Event Coreference Resolution using Mincut based Graph Clustering

Sangeetha.S[1] and Michael Arock[2]

Department of Computer Applications,
National Institute of technology, Tiruchirappalli, India
[1]sangeetha@nitt.edu [2]michael@nitt.edu

ABSTRACT

To extract participants of an event instance, it is necessary to identify all the sentences that describe the event instance. The set of all sentences referring to the same event instance are said to be corefering each other. Our proposed approach formulates the event coreference resolution as a graph based clustering model. It identifies the corefering sentences using minimum cut (mincut) based on similarity score between each pair of sentences at various levels such as trigger word similarity, time stamp similarity, entity similarity and semantic similarity. It achieves good B-Cubed F-measure score with some loss in recall.

KEYWORDS

Event coreference resolution, ACE event.

1. INTRODUCTION

Automatic content extraction (ACE) programme supports automatic processing of source language data. ACE [1, 21] defines three basic kinds of information to be extracted from natural language text such as entities, relations and events. An Event is a specific occurrence involving participants. An Event is something that happens. *Event extent* is the sentence in which the event is described. *Event trigger* is the word in the sentence that clearly expresses the occurrence of an event. An event comprises event participants, which are the entities that participate in the event with different roles.

To extract the participants of the event it is necessary to identify set of sentences describing the same event instance. The set of sentences referring to an event instance are said to be corefering each other. Generally, event coreference resolution problem is considered as either pair-wise event coreference resolution or coreference chaining. Pair-wise event coreference resolution identifies whether the given pair of sentences corefer each other or not whereas, coreference chaining identifies group of corefering sentences. As our motive is to identify all sentences describing an event instance for further processing i.e. to extract its participants, our approach treats the problem as coreference chaining problem.

Table.1 shows the set of sentences talking about the same event instance of event type *Contact: Meet* from which the participants such as [*Entity-arg:* Kofi Annan and Mohammed Al-Douri, *Time:* Thursday] are to be extracted. It is necessary to group all these corefering sentences to extract its participants.

The remaining sections of the paper are organized as follows. The next section describes the related work in the field of event coreference resolution. Section 3 provides a description of graph based modelling of our problem, section 4 explains the similarity score calculation, Section 5 explains experimental setup, results, discussions and Section 6 concludes the paper.

Sundarapandian et al. (Eds): CoNeCo,WiMo, NLP, CRYPSIS, ICAIT, ICDIP, ITCSE, CS & IT 07,
pp. 253–260, 2012. © CS & IT-CSCP 2012 DOI : 10.5121/csit.2012.2422

Table 1. The set of sentences talking about the same event instance of event type ***Contact: Meet***

Four Event Mentions (EM1) of type **{Contact: Meet}**
EM1 {This is my last word to you, he told hordes of journalists who chased him Thursday at his New York residence and U.N. headquarters seeking comment on his **talks** with the secretary-general, his future and the war}
EM2 {Annan also declined to comment on Thursday's **meeting** with Al-Douri}
EM3 {Annan said early Thursday before **meeting** Al-Douri that in talks with the envoy Monday, he didn't ask for an asylum or protection and he didn't ask me for help with his status}
EM4 {Iraq's U.N. Ambassador Mohammed Al-Douri, the first Iraqi official to concede defeat in the U.S. led war, **met** privately with Secretary-General Kofi Annan, but refused to talk about rumors that he was planning to leave New York}

2. BACKGROUND

Even though entity coreference resolution problem was attempted by many researchers such as [19], event coreference resolution was not received much attention. Earlier work on event coreference resolution [14] identified corefering sentences based on semantic score and attribute score after semantic role labeling which was understood as annotating participants of the event. Later David Ahn [8] treated the problem as a classification problem with rich set of features. Zhen Chen et al [6] proposed an agglomerative clustering approach to solve the event coreference resolution problem by considering the features based on event participants and attributes. An active mention is merged with the prior event based on probability identified using MaxEnt model. Naughton. M [18] solved coreference resolution using SVM classifier and hierarchical agglomerative clustering. Zhen Chen et al [4] treated this problem as graph clustering and no coreference resolution scores have been reported. They constructed coreference matrix by matching event participants and trigger word. C.A. Bejan et al [7] proposed an unsupervised Bayesian model for solving the problem using lexical, wordnet features and semantic features. The semantic features include the annotations of the text with argument roles. A. Elkhlifih et al [26] used a similarity measurement FSIM which combines the similarity between sentences and the distance between them. Peifeng Li et al [29] applied hierarchical clustering algorithm to cluster event mentions based on plain features and structured features.

