Cross-Document Coreference Resolution for Entities and Events

Ph.D. Thesis Proposal

Chris Tanner

May 16, 2018

Abstract of "Cross-Document Coreference Resolution for Entities and Events"

Abstract Here

Contents

Chapter 1
Introduction

Thesis Statement: I propose a novel, neural-based mention-pair model for cross-document coreference resolution for events, which uses few lexical features and addresses shortcomings of traditional clustering approaches. I will extend this work by jointly modelling both entities and events, while using structured information (e.g, parse trees). Last, we aim to improve mention detection, whereby we develop an all-inclusive, end-to-end system which jointly resolves mention boundaries and coreference predictions.

1.1 Motivation

Coreference Resolution remains a fundamental NLP task, as it is an essential component for any system that desires "understanding" textual data. That is, in order to accurately model meaning, one must at the very least understand which items are concerning the same underlying objects. As a simple example, if one performs a Web search for "President Barack Obama", some of the web page search results will contain sentences which only refer to him as "Obama", "he", or "The President," and correctly using this information is essential for returning relevant information to the user's query. Further, coreference resolution is useful for information extraction [17], question answering [23], topic detection [1], summarization [10], and more.

1.2 Problem Statement

Coreference resolution is the task of identifying – within a single text or across multiple documents – which *mentions* refer to the same underlying discourse object.

A **mention** is a particular instance of word(s) in a document which represent an *entity* or *event*, such as *Barack Obama*, *he*, or *announced*.

An **entity** may be a person, location, time, or an organization. The mentions which refer to them may be *named*, *nominal*, or *pronominal*:

- Named mentions are represented by proper names (e.g., André Benjamin or Pakse, Laos)
- Pronominal mentions are represented by pronouns (e.g., she or it)

- Nominal mentions are represented by descriptive words, not composed entirely of a named entity or pronouns (e.g., The well-spoken citizen)

An **event** can generally be thought of as a specific action. Quine [25] was the first to propose that an event refers to a physical object which is grounded to a specific time and location, and that two events are identical (i.e., co-referent) if they share the same spatiotemporal location. This definition has become the general consensus within the community[1]. Specifically, two co-referent events must share the same *properties* and *participants*. For example, in Figure 1.1, sentences #1 and #2 contain the co-referent events ("placed" and "put"), yet neither are co-referent with events in sentence #3. Often times, the participants (arguments) may be referred to in different ways, implied, or missing altogether.

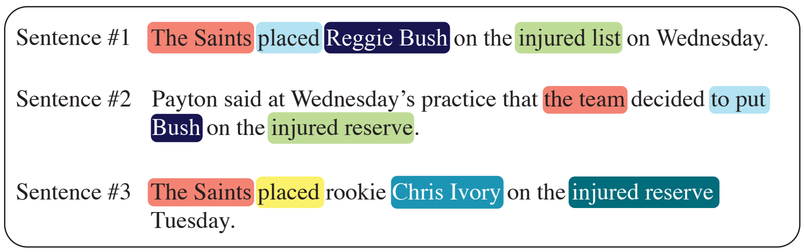

Figure 1.1: Sample of a coreference resolution corpus (ECB+), depicting gold coref mentions as having shared box colors.

Coreference resolution is concerned with linking either entities together and/or events together; that is, entities shall not be linked to events, and doing so would be considered an incorrect link. Although one may be interested in evaluating coreference systems by their ability to correctly link *pairs* of mentions [32], coreference resolution is ultimately a clustering task, whereby we wish to group all like-mentions together, as shown with colored boxes in Figure 1.1. Specifically, coreference systems aim to find a globally-optimal fit of mentions to clusters, whereby every mention m in the corpus is assigned to exactly one cluster C, such that every $m_i, m_j \in C$ are co-referent with each other. If a given m_i is not anaphoric with any other m_j, then it should belong to its own C with a membership of one.

Given a corpus of text documents, coreference resolution can be performed and evaluated on either a **within-document** or **cross-document** basis:

- **Within-document** is when each mention may only link to either (1) no other mention; or (2) other mentions which are contained in the same document. Even if the gold truth data denotes a mention should link with a mention from a different document, we ignore these links during the evaluation.

- **Cross-document** is when the entire corpus is available for linking; a mention is eligible to be co-referent with mentions in any other document, and the evaluation reflects the same. As described in [29], cross-document evaluation is normally conducted by transforming the entire corpus into a "meta-document."

[1] Hovy, et. al. [16] provide an in-depth study of varying definitions.

1.3 Coreference Systems

Coreference systems are predicated upon first having entity/event mentions identified, via a separate, distinct task called *mention detection*. Next, these identified mentions are used by coreference resolution models.

