The More Antecedents, the Merrier: Resolving Multi-Antecedent Anaphors

Hardik Vala, Andrew Piper, Derek Ruths
hardik.vala@mail.mcgill.ca, {andrew.piper, derek.ruths}@mcgill.ca

Introduction

One sub-problem of anaphor resolution has been largely untouched by prior work: the anaphor that has multiple antecedents, which we call *multi-antecedent anaphors* or *m*-anaphors (sometimes referred to in the literature as *split-antecedent* anaphors). For example,

Elizabeth *met* ***Mary*** *at the park and* ***they*** *began their stroll to the river.*

To avoid complexity, state-of-the-art coreference resolvers restrict anaphors to at most a single antecedent. Relaxing this constraint would pose serious problems in coreference chain-building, where each chain is intended to refer to a single entity. Moreover, multi-antecedent cases present a significant challenge given certain features well-suited for the single antecedent case do not apply (e.g. gender).

This work addresses multi-antecedent anaphors in NP anaphor resolution. While we frame the general question of multi-antecedent inference, we restrict our analyses to resolving the antecedents of the pronouns *they* and *them*. These pronouns best isolate the characteristics of *m*-anaphors (see the Scope section for details).

Contributions

1. **A generalization of the anaphor resolution problem to permit linking to multiple antecedents.**
2. **Preliminary insights into multi-antecedent anaphors based on their behaviour in linguistic environments.** (See the paper.)
3. **An entity-centric system for specifically resolving *m*-anaphors that outperforms a number of baseline methods.**
4. **A pairing of the proposed system with an existing coreference resolution system for the complete coreference resolution task, showing a gain of 0.6 points (CoNLL F1).**

Terminology

***m*-anaphor:** A special case of anaphor that links to multiple antecedents.

***m*-antecedent:** One of multiple antecedents of a particular *m*-anaphor.

siblings: Two or more *m*-antecedents linking to the same *m*-anaphor.

k-anaphor: An anaphor linking to exactly k antecedents (e.g. a 2-anaphor links to exactly 2 antecedents).

Elizabeth *met* ***Mary*** *at the park and* ***they*** *began their stroll to the river.*
m-antecedents that are siblings — *m*-anaphor that is a 2-anaphor

Problem Definition

We define the general NP anaphor resolution problem to account for *m*-anaphors as follows: Let $\mathcal{M}$ denote the set of all identified mentions in a document and let $M(x) \subseteq \mathcal{M}$ denote all mentions preceding a mention $x \ni \mathcal{M}$. The objective of the task is, for each $x \ni \mathcal{M}$, to find $C \subseteq \mathcal{M}$ such that all mentions in C are antecedent to x. If $C = \varnothing$, then x is non-anaphoric and if $|C| \geq 1$, then x is 1-anaphoric, and if $|C| > 1$, then x is *m*-anaphoric.

Scope

To constrain the scope of the study, we perform all our analyses on gold mentions, leaving the effect of imperfect mention detection as a problem for future work. Moreover, we only consider mentions of *they* and *them* that are known to be *m*-anaphoric for three reasons:

1. Non-pronomial *m*-anaphors, i.e. proper and common nouns, are much more susceptible to long-distance effects and may require external knowledge to resolve.
2. A host of very involved aspects of the complete *m*-anaphor resolution problem are circumvented, most notably, determining whether a mention is *m*-anaphoric, 1-anaphoric, or not anaphoric at all. For example, *you* may refer to one person or multiple, *who* can be used as an interrogative (non-anaphoric) or reflexive pronoun (anaphoric), pronouns such as *anyone* and *everyone* introduce many scoping difficulties, and pleonastic pronouns must be removed from the inference task entirely.
3. *they* and *them* are the most prevalent pronouns in our dataset.

Pronoun	# *m*-anaphors
they	278
them	165
we	140
you	43
everybody	12

(Counts of the most frequent *m*-anaphoric pronouns in P&P)

Method

#1 Identify candidate mentions.

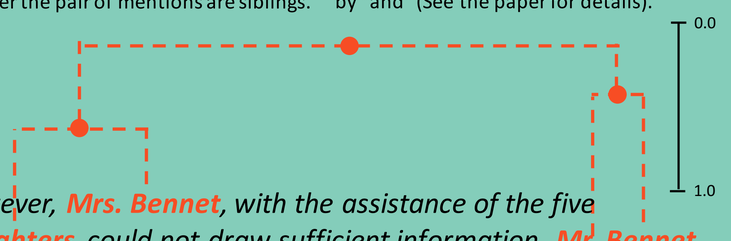
However, ***Mrs. Bennet****, with the assistance of the five* ***daughters****, could not draw sufficient information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.*

They *attacked him in various ways…*
***m*-anaphor to resolve** — **Candidate mentions**

#2 Perform agglomerative clustering using avg. linkage and similarity metric, $\sigma(\mathbf{w}^T\mathbf{x})$.

