Our code and datasets are available at https://github. com/acadTags/Automated-Social-Annotation

Joint Multi-Label Attention Networks for Social Text Annotation

UNIVERSITY OF LIVERPOOL

Hang Dong[1,2], Wei Wang[2], Kaizhu Huang[3], Frans Coenen[1],

1. University of Liverpool; 2. Xi'an Jiaotong-Liverpool University

Xi'an Jiaotong-Liverpool University 西交利物浦大学

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* relations 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]:

$$c_t = \sum_i \alpha_i h_i = \sum_i \frac{\exp(v_{wt} \bullet v_i)}{\sum_j \exp(v_{wt} \bullet v_j)} h_i$$

$$v_i = \tanh(W_t h_i + b_t)$$

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

Title-guided sentence-level attention mechanisms:

$$c_{ta} = \sum_r \alpha_r h_r = \sum_r \frac{\exp(c_t \bullet v_r)}{\sum_k \exp(c_t \bullet v_k)} h_r$$

$$v_r = \tanh(W_s h_r + b_s)$$

h_i and h_r denote the hidden state of word and sentence, respectively; The W_t, W_s, b_t, b_s are weights to be learned in training.

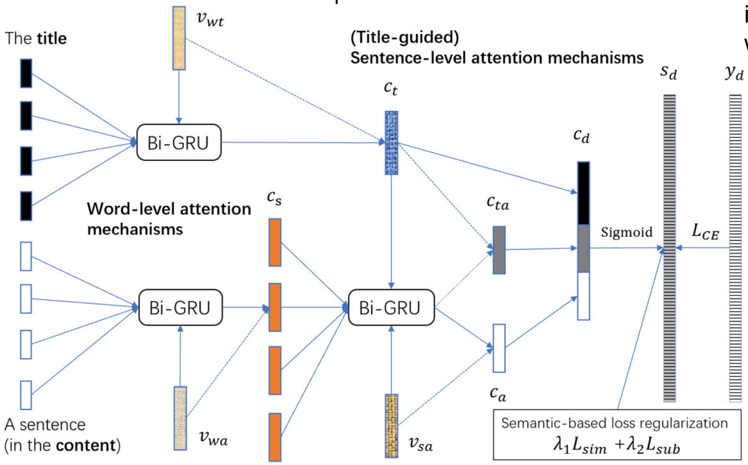
v_{wt}, v_{wa} and v_{wa} are global context vectors, i.e. "what is the informative word [or sentence]" to be learned.

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

$$c_d = [c_t, c_{ta}, c_a]$$

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: $L = L_{CE} + \lambda_1 L_{sim} + \lambda_2 L_{sub}$

$$L_{sim} = \frac{1}{2} \sum_d \sum_{(j,k)|T_j,T_k \in y_d} Sim_{jk} |s_{dj} - s_{dk}|^2$$

$$L_{sub} = \frac{1}{2} \sum_d \sum_{(j,k)|T_j,T_k \in y_d} Sub_{jk} R(s_{dj})(1 - R(s_{dk}))$$

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.

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

$Sub \in \{0,1\}^{|T|*|T|}$ can be obtained by grounding labels to knowledge bases (e.g. Microsoft Concept Graph, for the Bibsonomy dataset) or from crowd-sourced relations (for the Zhihu dataset).

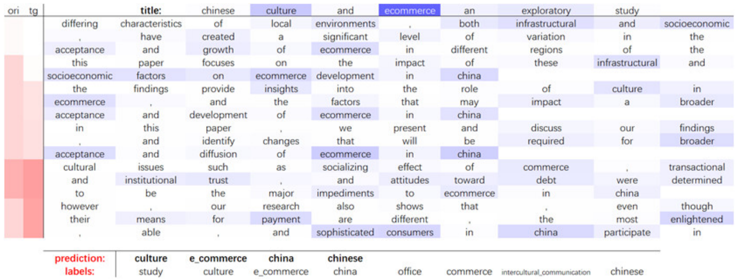
$R()$ is rounding function, $R(S_{dj}) = 1$ when $S\_dj \geq 0.5$, otherwise $R(S_{dj}) = 0$.

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.

Results

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



Dataset	$\|X\|$	$\|Y\|$	$\|V\|$	Ave	Σ_{Sub}
Bibsonomy (clean)	12,101	5,196	17,619	11.59	101,084
Zhihu (sample)	108,168	1,999	62,519	2.45	2,655

$|X|$, document size; $|Y|$, label size; $|V|$, vocabulary size; Ave, average number of labels per document; $\sum Sub$, number of subsumption relations.

Bibsonomy	Precision	Recall	F_1 Score	Time/Fold
Bi-GRU	.522±.020*	.217±.016*	.306±.019*	1480±92s
HAN	.572±.008*	.246±.012*	.344±.013*	1164±52s
JMAN-s-tg	.591±.010	.269±.006*	.370±.007*	1075±87s
JMAN-s-att	.586±.009	.269±.005*	.369±.006*	968±81s
JMAN-s	.586±.004	.282±.005	.380±.005	**894±55s**
JMAN	**.592±.009**	**.284±.006**	**.384±.007**	1044±73s

* Paired t-tests at 95 percent significance level against the JMAN model.

Zhihu	Precision	Recall	F_1 Score	Time/Fold
Bi-GRU	.238±.011*	.154±.009*	.187±.010*	1455±69s
HAN	.257±.012	.167±.010*	.203±.011*	1387±78s
JMAN-s-tg	.257±.005	.175±.003*	.208±.006**	1220±81s
JMAN-s-att	.254±.007**	.174±.005*	.207±.005*	1275±99s
JMAN-s	.257±.008	.177±.005	.210±.007	1147±44s
JMAN	**.260±.006**	**.179±.003**	**.212±.004**	**1135±52s**

* Paired t-tests at 95 percent significance level against the JMAN model.
** Paired t-tests at 90 percent significance level against the JMAN model.

Baselines (tested with models from 10-fold cross-validation):
Bi-GRU: Bidirectional Gated Recurrent Unit [1-2].
HAN: Hierarchical Attention Network [3-4].
JMAN-s: without semantic-based loss regularisers.
JMAN-s-tg: without semantic-based loss regularisers. and the title-guided sentence-level attention mechanism.
JMAN-s-att: without semantic-based loss regularisers and the original sentence-level attention mechanism.

References

[1] Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.
[2] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734.
[3] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489.
[4] Hebatallah A. Mohamed Hassan, Giuseppe Sansonetti, Fabio Gasparetti, and Alessandro Micarelli. 2018. Semantic-based tag recommendation in scientific bookmarking systems. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, pages 465–469, New York, NY, USA. ACM.

HangDong@liverpool.ac.uk; Wei.Wang03@xjtlu.edu.cn; Kaizhu.Huang@xjtlu.edu.cn; Coenen@liverpool.ac.uk