1.3.1 Mention Detection

This initial mention identification process is a separate line of research and has remained a fundamental task of NLP for several decades [22]. When concerned with entities, research is commonly referred to as *named entity recognition* or *entity recognition*. When concerned with events, research is commonly referred to as *event detection*.

Named Entity Recognition

The earliest work started in 1991 with the task of identifying company names [26]. In 1996, the MUC-6 conference [14] focused on Information Extraction tasks, which included coining the phrase "named entity" and drastically increasing attention to mention detection. Early work demonstrated state-of-the-art performance with Hidden Markov Models (HMMs) [4] and Conditional Random Fields (CRFs) [21]. Presently, the best performing systems use similar models — but from a deep learning framework — which include Bi-directional LSTMs [5, 15] and Convolution Neural Nets (CNNs) [20].

Event Detection

Event Detection has received significantly less attention than Named Entity Recognition; however, the task of semantic role labelling (SRL) addresses a similar and more encompassing problem; SRL is a shallow semantic parsing task, whereby the goal is to identify each predicate in a sentence, along with its constituents and how they fill a semantic role — specifically, to determine the role (e.g., Agent, Patient, Instrument, etc) and their adjuncts (Locative, Temporal, Manner, etc). [13, 24]. In short, both SRL systems and Event Detection systems have often relied on using many lexical and syntactical features, including those from constituency parsers [28], dependency parsers [18], etc. However, like entity recognition, recent state-of-the-art systems for Event Detection use Bi-directional LSTMs and CNNs [12].

1.3.2 Coreference Resolution

As mentioned, coreference systems aim to create the correct clusters of mentions; however, due to the number of possible mention-to-cluster combinations, finding a globally-optimal assignment of clusters is NP-Hard and thus computationally intractable. In attempt to avoid this, systems typically perform pairwise-mention predictions, then use those predictions to build clusters. The specific modelling strategies for such approximately fall into two categories: (1) mention-ranking / mention-pairs; and (2) entity-level / event-level, as described below; however, it is worth noting that there is no unanimously-dominate modelling paradigm, as state-of-the-art results often come from any of the ones listed below.

Mention-ranking models define a scoring function $f(m_i, m_j)$ which operates on a mention m_j and possible antecedent m_i, where m_i occurs earlier in the document and could be null (represented by ϵ and denoting that m_j is non-anaphoric); e.g., Wiseman, et. al.'s [31]. These models aim to find the ideal m_i

antecedent for every m_j mention. After every mention has decided to link to ϵ or a previous mention, it is common practice to define each cluster simply by joining together all mentions which are connected by a single path. This is a potential weakness, as it asserts the transitive property holds true (e.g., if m_3 predicts m_2 as its antecedent, and m_2 predicts m_1 as its antecedent, then $\{m_1, m2, m3\}$ are all connected, which could be a bad decision, as there is no direct consideration given to the relatedness between m_1 and m_3.

Mention-pair models score all pairs (m_i, m_j), in contrast to mention-ranking models which aim to find the ideal m_i antecedent for every m_j. After every pair of mentions has been scored, it is common practice to cluster mentions in a best-first or easy-first manner (e.g. agglomerative clustering). Because mention-pair models base their predictions on the information from just two mentions at a time, they are by definition less expressive than entity/event-level models. Yet, their inference can be relatively simple and effective, allowing them to be fast and scalable. Consequently, they have often been the approach used by many state-of-the-art systems [11, 27], including our work described in this proposal.

Entity/Event-level Instead of operating on a mention-level basis, these models differ in that they focus on building a global representation of each underlying entity or event, the basis of which determines each mention's membership [8, 30]. These models are attractive due to the intuitive nature of modelling each entity with its own representation; however, challenges include (1) deciding how to represent each entity as it is being developed; (2) decided how many entities to model.

The aim is for the above definitions and descriptions to provide sufficient background to understand all of our subsequent chapters in this proposal, including related research, our research thus far, and our proposed work of (1) combining entity and event coreference into one model; and (2) combining our own in-house mention detection with our coreference models.

Chapter 2

Related

2.1 Motivation

The seminal research on event coreference can be traced back to the DARPA-initiated MUC conferences, whereby the focus was on limited scenarios involving terrorist attacks, plane crashes, management succession, resignation, etc. [2, 17].

In the present day, Deep Learning is revolutionizing NLP. And, although coreference resolution has been researched for several decades, only recently have a few publications successfully applied deep learning to coreference – almost all of which have been for *entity* coreference. We attribute this relatively small amount of deep learning models to the fact that coreference resolution is inherently a clustering task, which tends to be a non-obvious modality for deep learning. Since our work so far (1) uses deep learning and (2) operates on events by using the ECB+ corpus, we organize the related research accordingly.

2.2 Deep Learning Approaches

To the best of our knowledge, there are six publications which apply deep learning to coreference resolution, five of which focus on entity coreference.