Weight vector learned using the standard cross-entropy loss function (with L2-regularization) in a maximum entropy model, where the decision variable is whether the pair of mentions are siblings.

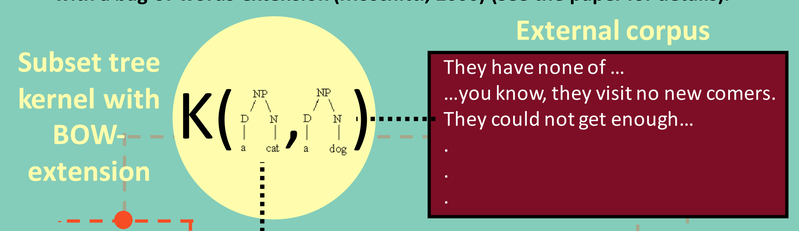
Feature vector defined over a pair of mentions, including morphosyntactic, grammatical, and semantic features, such as head match, word distance, and coordination by "and" (See the paper for details).



However, ***Mrs. Bennet****, with the assistance of the five* ***daughters****, could not draw sufficient information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.* ***They*** *attacked him in various ways…*

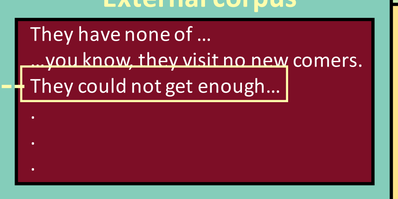
#3 Score each non-singleton cluster according to the prob. of coreference with the *m*-anaphor.

#3.1 Score the sentence(s) containing the cluster to each sentence containing *they* or *them* from an external corpus using a subset tree kernel (Collins and Duffy, 2002) with a bag-of-words-extension (Moschitti, 2006) (See the paper for details).



However, ***Mrs. Bennet****, with the assistance of the five* ***daughters****, could not draw sufficient information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.* ***They*** *attacked him in various ways…*

#3.2 Replace the sentence(s) containing the cluster with the sentence from the external corpus with the highest similarity.



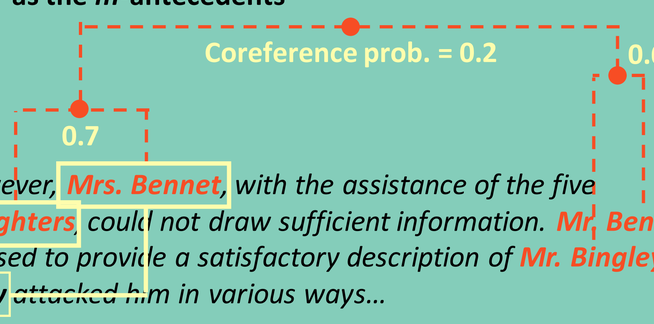
They *could not get enough information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.* ***They*** *attacked him in various ways…*

#3.3 Calculate the prob. of coreference between the *m*-anaphor and the counterpart *they* or *them* in the new sentence, using the classification mention-pair model described in Clark and Manning (2015).

Calculate coreference prob.

They *could not get enough information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.* ***They*** *attacked him in various ways…*

#4 Predict the cluster yielding the highest coreference prob. as the *m*-antecedents



However, ***Mrs. Bennet****, with the assistance of the five* ***daughters****, could not draw sufficient information.* ***Mr. Bennet*** *refused to provide a satisfactory description of* ***Mr. Bingley****.* ***They*** *attacked him in various ways…*

Data

Our dataset comprises of,

- The ***Pride and Prejudice*** **novel (P&P)** (121440 words) with all mentions of character fully resolved to their antecedents, including mentions referencing multiple characters.
- 36 short stories from the ***Scribner Anthology of Contemporary Short Fiction*** (Martone et al., 1999) **(Scribner)** (total of 216901 words), representing an eclectic collection of stories from the modern era. All mentions of *they* and *them* have been resolved (*m*-anaphoric, 1-anaphoric, and singleton), including those of non-person entities.

	they		*them*		Total	
	#	%	#	%	#	%
P&P	278	31.20	165	19.05	443	51.15
Scribner	243	12.96	79	4.21	322	17.17
Total	521	19.01	244	8.90	765	27.91

(# of *m*-anaphoric *they* and *them* mentions and % of all *they* and *them* mentions that are *m*-anaphors.)

These texts were annotated by three annotators and the inter-annotator agreement on the shared portion was 86.5%. Moreover, the dataset is partitioned according to a roughly, 60/20/20 split into training, validation, and testing sets.

