximizing submodular functions under a budget constraint. In particular, **Submodular1** implements the algorithm proposed in [20], and **Submodular2** implements the algorithm proposed in [21]. All above extractive baselines are implemented based on PKUSUMSUM summarization toolkit [22].

The state-of-the-art abstractive baselines include: **RA-MDS** [23], a sparse-coding-based approach based on a compression-based unified optimization framework; **ILPSumm** [24], a sentence-fusion based abstractive approach by using ILP based global optimization; and **PSM** [25], an abstraction-based MDS framework that constructs new sentences by noun/verb phrase selection and merging.

5.3 Performance Comparison

5.3.1 Results with ROUGE Evaluation

To prove the overall performance of our system, we compare it with several extractive baselines and several state-of-the-art abstractive baselines, and the 95% confidence intervals are also reported. For the abstractive approach RA-MDS, we directly use the results reported in [23], and the confidence intervals were not reported. The results on the DUC 2006 dataset in Table 2 show that our system outperforms all extractive baselines and abstractive baselines on ROUGE-2 and ROUGE-SU4, while achieves competitive performance on ROUGE-1 compared with several state-of-the-art extractive baselines. The results in Table 3 show that our system outperforms all extractive baselines and abstractive baselines on the DUC 2007 dataset. In particular, our system significantly outperforms the state-of-the-art abstractive MDS approaches ILPSumm [24] and PSM [25] over five out of six metrics on the two datasets, which verifies the effectiveness of our proposed approach. The experimental results also

TABLE 7
INFLUENCE OF SEMANTIC LINKS ON DUC 2006

System	ROUGE-1	ROUGE-2	ROUGE-SU4
Our System	0.39017 [0.37693-0.40205]	0.11033 [0.10162-0.11865]	0.14844 [0.14070-0.15613]
w/o Coherence	0.38691* [0.37163-0.39905]	0.09223* [0.08693-0.10805]	0.13713* [0.12924-0.14153]
w/o Link Freq.	0.38892 [0.37138-0.39405]	0.10121* [0.09193-0.10985]	0.13884* [0.13193-0.14651]
w/o Link Div.	0.38934 [0.37629-0.39511]	0.10132* [0.09383-0.11002]	0.13982* [0.13385-0.14705]

TABLE 8
INFLUENCE OF SEMANTIC LINKS ON DUC 2007

System	ROUGE-1	ROUGE-2	ROUGE-SU4
Our System	0.42765 [0.41593-0.43826]	0.14177 [0.13212-0.15102]	0.17481 [0.16686-0.18225]
w/o Coherence	0.41358* [0.40684-0.42225]	0.12406* [0.11632-0.13545]	0.16047* [0.15293-0.17045]
w/o Link Freq.	0.42043 [0.40763-0.43025]	0.12213* [0.10893-0.13905]	0.16749* [0.15438-0.17107]
w/o Link Div.	0.42113 [0.40792-0.43411]	0.12578* [0.11069-0.14053]	0.16867 [0.15191-0.17685]

demonstrate that representing text information based on SLN of concepts and events, and summarizing text information by summarizing SLN is more effective than summarizing text directly.

5.3.2 Results with Pyramid Evaluation

Since ROUGE metric evaluates summaries by strict string matching, we also use the pyramid evaluation metric which can measure 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 in [26]. Table 4 shows the evaluation results of our system and two abstractive baselines on both DUC 2006 and DUC 2007 (set threshold value to 0.6). The results show that our system significantly (paired t-test with $p < 0.05$) outperforms the two baselines on both datasets, which demonstrates that our system can generate more informative summaries.

5.4 Discussion

5.4.1 Semantic Representation of Documents

We first conduct experiments to demonstrate that SLN of concepts and events is powerful in representing semantic information of documents, and summarizing documents at SLN layer is effective. Experiments compared our approach with three most relevant baselines: (1) **ESLN**. Our previous work represents documents by an Event Semantic Link Network [27]. The event is defined and extracted based on syntactic structures of sentences. (2) **SRL**. It represents document information based on semantic role labeling and summarizes documents information by extracting important predicate-argument semantic roles [28]. (3) Another baseline, denoted as **Our System(w/oConcepts)**, is set by removing concept nodes and semantic links between concepts from the SLN representation of documents. Then, the SLN summarization component is similar to ESLN [27] except for some semantic features (e.g., link frequency, link diversity, etc.) are used.

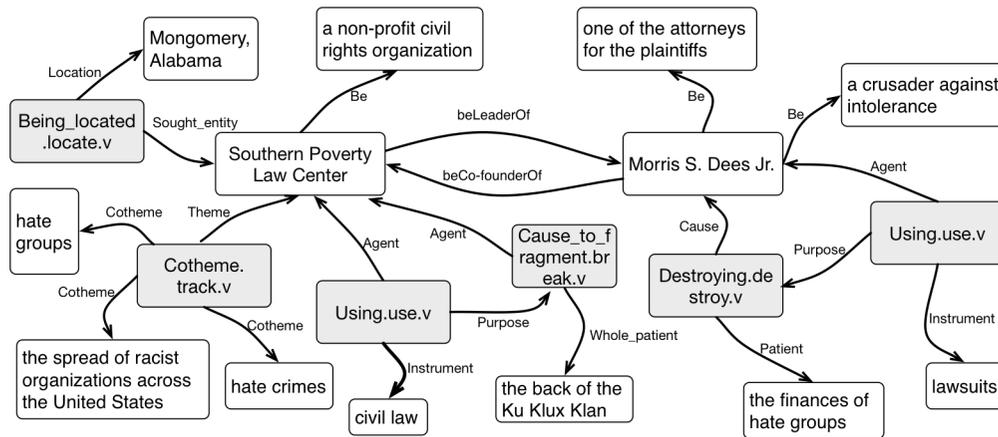


Fig. 4. An example part of the summary-SLN extracted by our method for topic “Southern Poverty Law Center”.