McConky et al [24] proposed an approach that finds event coreference based on the combination of the similarity of the two event descriptions based on vector space model, the similarity of location and time. They used dynamic weighting approach to combine the three similarity scores together. Heeyoung Lee et al [25] constructs clusters of entity and event mentions using features based on semantic role dependencies and linear regression for merging clusters.

Motivated by [6] which concluded that the contribution of event participants in event coreference resolution is lesser, by [27] which argue that the processing sentences individually for event extraction cause ambiguity and event participants needs to be identified after collecting all the sentences referring to the same event instance, we propose an approach that,

- Does not require the event sentences to be annotated with event participants as opposed to all the above related approaches except [23, 24, 26].
- Combines scores based on trigger words, entities, semantic similarity of nonentity words and timestamp and clusters documents using mincut clustering.
- Uses raw sentences with entities identified as a pre-processing step.

3. GRAPH BASED MODELLING

Let VS= $\{v_1, v_2.., v_n\}$ be the set of vertices that represents event sentences of same type within a document and ES= $\{ed_1, ed_2.., ed_m\}$ be the set of edges connecting the vertices and weight on each edge denotes the similarity score between the vertices connected by the edge. Now our problem is considered as dividing the graph into sub graphs which represent the group of sentences that represent an event instance.

A graph can be divided into sub graphs using the well known min cut based graph clustering approach. Our approach employs min cut based clustering [5, 11, 12, 20] for the following reasons

- It is a Bicriteria approach as it takes care of intercluster and intra cluster quality which produces best clusters.
- The time complexity is |V|
- “The minimum cut clustering technique creates clusters that have small inter cluster cuts and large intra cluster cuts. It provides strong connectedness within the cluster” [11].

Min-cut based clustering algorithm [11] is given in Table 2. As specified in [11] number of clusters created depends on the value of α. It is empirically set with average of all scores within the document.

Table 2. Cut clustering Algorithm

```
Cut Clustering Algorithm (G (V, E), α)
{
Let V’ = V U t
For all nodes v Є V
Connect t to v with edge of weight α
Let G (V’, E’) be the expanded graph after connecting t to V
Calculate the minimum-cut tree T’ of G
Remove t from T’
Return all connected components as the clusters of G
}
```

4. CONSTRUCTING COREFERENCE MATRIX

Coreference matrix is an N X N similarity matrix which is constructed for all pairs of same event type sentences in a document, where N is number of same event type sentences.

Maria Recasal et al [16] stated that “Event coreference resolution can be solved using paraphrase resolution. Single event can be reported in many newspapers in different ways keeping same NP such as name, date and numbers unchanged”. We symbolize entities in each sentence with entity code and then the task of constructing coreference matrix for every pair of sentences becomes the similarity score calculation between paraphrasing pairs. Then we add time stamp based on annotations of ACE [1,21] for each time based word in the sentence. Corefering sentences may have either all entities in common or some entities in common. Now the similarity scores between these paraphrases are identified [3, 9, 13] under four categories such as entity based score, semantic score trigger score and time based score using Jaccard coefficient [15].

Let
$E_i = \{e_{i1}, e_{i2},., e_{ix}\}$ be the set of entities in sentence 'i'
$T_i = \{t_{i1}, t_{i2}, t_{iy}\}$ be the set of time based words in sentence 'i'.
$W_i = \{w_{i1}, w_{i2}, w_{iy}\}$ be the set of other words in sentence 'i'.
$S_i = E_i \cup T_i \cup W_i$ where i =1. n.