2.2.1 Coreference Resolution for Entities

Sam Wiseman, et. al. [30, 31] trained a mention-ranking model with a heuristic loss function that assigns different costs based on the types of errors made, and their latter work used mention-ranking predictions towards an entity-level model.

Clark and Manning [7, 8] also built both a mention-ranking model and an entity-level model, the former of which was novel in using reinforcement learning to find the optimal loss values for the same four distinct error types defined in Wiseman's, et. al. [31] work.

Most recently, Lee, et. al. [19] developed the first end-to-end coreference system which is only trained on gold clusters and uses few features (speaker information, genre, span distance, mention width) to do both mention detection and coreference resolution on those mentions. Notably, this paper is the most similar to our proposed, desired goal as described in Section **??**.

2.2.2 Systems using ECB+ Corpus

For our research, we make use of the ECB+ corpus [9], which we further describe in Section **??**. This rich corpus provides annotations for both entities and events, yet most research chooses to focus on using *either* events *or* entities, not both. To the best of our knowledge, there are only two papers which focus on the event mentions of ECB+: The Hierarchical Distance-dependent Chinese Restaurant Process (HDDCRP) model by Yang, et. al. [33] (not a Deep Learning approach) and Choubey's and Huang's Iteratively-Unfolding approach [6] (a Deep Learning approach).

HDDCRP Model

Yang, et. al's HDDCRP model [33] uses a clever mention-pair approach, whereby they first use logistic regression to train parameters θ for the similarity function in Equation 2.1.

$$f_\theta(x_i, x_j) \propto \exp\{\theta^T \psi(m_i, m_j)\} \tag{2.1}$$

Then, in a Chinese-restaurant-process fashion, they probabilistically link together mentions based purely on the scores provided by this similarity function. That is, the value of $f(m_i, m_j)$ is directly correlated with the probability of (m_i, m_j) being chosen as a linked pair. Then, identical to Bengtson's and Roth's work [3], the HDDCRP model forms clusters by tracing through all linked pairs. All mentions that are reachable by a continuous path become assigned the same cluster. This hinges on the transitive property of coreference. For example, if $(m_1, m_3), (m_3, m_5)$ and (m_5, m_6) are each individually linked via the scoring function, then a cluster C_i is formed, where $C_i = \{m_1, m_3, m_5, m_6\}$, even though (m_1, m_5) or (m_3, m_6) may have had very low similarity scores. We aim to improve this shortcoming, as detailed in Section **??**.

Neural Iteratively-Unfolding Model

Recently, Choubey and Huang [6] introduced the first neural model for event coreference. Their system also fits into the mention-pair paradigm, whereby mentions are predicted by a feed-forward network. The authors asserted that when using the ECB+ corpus, within-doc coreference did not benefit from using mention context, which is an important finding. However, similar to the weakness of the HDDCRP model, they merge clusters which contain *any* mention-pair whose predicted score is below a given threshold, independent of mentions' relation to the cluster at large.

Bibliography

[1] James Allan, Jaime Carbonell, George Doddington, Jonathan Yamron, Yiming Yang, James Allan Umass, Brian Archibald Cmu, Doug Beeferman Cmu, Adam Berger Cmu, Ralf Brown Cmu, Ira Carp Dragon, George Doddington Darpa, Alex Hauptmann Cmu, John Lafferty Cmu, Victor Lavrenko Umass, Xin Liu Cmu, Steve Lowe Dragon, Paul Van Mulbregt Dragon, Ron Papka Umass, Thomas Pierce Cmu, Jay Ponte Umass, and Mike Scudder Umass. Topic detection and tracking pilot study final report. In *In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pages 194–218, 1998. Page 3.

[2] Amit Bagga and Breck Baldwin. Cross-document event coreference: Annotations, experiments, and observations. In *Proceedings of the Workshop on Coreference and Its Applications*, CorefApp '99, pages 1–8, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics. Page 7.

[3] Eric Bengtson and Dan Roth. Understanding the value of features for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 294–303, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. Page 8.

[4] Daniel M. Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. Nymble: a high-performance learning name-finder. In *In Proceedings of the Fifth Conference on Applied Natural Language Processing*, pages 194–201, 1997. Page 5.

[5] Jason P. C. Chiu and Eric Nichols. Named entity recognition with bidirectional lstm-cnns. *TACL*, 4:357–370, 2016. Page 5.

[6] Prafulla Kumar Choubey and Ruihong Huang. Event coreference resolution by iteratively unfolding inter-dependencies among events. In *EMNLP*, 2017. Page 8.

[7] Kevin Clark and Christopher D. Manning. Deep reinforcement learning for mention-ranking coreference models. *CoRR*, abs/1609.08667, 2016. Page 7.