The results on the DUC 2006 dataset in Table 5 show that our system outperforms the three baselines on ROUGE-2 and ROUGE-SU4, while achieves comparable performance with ESLN on ROUGE-1. The results in Table 6 show that our system outperforms the three baselines on the DUC 2007 dataset. Moreover, our system significantly outperforms the three baselines on both datasets under the Pyramid metric. The results demonstrate both the power of the SLN of concepts and events in representing document information and the effectiveness of our proposed summarization model. In particular, our system consistently outperforms the baseline **Our System (w/oConcepts)** on both datasets, which demonstrate the effectiveness of adding concepts and semantic links between concepts into the SLN representation.

5.4.2 Model Validation

We further validate different components of our system by removing the coherence constraints in the ILP-based SLN summarization algorithm (i.e. *w/o Coherence*). We also do model ablation evaluation by removing the link frequency feature (i.e. *w/o Link Freq.*), or removing the link diversity feature (i.e. *w/o Link Div.*) in the Semantic Link Network summarization model to validate the influence of semantic links during the summarization process.

The results in Table 7 and 8 show that our system with coherence constraints achieves significantly better performance than the counterpart without coherence constraints on both DUC2006 and DUC2007 datasets. So the coherence constraints are helpful in selecting more salient and coherent information. The results in Table 7 and Table 8 also show that our system without either the semantic link frequency features or the semantic link diversity features will lead to consistent performance decline. The results demonstrate the important role of semantic links in representing and understanding documents.

5.4.3 Case Study

To further study why our approach based on Semantic Link Network can achieve better performance, we show both the summarized SLN and the summary generated from it to see what semantic information have been extracted from documents and how they are organized. Taking the topic “Southern Poverty Law Center” from DUC2007 dataset as an example, part of the extracted

TABLE 9
COMPARISON OF THE SUMMARY RESULTS

<p>Summary by Our System: Southern Poverty Law Center, a non-profit civil rights organization, located in Montgomery, Alabama, which was founded in the 1970s to battle racial bias, tracks hate crimes, hate groups, the spread of racist organizations across the United States and Intelligence Report covers right-wing extremists. It publicized their activities in its Intelligence Report teaching tolerance. It won major legal fights against the Ku Klux Klan and other white supremacist groups. It used civil law to break the back of the Ku Klux Klan. It won a \$7 million verdict for the mother of Michael Donald in 1987. It won a large verdict against Klan groups and members for an assault on peaceful civil rights marchers in Forsyth County in 1989. Morris S. Dees Jr., co-founder and leader of the Southern Poverty Law Center, a crusader against intolerance, one of the attorneys for the plaintiffs and the chief trial counsel of the Law Center, used lawsuits to destroy the finances of hate groups. He represented Victoria Kee-nan and her son who were attacked by two security guards near the compound in July 1998, by introducing depositions to contradict the men's testimony. He hailed the jury's finding as a powerful message against racial intolerance. He won a series of civil rights suits against the Ku Klux Klan and other racist groups in a campaign to put them out of business. He expected the judgment to bring a quick end to the Aryan Nations and its racist, anti-Semitic message. He won a \$7 million verdict against a Ku Klux Klan organization in 1987 and forced the group to turn over its headquarters building. He won \$9 million in Portland, Ore., against the White Aryan Resistance in 1990.</p>
--

Human-written Summary: The Southern Poverty Law Center is a nonprofit 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 Christian Knights of the KKK; the White Aryan Resistance; and the Aryan Nations and its founder Richard Butler. The SPLC influenced funeral homes to provide equal services to blacks and whites, tried to discourage the sale and distribution of the racist book The Turner Diaries, and protected Vietnamese fishermen from Klan intimidation. The SPLC devotes much effort to raising the funds needed to help minorities litigate against civil rights abuses. It charges its clients nothing. Nearly all money from settlements goes to the victims, with less than 2 percent going to the SPLC.

Note: Comparison of the summary generated by our system and the gold human summary for the documents of topic “Southern Poverty Law Center”.

summary-SLN is shown in Fig. 4. As demonstrated in the figure, the summarized SLN extracted by our approach actually contains meaningful and informative concept and event information units as well as the semantic relations between them. The summarized SLN represents the core content of documents, reflecting both useful syntactic and semantic information.

Table 9 shows the comparison between the summary generated by our system based on the summary SLN and the reference human summary on the topic. The results show that our summary behaves similarly with the reference summary in the following aspects: (1) Aggregating

information from different places. For example, the description of concepts “Southern Poverty Law Center” and “Morris Dees” includes information from several documents through concept and event co-reference resolution and semantic links combination. (2) Organizing sentences coherently. The coherence constraints in the ILP-based network reduction component ensure the selected event and concept information to be coherent. (3) Clear pronoun reference. The adjacent sentences with the same subject in the summary are post-edited by replacing the 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 still have syntax errors. Most cases are because of non-accurate event extraction and event relation extraction, so more effective approaches need to be developed, which will further improve the summarization performance.

5.5 The Interface

Fig. 5 shows the main interfaces of our SLN-based text summarization system. The interface shown in Fig. 5 (a) displays the source text and the corresponding SLN which help users grasp overall structure of the documents. The top-ranked events and concepts extracted from the source text are also listed under the source text column, which help users focus on the key points of documents. Fig. 5 (b) shows the reduced SLN and the generated summary. The events and concepts in the reduced SLN are also listed under the text summary column. It helps users to obtain the main information of the original documents. The system intuitively demonstrates the SLN representation of documents and the main information extracted from documents (e.g. the reduced SLN and the corresponding summary), which can be developed toward an auxiliary text reading system.

6 RELATED WORK

Various summarization approaches can be summarized in a multi-dimensional classification space [29].

Extractive approaches. They produce a summary for a document set by directly selecting salient sentences. SLN has been used to represent documents in extractive summarization approaches [12], by specializing the semantic nodes as language units such as words, sentences, paragraphs and sections. An extractive MDS approach is proposed by extracting event mentions and temporal relations between event mentions [30]. The extractive approaches have the advantages of generating fluent sentences and preserving the meaning of the original documents but inevitably face the drawbacks of information redundancy and incoherence between sentences.