Entity based score (EBS) is calculated by finding the proportion of number of common entities with respect to average number of entities using eq.1.

$$EBS = \frac{2 * |E_i \cap E_j|}{|E_i| + |E_j|} \quad (1)$$

Semantic score (SS) is calculated by finding the words except entities and finding the proportion of number of common words with respect to average number of words plus the proportion of number words that belong to the same synset [10,17] with respect to the average number of words using eq.2

$$SS = \frac{2 * |W_i \cap W_j|}{|W_i| + |W_j|} + \frac{2 * |W_i \cap^a W_j|}{|W_i| + |W_j|} \quad (2)$$

where $\cap^a = same\ synset$

Trigger based score (TBS) is calculated by assigning weights to different levels of matching between the pairs. Weight w is assigned to same trigger words, weight 0.5w is assigned if trigger words belong to same synset and 0.25w is assigned if the semantic similarity score [22] of triggers are above as given in eq.3.

$$TBS = \begin{cases} W, & if\ common \\ 0.5W, & if\ in\ same\ synset \\ 0.25W, & if\ semantic\ similarity > T \end{cases} \quad (3)$$

where T is a threshold.

Time based similarity score (TimS) is calculated by assigning weights based on time stamp. More than one time stamp is associated with the event based on the time based words such as now, yesterday etc. The time based score is assigned by finding the proportion of number of exact timestamp including day and hour with respect to average number of timestamps, and proportion of number of same day timestamp with respect to average number of timestamps using eq.4.

$$TimBS = \frac{2 * |T_i \cap^b T_j|}{|T_i| + |T_j|} + \frac{2 * |T_i \cap^c T_j|}{|T_i| + |T_j|} \quad (4)$$

where $\cap^b = same\ day\ and\ hour$ $\quad \cap^c = same\ day$

The sum of all the four scores is the similarity score and it is assigned as edge weight of the graph.

5. EXPERIMENT AND RESULTS

5.1 Data and evaluation metric

For our experiment we have used ACE 2005 corpus [1, 21]. It contains 599 documents in different categories such as newswire, newsgroups, weblogs, broadcast news, broadcast conversation, telephone speech transcripts. All event sentences of a particular type within a document are given as input and the clusters that represent set of sentences describing about the same event instance are obtained as output.

As MUC metric does not give any credit for separating out singletons, we have adopted B-cubed measure for evaluation [2, 28], the best suited evaluation metric for coreference resolution. Assume ‘e’ is an event in a given document D and N is the total number of event instances in D. Let C_e be the number of correct elements in the output chain containing event e, and R_e is the number of total elements in the output chain containing event e, and T_e is the number of elements in the truth chain containing event e. Then the precision and recall are defined as follows.

$$Precision = \frac{|C_e|}{|R_e|} \quad (5)$$

$$Recall = \frac{|C_e|}{|R_e|} \quad (6)$$

The final precision and recall scores are computed by the following two formulae.

$$Final\ Precision = \sum_{i=1}^{N} wt_i \times Precision_i \quad (7)$$

$$Final\ Recall = \sum_{i=1}^{N} wt_i \times Recall_i \quad (8)$$

Where wt_i is the weight assigned to event i in D. As given in [2] the weight wt_i for entity i is computed such that $wt_i = 1/N$.

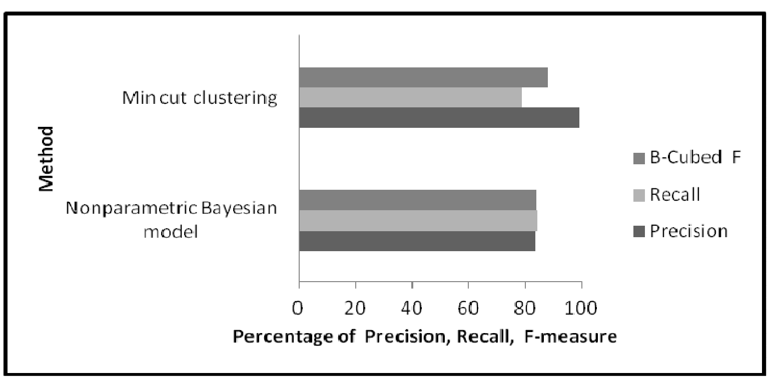

Figure 1. Comparison of results of Min cut clustering approach with Nonparametric Bayesian model

5.2 Results and Discussion

Fig.1 shows comparison of the results obtained by our approach with nonparametric Bayesian model [7]. Our approach shows promising result of 87.7% B-Cubed F measure. This work was motivated by the results of Z.Chen et al [6] which used features based on event participants, attributes and confirmed that the contribution of participants in event coreference is less. Our result confirms it and improves the coreference resolution task with the help of scores based on entity, trigger, semantic similarity and timestamp.

