Towards Integration of Discriminability and Robustness
For Document-Level Relation Extraction

Jia Guo1,2, Stanley Kok1, Lidong Bing2
1School of Computing, National University of Singapore
2DAMO Academy, Alibaba Group

Introduction

Sentence-level relation extraction
- Single sentence input
- Single entity mention
- Single entity pair
- Single label output

Document-level relation extraction
- Multiple sentence inputs
- Multiple entity mentions
- Multiple entity pairs
- Multiple label outputs

Challenges:
- Inadequate in effectively distinguishing relations
- Lack of sufficient learning for long-tailed relations
- Vulnerable to annotation errors or missing annotations

Our contributions:
- A novel model based on Pairwise moving-threshold loss, Entropy Minimization, and Supervised Contrastive Learning (PEMSCL)

Method

Our PEMSCL model
1. Pairwise moving-threshold loss with Entropy Minimization
2. Supervised Contrastive Learning for multi-labels and long-tailed relations
3. Negative label sampling strategy

Problem definition:
Inputs:
- A document: \( D = \{ w_i \}_{i=1} \)
- A set of entities
- Each entity \( e_i \) is associated with a set of mentions \( \mathcal{M}_e = \{ m_{i,j} \}_{j=1} \)

Outputs:
- For each entity pair \((e_i, e_j)\), the model predicts a subset of pre-defined relations \( R = \{ r_{i,j} \}_{i,j=1} \)
- If an entity pair does not express any relation, it is labeled as NA

Pairwise moving-threshold loss with entropy minimization
- The definition of information entropy:

Supervised contrastive learning for multi-labels & long-tailed relations
- The logit difference between positive and negative examples:

Experiments and Analysis

Benchmarks:
- DocRED [Tan et al., ACL’19] & Re-DocRED [Tan et al., EMNLP’22] (both \( |R| = 96 \))
- Two new data regimes:
  - OOG-DocRE / OOG-DocRE:
    - Original labels for the train set
    - Gold labels for the dev set
  - Modified as:

Our PEMSCL outperforms previous strong baselines on the original DocRED dataset and its cleaned version, the Re-DocRED dataset.

Our negative label sampling strategy is effective and robust in the noisy settings.