Abstractive approaches. Abstractive MDS research mainly includes: (1) Sentence fusion-based approaches ([24], [31], [32]), which firstly cluster sentences into several themes and then generate a new sentence for each cluster by extracting the common information of the cluster. (2) Information extraction-based approaches ([33], [34], [35], [36]), which firstly extract important information

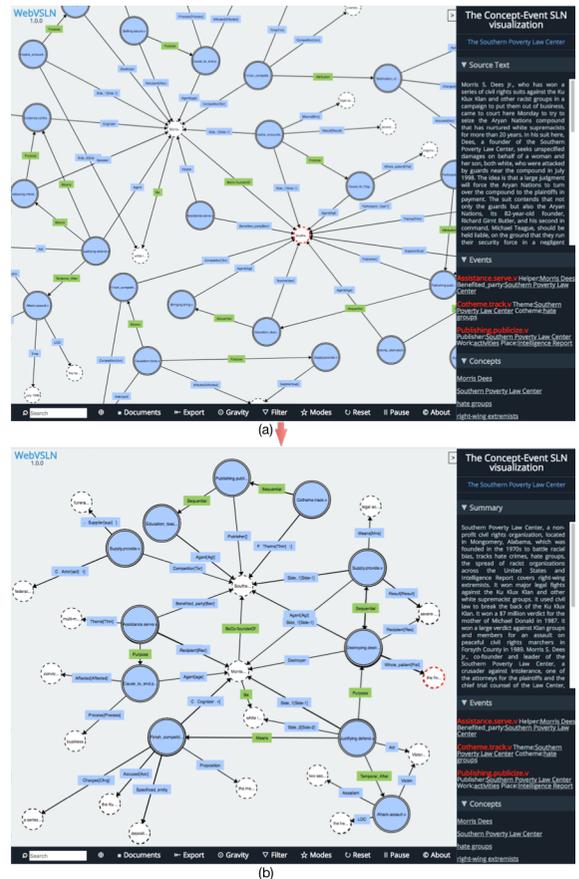


Fig. 5. System interfaces. (a) The system interface with source-SLN visualization (left-hand part) and source text (right-hand part). The top-ranked events and concepts extracted from the source text are also listed under the source text column. (b) The system interface with summary-SLN visualization (left-hand part) and text summary (right-hand part). Events and concepts in the summary-SLN are also listed under the text summary column.

units, such as Subject-Verb-Object triples, from documents and then generate new descriptions for them based on manually designed templates or patterns learned from corpus. (3) Paraphrasing based approaches ([37], [38], [39], [40]), which paraphrase the original sentences with sentence rewriting rules, such as sentence compression, phrases substitution and co-reference resolution, to generate more informative and concise summaries. These approaches mainly rely on statistical or syntactical analysis on text rather than semantics of documents.

Semantics-based Approaches. In recent years, there are a few MDS approaches attempting to use shallow semantic parsing techniques, such as semantic role labeling [41] and abstract meaning representation (AMR) [42], to improve document summarization. Semantic role information was used to enhance the graph-based ranking algorithm for extractive multi-document summarization [43]. A framework for abstractive MDS based on semantic role labeling was proposed [28]. Some recent works aim to generate abstractive summary by parsing the source text into AMR representation and then generate summary based on it [13], [44]. However, these shallow semantic

parsing techniques fail to capture the semantic relations contained in texts.

Neural-based approaches. Sequence-to-sequence models with attention mechanism have been applied to abstractive document summarization [45] [46] [47] [48]. Attempts to realize abstractive MDS under the framework have been made, e.g., generating English Wikipedia through multi-document summarization [49]. Several works apply pre-trained abstractive summarization model of single document summarization to multi-document summarization task [50][51][52] to overcome the lack of enough train examples for MDS tasks. Unsupervised neural abstractive MDS is to leverage large nonannotated corpus [53]. However, these approaches lack interpretability, controllability and tracibility.

SLN-based approaches. Research on SLN can be traced to the definition of inheritance rules for flexible retrieval of models in 1998 [64]. It was then developed into a systematic theory and model for semantics modeling in 2004 and 2012 [4]. The distinguished characteristics of SLN compared with other semantic representation approaches were introduced in [2][4]. The integration of SLN and the multi-dimensional resource space model was studied for modeling complex semantics [60][62]. SLN has been extended to support more advanced applications [3][29][61][63][65].

The summarization approach proposed in this paper is based on explicit SLN representation of document, which supports interpretability, controllability and tracibility for document summarization. The characteristics of SLN such as richness of semantic links and connectivity of SLN provide the basis for improving the quality of summary.

7 CONCLUSIONS

This paper proposes an abstractive multi-document summarization approach that transforms documents into a Semantic Link Network of concepts and events, and then transforms it into a summary by summarizing the SLN. Experiments show that the proposed text-SLN-text approach outperforms relevant state-of-the-art extractive and abstractive baselines, which verifies the effectiveness of the SLN-based summarization approach. This work further validates that the Semantic Link Network is effective in representing and understanding the semantics of documents, which is the basis of document summarization and other document-based information services.

ACKNOWLEDGMENT

This work was support in part by National Science Foundation of China (project no. 61876048).

*Professor Hai Zhuge is the corresponding author.

