Identifying the Most Dominant Event in a News Article by Mining Event Coreference Relations

Prafulla Kumar Choubey, Kaushik Raju, Ruihong Huang
Department of Computer Science and Engineering
Texas A&M University
(prafulla.choubey, kaushik.raju, huangrh)@tamu.edu

Abstract

Identifying the most dominant and central event of a document, which governs and connects other foreground and background events in the document, is useful for many applications, such as text summarization, storyline generation and text segmentation. We observed that the central event of a document usually has many coreferential event mentions that are scattered throughout the document for enabling a smooth transition of subtopics. Our empirical experiments, using gold event coreference relations, have shown that the central event of a document can be well identified by mining properties of event coreference chains. But the performance drops when switching to system predicted event coreference relations. In addition, we found that the central event can be more accurately identified by further considering the number of sub-events as well as the realis status of an event.

1 Introduction

According to the grounding principles (Grimes, 1975), a document consists of foreground events that form the skeleton of the story and move the story forward, and background events that add supportive information. Studies have shown that a foreground event tends to be the most important event in a sentence, which is usually the event that appears in the main clause, is active voiced, and has a high transitivity[1] (Decker, 1985). But among multiple foreground events, which one is most central to the overall story? We propose a new task of detecting the most dominant event in a news article, which is an event assumed to govern and connect other foreground events and background events. In other words, removal of the central event can break the entirety of a document and decompose the document into sections describing disjoint sets of situations. Identifying the central event of a document is clearly important for a wide range of NLP applications, including text summarization, storyline generation and text segmentation.

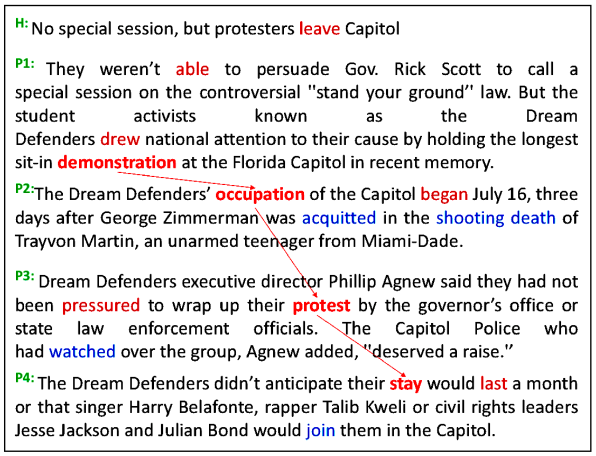

Figure 1: An example document to illustrate the central event of a document. Red colored words are foreground events, blue colored words are background events and mentions of the central event are in bold.

The intuitive observation is that the central event of a document usually has a large number of coreferential event mentions and those coreferential mentions are spread throughout the document. In Figure 1, the paragraphs 1-4 each describe a relatively independent subtopic and the repeated mentions of the central event "demonstration" throughout the document enable a smooth flow of information. For the same reason, identifying the central event facilitates partitioning text into coherent segments. But note that, the central event may not be the most newsworthy event that serves as the trigger for writing an article, and thus may not appear in the title or in the first sentence of a new article. As illustrated in this example, the trigger event is "protesters leave capitol", while

[1] High transitivity events have certain properties, are volitional, affirmative, realis etc.

Proceedings of NAACL-HLT 2018, pages 340–345
New Orleans, Louisiana, June 1 - 6, 2018. ©2018 Association for Computational Linguistics

the central event is "demonstration", the event that effectively connects other foreground events and background events and makes the story an entirety.

To systematically verify these observations, we annotated central events in news articles taken from two publicly available datasets, the richer event description (RED) (O'Gorman et al., 2016) and KBP 2015 (Mitamura et al., 2015) corpora. While whether each news article has only one central event is arguable, our two annotators agreed on the same central event in 97 out of 104 (93%) documents that we annotated. We then designed several rule-based methods to identify the central event by exploiting human annotated event coreference relations. Experimental results showed that indeed in around 75% of the documents in both corpora, the central event either has the largest number of coreferential event mentions or has the largest stretch size (i.e., the number of sentences between the first mention and the last mention of the central event) in the discourse. In addition, we found that the central event can be more accurately identified by further considering the number of sub-events as well as the realis status of an event, which indicate if an event is an actual specific event or a generic event etc. The evaluation shows that the insightful rules outperform several strong baseline approaches, including several heuristic based methods and random walk based event ranking methods, as well as two regression classifiers that integrate these rules as features.