[8] Kevin Clark and Christopher D. Manning. Improving coreference resolution by learning entity-level distributed representations, 2016. cite arxiv:1606.01323Comment: Accepted for publication at the Association for Computational Linguistics (ACL), 2016. Pages 6 and 7.

[9] Agata Cybulska and Piek Vossen. Using a sledgehammer to crack a nut? lexical diversity and event coreference resolution. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. European Language Resources Association (ELRA), 2014. Page 8.

[10] Naomi Daniel, Dragomir Radev, and Timothy Allison. Sub-event based multi-document summarization. In *Proceedings of the HLT-NAACL 03 on Text Summarization Workshop - Volume 5*, HLT-NAACL-DUC '03, pages 9–16, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. Page 3.

[11] Greg Durrett and Dan Klein. Easy victories and uphill battles in coreference resolution. In *EMNLP*, pages 1971–1982. ACL, 2013. Page 6.

[12] Xiaocheng Feng, Lifu Huang, Duyu Tang, Heng Ji, Bing Qin, and Ting Liu. A language-independent neural network for event detection. In *ACL*, 2016. Page 5.

[13] Daniel Gildea and Daniel Jurafsky. Automatic labeling of semantic roles. *Comput. Linguist.*, 28(3):245–288, September 2002. Page 5.

[14] Ralph Grishman and Beth Sundheim. Message understanding conference-6: A brief history. In *Proceedings of the 16th Conference on Computational Linguistics - Volume 1*, COLING '96, pages 466–471, Stroudsburg, PA, USA, 1996. Association for Computational Linguistics. Page 5.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997. Page 5.

[16] Eduard H. Hovy, Teruko Mitamura, Felisa Verdejo, Jun Araki, and Andrew Philpot. Events are not simple: Identity, non-identity, and quasi-identity. In *ACL*, 2013. Page 4.

[17] Kevin Humphreys, Robert Gaizauskas, and Saliha Azzam. Event coreference for information extraction. In *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, ANARESOLUTION '97, pages 75–81, Stroudsburg, PA, USA, 1997. Association for Computational Linguistics. Pages 3 and 7.

[18] Richard Johansson and Pierre Nugues. Lth: Semantic structure extraction using nonprojective dependency trees. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 227–230, Stroudsburg, PA, USA, 2007. Association for Computational Linguistics. Page 5.

[19] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197. Association for Computational Linguistics, 2017. Page 7.

[20] Xuezhe Ma and Eduard Hovy. End-to-end sequence labeling via bi-directional lstm-cnns-crf. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany, August 2016. Association for Computational Linguistics. Page 5.

[21] Andrew McCallum and Wei Li. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, CONLL '03, pages 188–191, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics. Page 5.

[22] David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1):3–26, January 2007. Publisher: John Benjamins Publishing Company. Page 5.

[23] Srini Narayanan and Sanda Harabagiu. Question answering based on semantic structures. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. Page 3.

[24] Vasin Punyakanok, Dan Roth, and Wen-tau Yih. The importance of syntactic parsing and inference in semantic role labeling. *Comput. Linguist.*, 34(2):257–287, June 2008. Page 5.

[25] W.V. O. Quine. Events and reification. In *Action and Events: Perspectives on the philosophy of Donald Davidson*, pages 162–171, 1985. Page 4.

[26] L. F. Rau. Extracting company names from text. In *Proc. of the Seventh Conference on Artificial Intelligence Applications CAIA-91 (Volume II: Visuals)*, pages 189–194, Miami Beach, FL, 1991. Page 5.

[27] Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. *Comput. Linguist.*, 27(4):521–544, December 2001. Page 6.

[28] Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. Joint learning improves semantic role labeling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 589–596, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. Page 5.

[29] Shyam Upadhyay, Nitish Gupta, Christos Christodoulopoulos, and Dan Roth. Revisiting the evaluation for cross document event coreference. In *COLING*, 2016. Page 4.

[30] Sam Wiseman, Alexander M. Rush, and Stuart M. Shieber. Learning global features for coreference resolution. In *HLT-NAACL*, pages 994–1004. The Association for Computational Linguistics, 2016. Pages 6 and 7.

[31] Sam Wiseman, Alexander M. Rush, Stuart M. Shieber, and Jason Weston. Learning anaphoricity and antecedent ranking features for coreference resolution. In *ACL (1)*, pages 1416–1426. The Association for Computer Linguistics, 2015. Pages 5 and 7.

[32] Travis Wolfe, Mark Dredze, and Benjamin Van Durme. Predicate argument alignment using a global coherence model. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings*, 2015. Page 4.

[33] Bishan Yang, Claire Cardie, and Peter I. Frazier. A hierarchical distance-dependent bayesian model for event coreference resolution. *TACL*, 3:517–528, 2015. Page 8.