Event though our approach achieves improvement in F-measure compared to nonparametric Bayesian model, it lags behind in recall with high precision [7]. This is because most of the incorrect clusters obtained by our approach are singletons.

Fig.2 shows the results obtained by our approach using different combination of scores in coreference matrix. When we form clusters based on TWS and TimBS F-measure is slightly low compared to all the other combinations. TWS and EBS show improvement of 0.4% over TWS and TimBS and a 0.2% improvement over the remaining two combinations. Hence, rather than having data annotated with participants such as Agent, Victim, Instrument etc , similarity scores between sentences based on EBS and TBS improves the coreference resolution.

In the above result, final f-measure gets reduced if we include TimBS. This is because, only limited number of the sentences in the data set have timestamps; So TimBS score in most of the sentence pair are 0. But TimBS is an important score which identifies the event that happen at the same time in real situation. If the percentages of sentences with time based annotations increases, then the result may further improve.

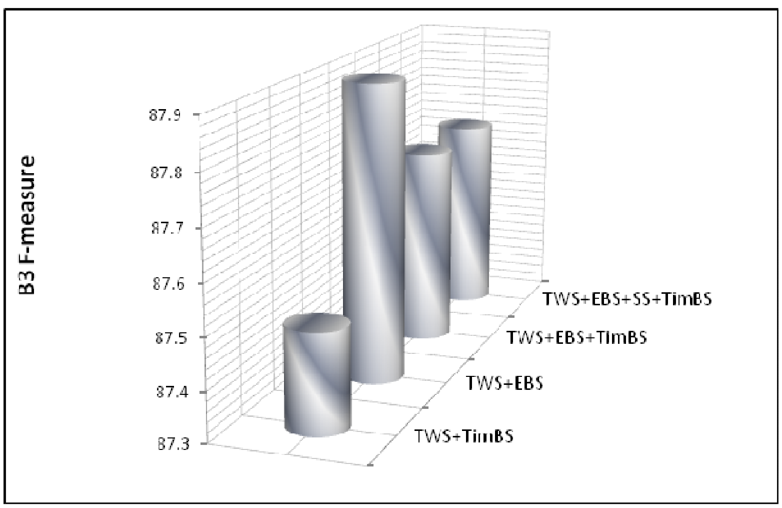

Figure 2. Results obtained by our approach using different combination of scores in coreference matrix

6. CONCLUSION

Our proposed approach formulates the event coreference resolution as a graph based clustering model. The clusters are found using minimum cut (mincut) based on similarity score between each pair of sentences at various levels such as trigger word similarity, time stamp match, semantic similarity and entity similarity. The main advantage of our approach is that it does not require event sentences to be annotated with participants of the event and it empirically achieves good F-measure score compared to other approaches. Hence we conclude that, rather than having data annotated (time consuming process) with participants such as Agent, Victim, instrument etc. similarity scores between sentences based on EBS and TBS improves the event coreference resolution.

REFERENCES

[1] ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3 2005.07.01, Linguistic Data Consortium. http.// www.ldc.upenn.edu

[2] Bagga, A. and Baldwin, B. (1998), "Algorithms for scoring coreference chains", *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pp.563--566.

[3] Bird Steven, Ewan Klein, Edward Loper, (2009), *Natural Language Processing with Python*, O'Reily Media.

[4] Chen, Z. and Ji. H. (2009a)," Graph-based Event Coreference Resolution", *In Proc. ACL-IJCNLP 2009 workshop on TextGraphs-4. Graph-based Methods for Natural Language Processing.*

[5] Chen, Z, (2010), "Graph-based Clustering and its Application in Coreference Resolution", *In Proceedings of the 2010 Workshop on Graph-based Methods for Natural Language Processing, ACL 2010,* pp.1–9.

