Introduction

- Social annotation, or tagging, is a popular functionality allowing users to assign "keywords" to online resources for better semantic search and recommendation. In practice, however, only a limited number of resources is annotated with tags.
- We propose a novel deep learning architecture for automated social text annotation with cleaned user-generated tags.

Research Questions

- How to model the impact of the title on social annotation? (see Title-Guided Attention Mechanisms)
- How to leverage both similarity and subsumption relationships among labels in neural networks to further improve the performance of multi-label classification? (see Semantic-Based Loss Regularizers)

Title-Guided Attention Mechanisms

Word-level attention mechanisms (for the title) [3-4]:

\[ c_t = \sum_{i} \alpha_i h_t = \sum_{i} \frac{\exp(v_{wt} \cdot v_i)}{\sum_{i} \exp(v_{wt} \cdot v_i)} h_i \]

\[ v_t = \tanh(W_h h_t + b_h) \]

Similarly we can obtain \( c_t \) (sentence representation) and \( c_a \) (content representation based on the original sentence-level attention mechanism in [3-4]).

**Title-guided** sentence-level attention mechanisms:

\[ c_a = \sum_{i} \alpha_i h_a = \sum_{i} \frac{\exp(v_{wa} \cdot v_i)}{\sum_{i} \exp(v_{wa} \cdot v_i)} h_i \]

\[ v_a = \tanh(W_h h_a + b_a) \]

\( h_t \) and \( h_a \) denote the hidden state of word and sentence, respectively. The \( W_t, W_s, b_t, b_a \) are weights to be learned in training.

\( v_{wt}, v_{wa} \) and \( v_{wa} \) are global context vectors, i.e. "what is the informative word [or sentence]" to be learned.

The final document representation is the concatenation of the title and the content representation.

\[ c_d = [c_t, c_a, c_a] \]

JMAN (Joint Multi-Label Attention Network)

The automated social text annotation task can be formally transformed into a multi-label classification problem.

**Semantic-based Loss Regularizers**

Users tend to annotate documents collectively with tags of various semantic forms and granularities.

The whole joint loss to optimize:

\[ L = \frac{1}{2} \sum_{j \in T} \sum_{i \in S} \text{Sim}(s_i, s_j) (1 - R(s_i, s_j)) \]

\( L_{sim} \) constrains similar labels to have similar outputs.

\( L_{sub} \) enforces each co-occurring subsumption pair to satisfy the dependency of the parent label on the child label.

\( \text{Sim} \in (0, 1) \) is a pre-computed label similarity matrix based on embeddings pre-trained from the label sets.

\( \text{Sim} \) can be obtained by grounding labels to knowledge bases (e.g. Microsoft Concept Graph, for the Bibsonomy dataset) or from crowdsourced relations (for the Zhihu dataset).

\( R(\cdot) \) is a rounding function. \( R(S_{ij}) = 1 \) when \( S_{ij} \geq 0.5 \), otherwise \( R(S_{ij}) = 0 \).

Conclusions & Future Studies

- Experiments show the effectiveness of JMAN with superior performance and training speed over the state-of-the-art models, HAN and Bi-GRU.
- It is worth to explore other types of guided attention mechanisms and to adapt the regularizers to pre-trained transferable models like BERT.

References


Results

Attention visualization of a document in Bibsonomy with the JMAN model: purple blocks show word-level attention weights; red blocks in "or" (original) and "tg" (title-guided) show sentence-level attention weights. Predicted labels and ground truth labels are also presented.