REFERENCES

- [1] Y. Chali and S. A. Hasan, "Query-focused multi-document summarization: Automatic data annotations and supervised learning approaches". *Natural Language Engineering*, vol.18, no.1, pp.109-145, 2012.
- [2] H. Zhuge, "Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning", *IEEE Transactions on Knowledge and Data Engineering*, vol.21, no.6, pp. 785-799, 2009.
- [3] H. Zhuge, "Interactive Semantics", *Artificial Intelligence*, vol.174, no.2, pp. 190-204, 2010.
- [4] H. Zhuge, "Semantic Link Network", Chapter 2 in *The Knowledge Grid: Toward Cyber-Physical Society*. World Scientific, 2012 (2nd ed).
- [5] M. Collins, "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms", In *Proceedings of EMNLP*, pp.1-8, 2002.
- [6] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit", In *Proceedings of ACL*, pp.55-60, 2014.
- [7] S. Schuster, R. Krishna, A. Chang, L. Fei-Fei, and C. D. Manning, "Generating semantically precise scene graphs from textual descriptions for improved image retrieval", In *Proceedings of the Fourth Workshop on Vision and Language*, pp.70-80, 2015.
- [8] C. Shen, F. Liu, F. Weng, and T. Li, "A Participant-based Approach for Event Summarization Using Twitter Streams", In *Proceedings of HLT-NAACL*, pp.1152-1162, 2013.
- [9] S. Liu, Y. Chen, S. He, K. Liu, and J. Zhao, "Leveraging framenet to improve automatic event detection", In *Proceedings of ACL* 2134-2143, 2016.
- [10] W. Li, D. Cheng, L. He, et al. "Joint Event Extraction Based on Hierarchical Event Schemas from FrameNet". *IEEE Access*, vol. 7, pp.25001-25015, 2019.
- [11] C. Chen and V. Ng. Joint modeling for Chinese event extraction with rich linguistic features. In *Proceedings of COLING*, pp. 529-544, 2012.
- [12] X. Sun and H. Zhuge, "Summarization of Scientific Paper through Reinforcement Ranking on Semantic Link Network", *IEEE Access*, vol. 6, pp.40611-40625, 2018.
- [13] F. Liu, J. Flanigan, et al. "Toward Abstractive Summarization Using Semantic Representations". In *HLT-NAACL* 2015.
- [14] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries", *ACL Text Summarization Branches Out*, 2004, available at <https://www.aclweb.org/anthology/W04-1013>.
- [15] A. Nenkova and R. Passonneau, "Evaluating content selection in summarization: The pyramid method", In *Proceedings of HLT-NAACL*, pp.145-152, 2004.
- [16] D. R. Radev, H. Jing, M. Styś and D. Tam, "Centroid-based summarization of multiple documents". *Information Processing & Management*, vol.40, no.6, pp.919-938, 2004.
- [17] R. Mihalcea and P. Tarau, "TextRank: Bringing order into text", In *Proceedings of EMNLP*, pp.404-411, 2004.
- [18] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization", *Journal of Artificial Intelligence Research*, vol. 22, pp.457-479, 2004.
- [19] X. Wan and J. Xiao, "Graph-Based Multi-Modality Learning for Topic-Focused Multi-Document Summarization", In *Proceedings of IJCAI*, pp.1586-1591, 2009.
- [20] H. Lin and J. Bilmes, "Multi-document summarization via budgeted maximization of submodular functions", In *HLT-HAACL*, pp.912-920, 2010.
- [21] J. Li, L. Li and T. Li. "Multi-document summarization via submodularity". *Applied Intelligence*, vol. 37, no.3, pp.420-430, 2012.
- [22] J. Zhang, T. Wang and X. Wan. "PKUSUMSUM: a Java platform for multilingual document summarization". In *Proceedings of COLING*, pp.287-291, 2016.
- [23] P. Li, L. Bing, W. Lam, H. Li, and Y. Liao, "Reader-aware multi-document summarization via sparse coding". In *Proceedings of IJCAI*, pp.1270-1276, 2015.
- [24] S. Banerjee, P. Mitra, and K. Sugiyama, "Multi-document abstractive summarization using ilp based multi-sentence compression". In *Proceedings of IJCAI*, pp.1208-1214, 2015.
- [25] L. Bing, P. Li, Y. Liao, W. Lam, W. Guo, and R. J. Passonneau, "Abstractive Multi-Document Summarization via Phrase Selection and Merging", In *Proceedings of ACL*, pp.1587-1597, 2015
- [26] R. J. Passonneau, E. Chen, W. Guo, and D. Perin, "Automated Pyramid Scoring of Summaries using Distributional Semantics", In *Proceedings of ACL*, pp.143-147, 2013.
- [27] W. Li, L. He, and H. Zhuge, "Abstractive News Summarization based on Event Semantic Link Network". In *Proceedings of COLING*, pp.236-246, 2016.
- [28] A. Khan, N. Salim, and Y. J. Kumar, "A framework for multi-document abstractive summarization based on semantic role labeling", *Applied Soft Computing*, vol.30, pp.737-747, 2015.

