Hierarchical Attention Network with Pairwise Loss for Chinese Zero Pronoun Resolution

Peiqin Lin (linpq3@mail2.sysu.edu.cn) and Meng Yang (yangm6@mail.sysu.edu.cn)
School of Data and Computer Science, Sun Yat-sen University

Task Description

A zero pronoun refers to the component that is omitted because of the coherence of language. A zero pronoun can be an anaphoric zero pronoun (AZP) if it corefers to one or more mentions in the associated text, or a non-anaphoric one if there are no such mentions. In the example, *pro*₁ is anaphoric and corefers to the mention “The police”, while *pro*₂ is non-anaphoric.

Motivation

- Previous methods either did not consider any interaction between zero pronouns and candidate antecedents (Chen and Ng 2016; Yin et al. 2018a) or just employed unidirectional attention from the representations of zero pronouns to candidate antecedents (Liu et al. 2016; Yin et al. 2018b), weakening the representation ability of the learned features.
- Previous methods simply formulate the resolution task as a classification task (e.g., whether a candidate is the antecedent of a zero pronoun), which ignores the relationship between different candidates of a zero pronoun (e.g., the correct candidates are similar and their scores should be larger than those of wrong candidates by a large margin).

Example:
[警方] 怀疑 这是 一起 黑枪 案件，*pro*₁将 枪械 和 皮包 交送 市里, *pro*₂以 清理 案情
[The police] suspected that this is a criminal case about illegal guns, *pro*₁ brought the guns and bags to the city *pro*₂ to deal with the case.
Hierarchical Attention Network with Pairwise Loss for Chinese Zero Pronoun Resolution

Peiqin Lin (linpq3@mail2.sysu.edu.cn) and Meng Yang (yangm6@mail.sysu.edu.cn)
School of Data and Computer Science, Sun Yat-sen University

Our Model (HAN-PL)

Hierarchical Attention Network (HAN) employs interactive attention and self attention to better model zero pronouns and candidate antecedents.

Pairwise Loss (PL), which is based on a pairwise-margin loss and a similarity constraint, instead of cross entropy loss used in previous methods, is designed to guide the optimization of the model.

- We take each correct antecedent and each wrong antecedent in the candidate set as a pair, and then compute the pairwise-margin loss between them;
- We take correct antecedents in pair, and then design a similarity constraint for better training the model.

Note: Peiqin Lin is applying Ph.D. now, more information in lpq29743.github.io