Literary works were chosen over other textual modalities, e.g. news articles, because they showed a higher density of *m*-anaphors (a preliminary annotation exercise showed that literary works contained 37% more m-anaphors per word).

The external corpus was built from texts comparable to our dataset. 651,108 sentences containing one of *they* or *them* were mined from a larger corpus of 798 literary texts spanning the nineteenth and twentieth centuries.

Evaluation

Accuracy is measured in terms of the number of mention pairs correctly grouped as *m*-antecedents for a given *m*-anaphor.

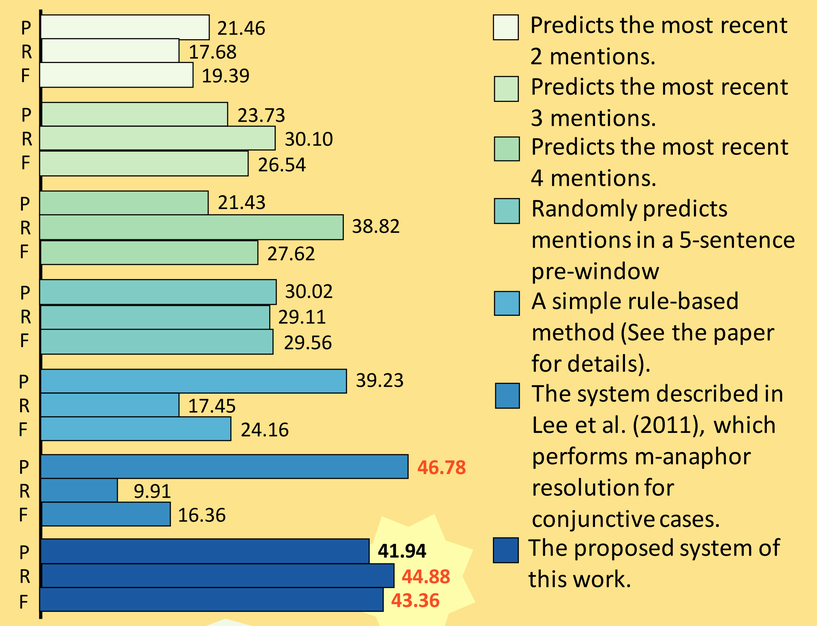
Let $n_1, n_2, \ldots, n_N$ represent the number of gold *m*-antecedents for *m*-anaphors $g_1, g_2, \ldots, g_N$ in a document, and $m_1, m_2, \ldots, m_N$ are predicted, of which $k_1, k_2, \ldots, k_N$ are correct.

Precision $= \Sigma_i k_i / \Sigma_i m_i$
Recall $= \Sigma_i k_i / \Sigma_i n_i$

Experiments

#1 System Comparison

Test set performance of each system on the m-anaphor resolution task.



The proposed system outperforms all other systems, but exhibits a bias towards 2-anaphors, recent mentions, and mentions coordinated by conjunction. This is not surprising given such cases are easiest to resolve.

#2 Full Coreference Resolution

The proposed system is integrated with the coreference resolution system of Clark and Manning (2015), and its prediction threshold raised to 0.89, at which point the precision on the validation set is 78.9. The Clark and Manning (2015) system is first run over the test set, producing coreference chains which are then filtered for character entities using the approach of Vala et al. (2015). Our adjusted system is then applied over all *they* and *them* mentions. Each such mention predicted as *m*-anaphoric is added to the coreference chains of the entities corresponding to the *m*-antecedent mentions. To evaluate the accuracy against the gold mention clusters, each *m*-anaphoric *they* and *them* is added to each cluster containing a gold *m*-antecedent.

	MUC	B^3	$CEAF_e$	Avg.
CLARK	42.3	39.5	32.4	38.1
CLARK + PROPOSED	43.4	40.0	**31.9**	38.7

(CoNLL metric scores for coreference resolution on the test portion of P&P for the Clark and Manning (2015) system), with (CLARK + PROPOSED) and without (CLARK) the pairing with the proposed system.

References

Clark, Kevin, and Christopher D. Manning. "Entity-centric coreference resolution with model stacking." *Association of Computational Linguistics (ACL)*. 2015.

Collins, Michael, and Nigel Duffy. "New ranking algorithms for parsing and tagging: Kernels over discrete structures, and the voted perceptron."*Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002.

Lee, Heeyoung, et al. "Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task." *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*. Association for Computational Linguistics, 2011.

Michael Martone, Lex Williford, and Rosellen Brown. 1999. *The Scribner Anthology of Contemporary Short Fiction: Fifty North American Stories Since 1970*. Touchstone.

Alessandro Moschitti. 2006. Making tree kernels practical for natural language learning. In *EACL*, volume 113, page 24.

Hardik Vala, David Jurgens, Andrew Piper, and Derek Ruths. 2015. Mr. bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 769–774.