- [29] H. Zhuge, "Multi-Dimensional Summarization in Cyber-Physical Society", Morgan Kaufmann, 2016.
- [30] G. Glavaš, J. and Šnajder, "Event graphs for information retrieval and multi-document summarization", *Expert Systems With Applications*, vol.41, no.15, pp.6904-6916, 2014.
- [31] R. Barzilay, and K. R. McKeown, "Sentence fusion for multidocument news summarization", *Computational Linguistics*, vol. 31, no.3, pp.297-328, 2005.
- [32] K. Filippova, "Multi-sentence compression: Finding shortest paths in word graphs". In *Proceedings of COLING 2010*, pp.322-330, 2010.
- [33] P. E. Genest, and G. Lapalme, "Framework for abstractive summarization using text-to-text generation", In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pp.64-73, 2011.
- [34] L. Wang and C. Cardie, "Domain-Independent Abstract Generation for Focused Meeting Summarization", In *Proceedings of ACL*, pp.1395-1405, 2013.
- [35] D. Pighin, M. Cornolti, E. Alfonseca, and K. Filippova, "Modelling Events through Memory-based, Open-IE Patterns for Abstractive Summarization", In *Proceedings of ACL*, pp.892-901, 2014.
- [36] W. Li, "Abstractive Multi-Document Summarization with Semantic Information Extraction". In *Proceedings of EMNLP*, pp.1908-1913, 2015.
- [37] T. Cohn and M. Lapata, "Sentence compression as tree transduction", *Journal of Artificial Intelligence Research*, vol. 34, pp.637-674, 2009.
- [38] T. Berg-Kirkpatrick, D. Gillick, and D. Klein, "Jointly learning to extract and compress". In *Proceedings of HLT-NAACL*, pp.481-490, 2011.
- [39] T. Hirao, M. Nishino, Y. Yoshida, J. Suzuki, N. Yasuda, and M. Nagata, "Summarizing a document by trimming the discourse tree", *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol 23, no.11, pp.2081-2092, 2015.
- [40] L. Wang, H. Raghavan, V. Castelli, R. Florian, and C. Cardie, "A sentence compression-based framework to query-focused multi-document summarization", arXiv preprint arXiv:1606.07548, 2016.
- [41] X. Carreras, and L. Márquez, "Introduction to the CoNLL-2005 shared task: Semantic role labeling", In *CoNLL*, pp.152-164, 2005.
- [42] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, and N. Schneider, "Abstract meaning representation for sembanking", In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 178-186, 2013.
- [43] S. Yan, and X. Wan, "SRRank: leveraging semantic roles for extractive multi-document summarization", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol.22, no.12, pp.2048-2058, 2014.
- [44] K. Liao, L. Lebanoff, and F. Liu. "Abstract Meaning Representation for Multi-Document Summarization". arXiv preprint arXiv:1806.05655, 2018.
- [45] A. See, P. J. Liu, and C. D. Manning. "Get to the point: Summarization with pointer-generator networks". In *Proceedings of ACL 2017*, <https://arxiv.org/abs/1704.04368>.
- [46] W. Li, X. Xiao, Y. Lyu, and Y. Wang. "Improving Neural Abstractive Document Summarization with Explicit Information Selection Modeling". In *Proceedings of EMNLP*, pp.1787-1796, 2018.
- [47] W. Li, X. Xiao, Y. Lyu and Y. Wang. "Improving Neural Abstractive Document Summarization with Structural Regularization". In *Proceedings of EMNLP 2018*, pp.4078-4087, 2018.
- [48] Y. C. Chen and M. Bansal. "Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting". In *Proceedings of ACL*, pp.675-686, 2018.
- [49] P. J. Liu, M. Saleh, E. Pot, B. Goodrich, R. Sepassi, L. Kaiser and N. Shazeer. "Generating wikipedia by summarizing long sequences". In *Proceedings of ICLR 2018*. <https://arxiv.org/abs/1801.10198>.
- [50] L. Lebanoff, K. Song and F. Liu. "Adapting the Neural Encoder-Decoder Framework from Single to Multi-Document Summarization". In *Proceedings of EMNLP*, pp. 4131-4141, 2018.
- [51] J. Zhang, J. Tan and X. Wan. "Towards a Neural Network Approach to Abstractive Multi-Document Summarization". arXiv preprint arXiv:1804.09010, 2018.
- [52] T. Baumel, M. Eyal and M. Elhadad. "Query Focused Abstractive Summarization: Incorporating Query Relevance, Multi-Document Coverage, and Summary Length Constraints into seq2seq Models". arXiv preprint arXiv:1801.07704, 2018.
- [53] E. Chu and J. P. Liu. "Unsupervised Neural Multi-Document Abstractive Summarization". arXiv preprint arXiv:1810.05739, 2018.
- [54] Q. Li, H. Ji and L. Huang. "Joint event extraction via structured prediction with global features. In *Proceedings of ACL*, pp.73-82, 2013.
- [55] T. H. Nguyen, K. Cho, K. and R. Grishman. Joint event extraction via recurrent neural networks. In *Proceedings of NAACL*, pp.300-309, 2016.
- [56] M. T. Pilehvar, D. Jurgens, and R. Navigli, "Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity". In *Proceedings of ACL*, pp.1341-1351, 2013.
- [57] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *Journal of Machine Learning Research*, vol. 12 (Jul), pp.2121-2159, 2011.
- [58] J. Pennington, R. Socher, C. Manning. Glove: Global vectors for word representation. In *Proceedings of EMNLP*, pp.1532-1543, 2014.
- [59] V. W. Feng and G. Hirst. Text-level discourse parsing with rich linguistic features. In *Proceedings of ACL*, pp.60-68, 2012.
- [60] H. Zhuge, *The Web Resource Space Model*, Springer, 2008.
- [61] H. Zhuge, Semantic linking through spaces for cyber-physical-social intelligence: A methodology, *Artificial Intelligence*, vol.175, No. 5-5, pp.988-1019, 2011.
- [62] H. Zhuge and Y. Xing, Probabilistic Resource Space Model for Managing Resources in Cyber-Physical Society, *IEEE Transactions on Service Computing*, vol. 5, no.3, pp.404-421, 2012.
- [63] H. Zhuge, *Cyber-Physical-Social Intelligence on Human-Machine-Nature Symbiosis*, Springer, 2019.
- [64] H. Zhuge, "Inheritance rules for flexible model retrieval," *Decision Support Systems*, vol. 22, no. 4, pp. 379-390, 1998.
- [65] H. Zhuge, "Socio-Natural Thought Semantic Link Network: A Method of Semantic Networking in the Cyber Physical Society". Keynote, in *Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, pp.19-26, 2010.
- [66] H. Zhuge, "Active e-document framework ADF: model and tool". *Information & Management*, vol.41, no.1, pp.87-97, 2003.



Wei Li is a PhD candidate of the Institute of Computing Technology in Chinese Academy of Sciences and the School of Computer and Control Engineering in University of Chinese Academy of Sciences, China. His current research interests include Text Summarization, Semantic Link Network and Natural Language Processing.



Hai Zhuge is a Distinguished Scientist of the ACM, a Fellow of British Computer Society, and a Senior Member of IEEE. He has made systematic contribution to semantics and knowledge modeling through lasting fundamental research on the [Semantic Link Network](#) and the [Resource Space Model](#) based on a multi-dimensional methodology. He has also made fundamental contribution to the study of Cyber-Physical-Social Intelligence. As a Chair in Computer Science in Aston University, he leads the International Research Network on Cyber-Physical-Social Intelligence consisting of Guangzhou University, Key Lab of Intelligent Information Processing at the Institute of Computing Technology in Chinese Academy of Sciences, Systems Analytics Research Institute at Aston University, and University of Chinese Academy of Sciences. He is serving as an Associate Editor of *IEEE Intelligent Systems*.
Personal webpage: <http://www.knowledgegrid.net/~h.zhuge>.