2 Related Work

Many previous works studied the parameters that determine the overall quality of an individual event, including actualization (Tasaku, 1981), transitivity (Hopper and Thompson, 1980; Tsunoda, 1985) and the broader concept of eventiveness (Monahan and Brunson, 2014). However, these atomic qualities defined for an individual event are inadequate in distinguishing the key foreground event in a document.

In concurrent works, Decker (1985); Kay and Aylett (1996) focused on distinguishing foreground events from background events in a sentence and proposed that the most important event within a sentence is usually the event that appears in the main clause, is active voiced, and has a high transitivity. Upadhyay et al. (2016) applied these rules to identifying the trigger event of a news article by identifying the most important event in a human-generated document summary.

Recognizing document-level central events has been shown important for text summarization. Filatova and Hatzivassiloglou (2004a,b) used normalized frequencies of co-referential event mentions as parameters to prioritize events to be included in a summary and found that this helped in generating better text summaries, despite its being an elementary measure. Our experiments showed that in addition to the number of co-referential event mentions, discourse layout features including both the stretch of an event chain and early presences of event mentions are key factors in identifying the central event of a document.

Graph-based methods (Mihalcea and Tarau, 2004) have been widely used to identify keywords and phrases in a document by constructing a word/phrase graph and applying random walk algorithms (Brin and Page, 2012) on the graph. We implemented random walk based methods for identifying the central event as well, which however did not perform well. Mainly, the random walk based ranking strategy determines the importance of an events based on the importance of its related events in a document graph, which does not effectively capture discourse layout features of co-referential event mentions, which are important for identifying the central event of a document.

3 Central Event Annotations

We annotated central events for 30 news articles from the RED corpus[2] and 74 news articles from the KBP 2015 corpus[3]. We asked two annotators to identify the most dominant event that connects other foreground and background events. Both the documents and the gold event mentions for each document inherited from the previous RED and KBP annotations were provided to annotators. The annotators were instructed to select only one event as the central event. For 26 documents from the RED corpus and 71 documents from the KBP

[2]The RED corpus contains 95 documents in total. However, 65 documents are news summaries, discussion forum posts or web posts. The central event as defined should only be considered for natural coherent texts, therefore, the annotations were only conducted for the 30 news articles in the corpus.

[3]The KBP 2015 corpus contains 158 documents, where 81 are news articles and the remaining are discussion forum posts. Then in 7 out of the 81 news articles, annotators unanimously found that the central event was not of one of the interested event types in KBP and was not tagged in the KBP annotations. Therefore, our annotators skipped the 7 documents.

corpus, both annotators identified the same central event. For the other 7 documents, where the two annotators disagreed on the central event, we kept the annotations from the first annotator.

4 Characteristics of Central Events

We analyzed the distributional properties of central events in the first 10 documents from the RED corpus. The findings are summarized below.

Frequent and Extended Repetitions: As shown in Figure 1, the central event is usually repeated throughout the document. This observation can also be accounted to the way humans produce and comprehend language. Language is inherently sequential and a writer repeats the same event to remind the readers about the main event. Therefore, the frequent and extended repetitions of the central event facilitate to minimize the cognitive effort needed by the reader for understanding a text.

Early Presences: News articles mostly begin with a summary of important events and continue to elaborate them in subsequent paragraphs. To some extent, the objective of initial paragraphs is to direct readers' attention toward the main subject. Therefore, while the central event may not always appear in the title or in the first sentence of a new article, the central event often appears early in the beginning paragraphs.

Sub-events: Being the most dominant event in a document, the central event often has many sub-events that are present to elaborate and support the central event.

Event Realis Status: Central events are usually specific and have actually occurred. This event attribute has been defined as the contextual modality in RED corpus[4] and realis status in KBP corpus[5] and we observed that this attribute is "Actual" for the majority of central events.

5 Central Event Identification

Inspired by the identified characteristics of central events, we designed rule-based classifiers that rely on the following four ranking critera.

Size Rank: calculated using the number of coreferential event mentions in a event coreference chain. The event having the largest number of coreferential mentions is ranked the highest.