[6] Chen Z., Ji, H. Haralick, R, (2009b), "A Pairwise Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution", *In Proc. RANLP 2009 workshop on Events in Emerging Text Types*.

[7] Cosmin Adrian Bejan and Sanda Harabagiu, (2010), "Unsupervised Event Coreference Resolution with Rich Linguistic Features", *In Proc. of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden

[8] David Ahn. (2006), "The stages of event extraction", *In Proc. of the Workshop on Annotating and Reasoning about Time and Events*, pp. 1–8.

[9] Edward Loper, Steven Bird. (2002), "NLTK the Natural Language Toolkit", *In Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics*, 63-70.

[10] Fellbaum, C. (Ed.) (1998), *WordNet an electronic lexical database*, MIT Press http://wordnet.princeton.edu

[11] Flake, G. W. Tarjan, R. E. and Tsioutsiouliklis K, (2003), "Graph clustering and minimum cut trees". *Internet Mathematics*, Vol.1 no.4, pp.385–408.

[12] Gomory, R. E. Hu,T. C, (1961), "Multi-terminal network flows", *Journal of the Society for Industrial and Applied Mathematics*, 9.

[13] http://www.nltk.org

[14] Humphreys, K. Gaizauskas, R. Azzam. S, (1997), "Event coreference for information extraction", *In Proc of the ACL Workshop on Operational Factors in Practical Robust Anaphora Resolution for Unrestricted Texts*

[15] Jaccard P. (1901), "Etude comparative de la distribution florale dans une pard, ortion des Alpes et des Jura", Bulletin de la Société Vaudoise des Sciences Naturelles, 37, pp.547–579.

[16] Maria Recasal, Marta Vila. (2010), "On Paraphrase and Coreference", *Computational Linguistics*, Vol.36 no.4, pp. 639-647, MIT Press.

[17] Miller, G. (1995), "Wordnet, A lexical database for English", Communications of the ACM 38, pp.39-41.

[18] Naughton, M. (2009), *Sentence-Level Event Detection and Coreference Resolution.* Doctoral Thesis, School of Computer Science and Informatics, University College Dublin.

[19] Nicolae C. and Nicolae G. (2006), "Bestcut. A graph algorithm for coreference resolution", *In proceedings EMNLP*, 275–283, Sydney.

[20] Schaeffer, S. E. (2007), "Graph clustering", *Computer Science Review*, 1(1).27–64.

[21] Walker, C. Strassel, S. Medero, J., (2006), Consortium, L. D. *ACE 2005 Multilingual Training Corpus.* Linguistic Data Consortium, University of Pennsylvania.

[22] Wu, Zhibiao and Martha Palmer. (1994), Verb semantics and lexical selection, *In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 133–138.

[23] Bagga and B. Baldwin, “Cross-Document Event Coreference: Annotations, Experiments, and Observations,” *In Proc. ACL-99Workshop on Coreference and Its Applications*, 2009, pp. 1-8.

[24] Katie -McConky, Rakesh Nagi, Moises Sudit and William Hughes (2012), “Improving Event Co-reference By Context Extraction and Dynamic Feature Weighting”, *2012 IEEE International MultiDisciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*, New Orleans, LA, pp.38-43

[25] Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu, Dan Jurafsky,’ (2012),”Joint Entity and Event Coreference Resolution across Documents”, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,* pp. 489–500,

[26] A. Elkhlifi and A. Faiz (2009),”Automatic Annotation Approach of Events in News Articles,” *International Journal of Computing & Information Sciences*. 7(1). pp.40-50.

[27] Huang, R. and Riloff, E. (2012) "Modeling Textual Cohesion for Event Extraction", *Proceedings of the 26th Conference on Artificial Intelligence (AAAI 2012)*

[28] http://nlp.cs.qc.cuny.edu/kbp/2011/scoring.html (Original B-Cubed)

[29] Peifeng Li, Qiaoming Zhu, Xiaoxu Zhu, (2011) “A Clustering and Ranking Based Approach for Multi-document Event Fusion “, *12th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*,pp.161-165