TABLE A.1
THE SET OF CONCEPT DESCRIPTION AND SENTENCE STRUCTURING RULES

CONCEPT DESCRIPTION RULES	
For concept $a \left(\overset{r_1}{\rightarrow} c_1, \overset{r_2}{\rightarrow} c_2, \dots, \overset{r_n}{\rightarrow} c_n \right)$, the description of concept a can be:	
1)	Appositive modifier “ a, c_1, \dots, c_n ”, for example “ <i>Morris Dees, civil rights lawyer, co-founder of Southern Poverty Law Center, ...</i> ”
2)	Attributive clause “ a who/which/that $r_1 c_1, \dots, \text{and } r_n c_n$ ”, for example “ <i>Morris Dees who was the co-founder of Southern Poverty Law Center and a civil rights lawyer...</i> ”
3)	Appositive modifier mixed with attributive clause “ c_i a who was $c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_n$ ”, for example “ <i>Civil rights lawyer Morris Dees who was the co-founder of Southern Poverty Law Center ...</i> ”
SENTENCE STRUCTURING RULES	
if $e_1 \xrightarrow{\text{AFTER/BEFORE/OVERLAP}} e_2$, then generate “ e_1 after/before/when e_2 ”;	if $e_1 \xrightarrow{\text{ce}} e_2$, then generate “Because e_1, e_2 ” and “ e_2 because e_1 ”;
if $e_1 \xrightarrow{\text{pur}} e_2$, then generate “ e_1 to e_2 ”, “ e_1 in order to e_2 ” and “ e_1 so that e_2 ”;	if $e_1 \xrightarrow{\text{mea}} e_2$, then generate “ e_1 by e_2 ” and “ e_1 by the way that e_2 ”;
if $e_1 \xrightarrow{\text{attri}} e_2$, then generate “ $e_1 e_2$ ”, “ e_1 about/on/in/with/at e_2 ” and “ e_1 that e_2 ”;	if $e_1 \xrightarrow{\text{seq}} e_2$, then generate “ e_1, e_2 ” and “ e_1 and e_2 ”;

APPENDIX

A SUMMARY GENERATION

The summary of *SLN* contains selected concepts, events and semantic links between them. To generate fluent sentences based on the summary-*SLN*, a set of concept description templates, event description templates and sentence structuring rules are extracted from the annotated corpus. To improve the linguistic quality of summary, we firstly generate large numbers of candidate sentences based on the extracted templates, and then use a greedy algorithm to select sentences with the best linguistic quality and no information overlap to compose a summary.

A.1 Sentence Generation

To generate a complete and correct sentence, both the description of concepts, events and the organization of sentence structures are settled in the following ways:

- For each concept that is semantically linked to other concepts, different expressions can be generated by using the concept description rules in Table A.1.
- For each event node, different sentences are generated based on a set of event description templates extracted from the annotated corpus. Example templates for “ATTACK” event are shown in Table A.2.
- If two events have semantic links, they can be merged to generate one unified sentence by using the sentence structuring rules in Table A.1. If two events share the same or coreferential argument concepts, pronouns will be used to avoid repeated description of the same concept.

A.2 Greedy Selection

After the above sentence generation step, large number of candidate summary sentences can be obtained. Some of them may have information overlap with each other if they are generated from the same event or concept node. To improve the linguistic quality of the summary, we iteratively select a sentence with the best linguistic quality and delete sentences that have information overlap with it from the candidate set. The linguistic quality of a sentence $s = \{w_1, w_2, \dots, w_L\}$ is defined similarly as [24]:

$$LQ(s) = 1 / \left(1 - \left(\log_2 \prod_{t=1}^W P(w_t | w_{t-1}, w_{t-2}) \right) / W \right) \quad (19)$$

where W is the total number of words in sentence s ; w_0 and w_1 both represent the beginning of sentence s . $P(w_1 | w_0, w_{-1})$ indicates the probability of w_1 as the first word of a sentence. The 3-gram language model $P(w_t | w_{t-1}, w_{t-2})$ is learnt from the English Gigaword corpus.

The network-flow based coherence constraints ensure that the summary-*SLN* is connected and has a flow from the root node to all other nodes. The selected summary sentences are ordered based on the direction of the flow to form a coherent summary.

B EVALUATION OF SEMANTIC LINK NETWORK CONSTRUCTION

B.1 Dataset

Both the annotated exemplars and full text annotations in FrameNet 1.7 corpus are transformed into annotated datasets for event extraction by filtering non-event frames. Gold relations between events in full text annotations are manually annotated by using brat (<http://brat.nlplab.org>). All exemplars are used as training data for event extraction model. The full text annotations are split into training set and test set. The training set is used for training the event extraction model and event relation prediction model. The models are evaluated on the test set. The details of the datasets are shown in Table B.1. We also evaluate the performance of event extraction on the ACE 2005 corpus. To compare with state-of-the-art event extraction systems on the ACE 2005 corpus, we use the same evaluation settings in previous work [54, 55] and use 40 news-wire documents of ACE as our test set.

B.2 Evaluation of FrameNet-based Event Extraction

We compare the event schema definition in our work to both predefined ACE and ERE^[1] (Entities, Relations and Events) event schemas. ERE was designed as a lighter-weight version of ACE and a simple approach to entity, relation and event annotation. As shown in Table B.2, our event schema definition contains 655 event types with

TABLE B.1
THE ANNOTATED DATASET FOR EVENT EXTRACTION

	Count	
Exemplar sentences	86962	
Full text annotations	Training set	Test set
Sentences	3402	717
Events	8079	1755
Event Relations	2437	1003

The details of exemplars and full text annotations after filtering non-event frames from FrameNet 1.7 Release.

2050 argument roles (604 distinct roles) in total, which is a few orders of magnitude larger than the definitions in ACE and ERE. More definition of event types will make our event extraction system extract richer event information from text, which facilitates the text summarization process.