[4]defines 4 types of contextual modality, namely, actual, hypothetical, uncertain/ hedged and generic

[5]defines 3 realis status types, namely, actual, generic and other

Stretch Rank: based on the number of sentences between the first and the last mention of an event. The event with the largest stretch size is ranked the highest.

Position Rank: based on the sentence number in which an event was first mentioned. This measure is to capture the characteristic that central events tend to appear early in a document.

Enriched Size Rank: This rank is based on the sum of the number of coreferential mentions for an event and the number of its sub-events.

Input: central event candidates, E_Z, E_T, E_P, E_E, E_R
Output: E_{center}
$E_{center} := \phi$
Coreference
$E_{center} := E_Z \cap E_T \cap E_P$
if $E_{center} == \phi$: $E_{center} := (E_Z \cup E_T) \cap E_P$
if $E_{center} == \phi$: $E_{center} := E_P$
return E_{center}
Coreference + Subevent
$E_{center} := E_Z \cap E_T \cap E_P \cap E_E$
if $E_{center} == \phi$: $E_{center} := (E_Z \cup E_T) \cap E_P \cap E_E$
if $E_{center} == \phi$: $E_{center} := E_P \cap E_E$
if $E_{center} == \phi$: $E_{center} := E_P$
return E_{center}
Coreference + Subevent + Realis
$E_{center} := E_Z \cap E_T \cap E_P \cap E_E \cap E_R$
if $E_{center} == \phi$: $E_{center} := (E_Z \cup E_T) \cap E_P \cap E_E \cap E_R$
if $E_{center} == \phi$: $E_{center} := E_P \cap E_E \cap E_R$
if $E_{center} == \phi$: $E_{center} := E_P \cap E_E$
if $E_{center} == \phi$: $E_{center} := E_P$
return E_{center}

5.1 Rule Based Classifiers

First, we identify central event candidates by requiring their size rank in the top three positions. Note that more than three events may be selected if there are ties in any of the top three positions. Then, we identify the central event in the candidate set by requiring different combinations of the highest ranks, including the highest size rank E_Z, highest stretch rank E_T, highest position rank E_P and highest enriched size rank E_E. In addition, we identify an event set E_R which includes events whose contextual modality or realis status is "Actual" and use the set for constraining central event identification. Specifically, we define three rule based classifiers which begin with strict rules followed by relaxed rules in subsequent passes. The system **Coreference** uses size, stretch and position ranks, **Coreference + Subevent** considers enriched size rank as well, and **Coreference + Subevent + Realis** further combines realis status with each rank in favor of specific events.

5.2 Statistical Regression Classifiers

We trained a linear as well as a nonlinear regression classifier to integrate the same set of ranking rules as features for identifying central events, by using the standard ordinary least squares **linear regression** (Galton, 1886) model and the epsilon-support vector regression (**SVR**) (Vapnik, 1995) model with radial basis function kernel respectively. Input to both the linear and nonlinear regression classifiers consists of 20 (19) dimensional vector, 4 dimensional categorical vector for each of the size, stretch, position and enriched size ranks and 4 (3) dimensional categorical vector for realis attribute for RED (KBP) corpus. The models were implemented using scikit-learn module (Pedregosa et al., 2011). The SVR classifier uses **rbf** kernel with γ coefficient of 0.05 and all other parameters are left to be the default values.

5.3 Coreference: Predicted

We further used system predicted coreference relations to calculate size, stretch and position ranks and used them to identify central events, where coreference relations were predicted by a neural network based pairwise classifier using event lemmas, parts-of-speech tags and event arguments as features. The classifier was trained on the corpus used in the Event Nugget Detection and Coreference task in the TAC KBP 2016 (Ellis et al., 2015).

Specifically, the classifier uses a common neural layer shared between two event mentions that embed event lemma and parts-of-speech tags and then calculates cosine similarity, absolute and euclidean distances between two event embeddings. Classifier also includes a neural layer component to embed event arguments that are overlapped between the two event mentions. Its output layer takes the calculated cosine similarity, euclidean and absolute distances between event mention embeddings as well as the embedding of the overlapped event arguments as input, and output a confidence score to indicate the similarity of the two event mentions[6]. We used 300 dimensional word embeddings (Pennington et al., 2014) and one hot 37[7] dimensional pos tag embeddings. In addition, we used (Lewis et al., 2015) for semantic role labeling to obtain event arguments.

