

Gendered Pronoun Resolution

Dhruv Patel
IISc, Bangalore
dhruvpatel@iisc

Pratik Sachan
IISc, Bangalore
pratiksachan@iisc

Abstract

Abstract here.

1 Datasets and Metrics

1.1 Datasets

We have used two datasets in our experiments.

1. **GAP** coreference dataset([Webster et al., 2018](#)) is gender-balanced dataset divided into three sets for development, test and validation. Both development and test sets contain 2000 sentences each. Validation dataset contains 454 sentences. In each sentence there are two possible candidates denoted by A and B respectively. There is one pronoun per sentence. This pronoun can refer to either A or B or neither. Below is an example sentence from dataset. Bold italic is a pronoun. Bold words are candidate pronouns. Underlined noun is correct pronoun.

Kathleen first appears when **Theresa** visits *her* in a prison in London.

2. **DPR** Definite Pronoun Resolution ([Rahman and Ng, 2012](#)) dataset is divided into two sets for training and testing. There are 1886 sentences in total. Although the original dataset has only one candidate per sentence, there are two sentences having same actors in common(i.e. there are 943 pairs of sentences). So we combined two actors to play as candidates. The resulting dataset is similar to GAP. Below is an example pair.

- James asked **Robert** for a favor, but *he* refused.
- **James** asked Robert for a favor, but *he* was refused.

We have used 2000 sentences from GAP to train, while keeping others aside for validation and test. When we used DPR in addition to GPR, we used both train and test sets for training. Validation in this case was still done on GAP validation set.

1.2 Data Augmentation

Earlier we tried our models without any data augmentation. But since GAP has only 2000 sentences in development set, our models overfitted. An SVM trained on these 2000 sentences outperformed neural network architectures that we tried. To compare our later modifications, we will use SVM as a baseline.

To mitigate the situation we applied data augmentation. Our hypothesis is that, since input to our network is just a pair of candidate nouns, it doesn't matter what these nouns are. If all occurrences of noun 'Firstname Lastname' were to be replaced by some other plausible pair of first name and last names, sentence should make perfect sense. To neural network "Jon Snow doesn't know anything." should be similar to "John Wick doesn't know anything."

To augment data we applied simple rule. If both candidate A and candidate B had less than four words then with probability 0.7 we would pick random noun with same number of words. That is if A had three words and B had two words, we would pick alternative A and B with three words and two words respectively. If pronoun is male then only male names are proposed as alternatives. Alternative name for B was chosen such that no word of it was a substring of alternative A. Also none of the alternatives had any overlap with original nouns. Below are the examples of augmented sentences. First sentence is an original, other are augmented. Here Margaret Ray is candidate A and Betsy is candidate B.

- Tony Markham, a high school senior and the “Tall Dark Stranger” Betsy fell in love with as a freshman, who has since become a good friend not only to Betsy but the entire Ray family. Mrs. Ray, Betsy’s mother. Mr. Ray, Betsy’s father, who owns a shoe-store. **Margaret Ray**, Betsy’s sister who is five years younger than she is.
- Tony Markham, a high school senior and the “Tall Dark Stranger” Booth fell in love with as a freshman, who has since become a good friend not only to Booth but the entire Delgado family. Mrs. Delgado, Booth’s mother. Mr. Delgado, Booth’s father, who owns a shoestore. Pam Delgado, Booth’s sister who is five years younger than she is.
- Tony Markham, a high school senior and the “Tall Dark Stranger” Alyssa fell in love with as a freshman, who has since become a good friend not only to Alyssa but the entire Jolie family. Mrs. Jolie, Alyssa’s mother. Mr. Jolie, Alyssa’s father, who owns a shoe-store. Angelina Jolie, Alyssa’s sister who is five years younger than she is.

The pool for alternative names was extracted from dataset for stage2 of Kaggle competition. It has around 12K sentences. Figure 1 shows, distribution of names for both genders. One word names are most common in dataset followed by two word names.

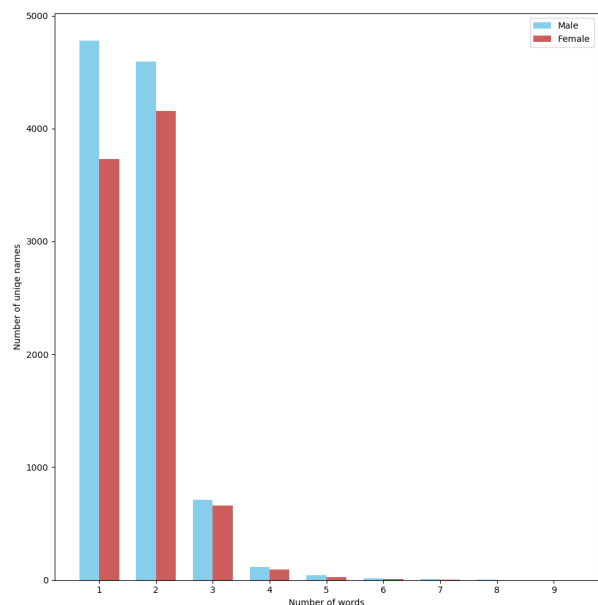


Figure 1: Distribution of names.

References

- Altaf Rahman and Vincent Ng. 2012. Resolving complex cases of definite pronouns: the winograd schema challenge. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 777–789. Association for Computational Linguistics.
- Kellie Webster, Marta Recasens, Vera Axelrod, and Jason Baldridge. 2018. Mind the gap: A balanced corpus of gendered ambiguous. In *Transactions of the ACL*, page to appear.