The performance of event extraction results of our system is compared with the rule-based system in our previous work ESLN [27]. We mainly evaluate the performance of event trigger identification and event argument identification. An event trigger is correctly identified if its offsets match those of a gold-standard trigger; an event argument is correctly identified if its offsets and event trigger match those of any of the reference argument mentions in the document. Table B.3 shows that the results of our system obtain not only much better recall but also significantly higher precision on both event trigger identification and event argument identification. The results demonstrate that our event extraction system is able to extract much richer event information from text. Moreover, our system can recognize the event types and argument roles correctly, which helps both the semantic representation of documents and the document summarization process.

We further compare the coverage of the event extraction results of our system with several ACE-based state-of-the-art systems on the ACE 2005 test set: (1) **Structure-Joint**[Ⓜ]: A structured perception joint model based on symbolic global semantic features [54]. (2) **JRNN**[Ⓜ]: A joint event extraction model via recurrent neural network [55]. The results of Structure-Joint and JRNN are computed by publicly released models. As our system and the ACE-based models use different event schema definition, we cannot compare the accuracy of event type classification and argument role classification directly. So, we mainly evaluate the performance of event trigger identification and event argument identification. The results in Table B.4 show that the recall of our system on the ACE test set is significantly higher than the ACE-based models. The main reason is that our event schemas have much wider coverage than ACE and our annotated dataset is much larger than ACE annotations.

B.3 Evaluation of Event Relation Extraction

We compare the event relation extraction component in this work with our previous work ESLN [27]. Besides the

[Ⓜ] <https://github.com/Aureliu/BIU-RPI-Event-Extraction-Project>

[Ⓜ] <https://github.com/bishanyang/EventEntityExtractor>

TABLE B.2
COMPARING THE COVERAGE OF EVENT SCHEMA DEFINITION WITH ACE AND ERE

Data	Our Event Schemas	ACE	ERE
#of Event Types	655	33	38
#of Argument Roles	2050	159	185
#of Distinct Arg. Roles	604	28	26

syntactic and statistic features used in the ESLN, more semantic features such as event types and relations of two event types in FrameNet are used to help predict the semantic relations between events in this work. The results in Table B.5 show that our system significantly improves the performance of event relation prediction.

B.4 Evaluation of Concept Relation Identification

The concept extraction process is the same as our previous work ESLN, which is able to cover 98.4% of the gold-standard argument mentions in the test set. The precision, recall, and F1 score of concept extraction on the test set are 83.29%, 79.28% and 81.23%, respectively. The relations between concepts are restricted with syntactic patterns “be”, “be-NP-prep” and “be-AP-prep” where NP indicates noun phrase and AP indicates adjective phrase. Several dependency syntactic rules (e.g. dependency labels “nsubj”, “appos”, “nmod:of” and “nmod:poss”) are used to extract concept relations from text. Representative examples of dependency patterns for concept relation extraction are shown in Figure B.1. We manually analyze 100 random sentences, and the results show that the three patterns cover over 90% relations except for the relations represented with verbs that can be recognized as events. The precision, recall, and F1 score of concept relation extraction on the test set are 75.12%, 69.63% and 72.27%, respectively. The errors during concept extraction and concept relation extraction mainly come from non-accurate dependency parsing results. The error of concept extraction is mainly boundary error, for example “a non-profit organization in Birmingham, Ala, tracks”. The error of concept extraction will also result in the error of concept relation extraction.

C FEATURE WEIGHTS LEARNING FOR SEMANTIC LINK NETWORK SUMMARIZATION

We learn feature weights θ and ψ by training on a set of source *SLNs* paired with gold summary *SLNs*. The source *SLN* is constructed from source texts whereas the gold summary *SLN* is constructed from reference summaries and then mapped into the source *SLN* by text similarity method [56]. We formulate our estimation problem as follows:

$$-score(G^*) + \max_c(score(G) + cost(G; G^*)) \quad (18)$$

TABLE B.3
COMPARING THE PERFORMANCE OF EVENT EXTRACTION ON THE TEST SET

Data	Our Event Extraction System			Rule-based System in ESLN [27]		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Event Trigger Identification	92.51	77.38	84.27	87.96	68.56	77.06
Event Type Classification	82.96	69.39	75.57	N/A	N/A	N/A
Event Argument Identification	80.99	58.43	67.88	67.27	44.42	53.51
Argument Role Classification	72.72	52.46	60.95	N/A	N/A	N/A

Note: the rule-based system doesn't recognize event types and argument roles in the FrameNet-based event schemas.

TABLE B.4
COMPARING THE COVERAGE OF EVENT EXTRACTION RESULTS ON THE ACE2005 TEST SET

Data	Our System	Structure-Joint [54]	JRNN [55]
	Recall	Recall	Recall
Event Trigger Identification	85.38	79.48	78.30
Event Argument Identification	47.94	32.30	23.92

Note: as ACE has only annotated part of events in text, so only recall is compared on the ACE test set.

where G^* denotes the gold summary- SLN . $score()$ is defined in Equation 7. $cost(G; G^*)$ penalizes each event or concept in G but not in G^* , which can be easily incorporated into the linear objective in Equation 8. We optimize our goal using AdaGrad [57] with l_2 regularization ($\lambda = 0.01$), with an initial step size 0.1. The ILP model is solved using Gurobi 6.5.2.

TABLE B.5
EVALUATION OF EVENT RELATION EXTRACTION ON THE TEST SET

	Precision	Recall	F1-score
Our System	0.6913	0.6732	0.6821
ESLN [27]	0.5894	0.6222	0.6054

D FEATURES

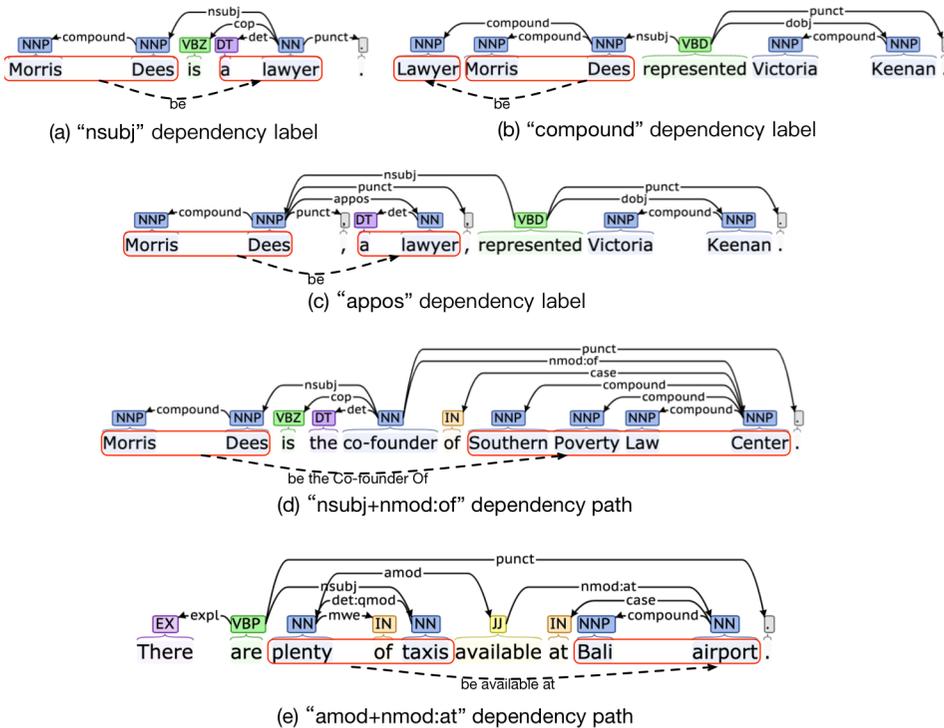


Figure B.1 Representative examples of dependency patterns for concept relation extraction.

The features for event type classification, event argument extraction and event relation prediction are shown in Table D.1, D.2 and D.3 respectively. The features for SLN summarization are shown in Table D.4.

TABLE D.1
THE FEATURES FOR EVENT TYPE IDENTIFICATION

Lexical features	Word and lemma of the trigger word, its parent and children in dependency tree
POS-tag features	part-of-speech tags of the trigger word, its parent and children in dependency tree
Syntactic features	the set of syntactic dependencies of the trigger word; the set of dependency labels of its children; the dependency label connecting it and its parent
Word vector features	100-dimensional GloVe word vector [58] of the trigger word

All incorporate with the event type being scored.

TABLE D.2
THE FEATURES FOR EVENT ARGUMENTS EXTRACTION

Ordering features	the relative position of concept span s with respect to trigger t : before, after or overlap; the distance between nearest word of s and t ;
Lexical features	word, lemma, and POS-tag of the first, last and head word of s and trigger word t ; the number of words in s ;
Syntactic features	The syntactic dependency label between head word of s and trigger t ; the syntactic dependency of the first word of s with respect to its head; all syntactic dependents and labels of the head word of s and trigger t ; sequence of labeled, directed edges from the head word of s to head word of t and its length.
Word vector features	100-dimensional GloVe word vector [58] of head word of s and t ; the summation and average of word vectors of all words in s ; the summation and average of word vectors of all words in t .

All incorporate the event type f and argument role r being scored.

TABLE D.3
THE FEATURES FOR EVENT RELATIONS CLASSIFICATION

Event type features	Event types of two events; Relations of two event types in FrameNet; whether two event types belong to the same event scenario; the common ancestor of two event types in hierarchical event schemas.
Event arg. features	Whether two events share the same concept as argument; whether the argument concepts of two events has semantic links with each other.
Position features	the set of features that measure the distance between event triggers (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 triggers as well as features indicating whether the word forms are the same, the semantic similarity between trigger words, the word and lemma of each token between the trigger words;
Syntactic features	syntactic path between the triggers (dependency labels on the syntactic path between the triggers), 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 triggers
Modifier features	the set of features that describe the modal, auxiliary, negation, and determination modifiers of both event triggers
Word vector features	100-dimensional GloVe word vector [58] of both event triggers words
Discourse features	the discourse relations between event mentions. We use the text-level discourse analysis method [59] to extract the discourse relations between events.

TABLE D.4
CONCEPT AND EVENT FEATURES FOR SUMMARIZING SLN

Concept Features	Concept Type	whether it is NE, proper nouns, contains prepositional phrase, or occurs in the topic description
	Concept Frequency	one binary feature for each frequency threshold ($t=0/1/2/5/10$) of the concept occurring in text
	Concept Head	word, lemma, POS-tag and frequency of head word in text (binarized using 5 thresholds)
	Concept Position	average and foremost position of sentences containing the concept (binarized using 5 thresholds)
	Concept Span	word, lemma, POS-tag of the concept words
	Link Frequency	total number of links from and to the concept node in <i>SLN</i> (binarized using 5 thresholds)
Event Features	Link Diversity	total types of links from and to the concept node in <i>SLN</i> (binarized using 5 thresholds)
	Trigger Word	word, lemma, POS-tag of the trigger word and whether it occurs in the topic description
	Trigger Frequency	one binary feature for each frequency threshold ($t=0/1/2/5/10$) of the trigger word occurring in text
	Trigger Position	average and foremost position of sentences containing the trigger (binarized using 5 thresholds)
	Event Frequency	one binary feature for each frequency threshold ($t=0/1/2/5/10$) of the event mention occurring in text
	Event Position	average and foremost position of sentences containing the event mention (binarized using 5 thresholds)
	Time&Loc Arg.	one binary feature indicates whether it contains time argument and location argument
Argument Roles	event type; ratio of argument roles that are filled (binarized using 5 thresholds)	
Link Frequency	total number of links from and to the event node in <i>SLN</i> (binarized using 5 thresholds)	
Link Diversity	total types of links from and to the event node in <i>SLN</i> (binarized using 5 thresholds)	