[6]We implemented our classifier using the Keras library (Chollet, 2015)

[7]Corresponding to the unique 36 POS tags based on the Stanford POS tagger (Toutanova et al., 2003) and an additional 'padding'.

6 Evaluation

6.1 Baseline Systems

Three Heuristics Based Classifiers: The three systems *Main event: Headline*, *First event: First sentence* and *Main event: First sentence* chose the main event (syntactic root) in headline, the first event in the first sentence and the main event (syntactic root) in the first sentence as the center event respectively.

Random Walk Based Ranking Systems: implemented the random walk based vertex ranking algorithm (Mihalcea and Tarau, 2004) on graphs generated using human annotated event relations. The motivation is to decide the importance of an event mention within an event graph of a document[8] based on the importance of its related event mentions[9]. The system *Random walk: All Relations* uses coreference, sub-event, set/ member, temporal and causal relations to build the graph while the system *Random walk: Coref+SE* only considers event coreference and sub-event relations. We evaluate both systems on documents from the RED corpus only as it extensively annotates event relations which yields a connected graph for each document. However, the graphs generated for documents in the KBP corpus often contain many disconnected components and thus are not suitable for these systems.

6.2 Results

We evaluated all the systems using the rest 20 documents from the RED corpus and all the 74 docu-

[8]We build an event graph for a document by using undirected edges for coreference relations and directed edges for other relations including set/ member, sub-event, temporal and causal relations. This is mainly meant to retain the symmetrical property of coreference relations. Moreover, since coreference link can easily create cycles in the graph, we utilize its transitivity property and link all the coreferent event mentions to its first instance in the document only.

[9]We rank event mentions by using the vertex scoring algorithm proposed in Brin and Page (2012).

$$S(V_i) = (1-d) + d \sum_{j=IN(V_i)} \frac{1}{|OUT(V_j)|} S(V_j) \quad (1)$$

where $IN(V_i)$ and $OUT(V_j)$ represent the set of event mentions that are predecessors and successors to V_i respectively. Also, d is a damping vector that is kept 0.85 in our experiments. We initially assign random values to all the event mentions in an event graph and then update scores for all event nodes using equation 1 after each iteration. Computation stops when the sum of differences between the scores computed for all event mentions at two successive iterations reduces below 0.01.

Model	Rec	Prec	F1
Richer Event Description (RED)			
Main event: Headline	45.0	45.0	45.0
First event: First sentence	10.0	10.0	10.0
Main event: First sentence	40.0	40.0	40.0
Random walk: All Relations	40.0	40.0	40.0
Random walk: Coref+SM	45.0	45.0	45.0
Coreference	75.0	55.55	63.82
Coreference + Subevent	75.0	62.5	68.18
Coreference + Subevent + Realis	80.0	66.67	72.73
Linear Regression	63.33	63.33	63.33
SVR	66.67	66.67	66.67
Coreference: Predicted	45.0	45.0	45.0
KBP 2015			
Main event: Headline	45.94	45.94	45.94
First event: First sentence	39.19	39.19	39.19
Main event: First sentence	39.19	39.19	39.19
Coreference	77.03	54.81	64.04
Coreference + Subevent	77.03	60.0	67.46
Coreference + Subevent + Realis	78.37	66.67	72.05
Linear Regression	66.21	61.25	63.63
SVR	67.56	62.5	64.93
Coreference: Predicted	48.65	45.56	47.05

Table 1: Experimental Results.

ments from the KBP 2015 corpus. The two regression classifiers were evaluated using 5-fold cross-validation on each corpus. We expect a system to identify only one central event for each document. If a system predicts more than one central event, we will penalize the system on precision strictly and treat each wrongly predicted event as a false hit. Table 1 shows the comparison results.

The heuristic based systems obtained a low recall on both corpora, which indicates that simple heuristics miss a large proportion of cases. Both random walk based systems suffered from a low recall of 40-45% as well when applied to the RED corpus, due to the fact that graph-based ranking models do not effectively capture discourse layout features of co-referential event mentions.

In contrast, the rule based system **Coreference** achieved the recall above 75% on both corpora when using annotated event coreference relations. The system **Coreference + Subevent + Realis** further improves the precision of central event identification by over 11% on both corpora after considering subevents and the realis status in the rules, which facilitate accurate identification of the central event among multiple foreground events. The high recall and precision indicate that the insightful rules exploiting properties of event chains are able to capture the overall texture in the discourse. Then compared with rule based systems, the two statistical classifiers that integrate the same set of rules as features do not further improve the central event identification performance. But when using system predicted noisy event coreference relations, the rule based system **Coreference: Predicted** performed dramatically worse than its counterpart using gold event chains (system **Coreference + Subevent + Realis**). This is unsurprising though considering the relatively low performance of current event coreference resolution systems.

6.3 Analysis

To gain a better understanding of how noise in system predicted event coreference links influences central event identification performance, we analyzed the documents where the system **Coreference: Predicted** failed to identify the central event. We found that both types of event coreference resolution errors, missed coreference links as well as wrong links, cause problems, especially in calculating the Size Rank and the Stretch Rank for an event. Specifically, the first type of errors can demote both ranks of the correct central event while the second type of errors can wrongly promote one of the two ranks for non-central events.

7 Conclusions

We have presented a new task of identifying the central event for a document. Based on our annotations, we discussed the role of central events in enabling a coherent discourse and the distributional characteristics of central events. We especially emphasized on the importance of event coreference in identifying central events. Inspired by these observations, we designed a rule-based classifier that achieved high recall and precision in identifying central events. The low performance of the classifier using system predicted event coreference relations indicates that significant efforts are needed to further improve event coreference resolution performance in the future.

Acknowledgments

This work was partially supported by the National Science Foundation via NSF Award IIS-1755943. Disclaimer: the views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF or the U.S. Government.

References

Sergey Brin and Lawrence Page. 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks* 56(18):3825–3833.

François Chollet. 2015. Keras. https://github.com/fchollet/keras.

Nan Decker. 1985. The use of syntactic clues in discourse processing. In *Proceedings of the 23rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pages 315–323.

Joe Ellis, Jeremy Getman, Dana Fore, Neil Kuster, Zhiyi Song, Ann Bies, and Stephanie Strassel. 2015. Overview of linguistic resources for the tac kbp 2015 evaluations: Methodologies and results. In *Proceedings of TAC KBP 2015 Workshop, National Institute of Standards and Technology*. pages 16–17.

Elena Filatova and Vasileios Hatzivassiloglou. 2004a. Event-based extractive summarization. In *Proceedings of ACL Workshop on Summarization*. Barcelona, Spain., volume 111.

Elena Filatova and Vasileios Hatzivassiloglou. 2004b. A formal model for information selection in multi-sentence text extraction. In *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, page 397.

Francis Galton. 1886. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15:246–263.

Joseph Evans Grimes. 1975. *The thread of discourse*, volume 207. Walter de Gruyter.

Paul J Hopper and Sandra A Thompson. 1980. Transitivity in grammar and discourse. *language* pages 251–299.

Roderick Kay and Ruth Aylett. 1996. Transitivity and foregrounding in news articles: Experiments in information retrieval and automatic summarising. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '96, pages 369–371. https://doi.org/10.3115/981863.981918.

Mike Lewis, Luheng He, and Luke Zettlemoyer. 2015. Joint a* ccg parsing and semantic role labelling. In *EMNLP*. pages 1444–1454.

Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *EMNLP*. volume 4, pages 404–411.

Teruko Mitamura, Zhengzhong Liu, and Eduard Hovy. 2015. Overview of tac kbp 2015 event nugget track. In *Text Analysis Conference*.

Sean Monahan and Mary Brunson. 2014. Qualities of eventiveness. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. pages 59–67.

Tim O'Gorman, Kristin Wright-Bettner, and Martha Palmer. 2016. Richer event description: Integrating event coreference with temporal, causal and bridging annotation. *Computing News Storylines* page 47.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *EMNLP*. volume 14, pages 1532–1543.

Tsunoda Tasaku. 1981. Split case-marking patterns in verb-types and tense/aspect/mood. *Linguistics* 19(5-6):389–438.

K. Toutanova, D. Klein, C. Manning, and Y. Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*.

Tasaku Tsunoda. 1985. Remarks on transitivity. *Journal of linguistics* pages 385–396.

Shyam Upadhyay, Christos Christodoulopoulos, and Dan Roth. 2016. " making the news": Identifying noteworthy events in news articles. In *Proceedings of the Fourth Workshop on Events*. pages 1–7.

Vladimir N. Vapnik. 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA.