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

Joint Multi-Label Attention Networks for Social Text Annotation



Hang Dong^{1,2}, Wei Wang², Kaizhu Huang³, Frans Coenen¹, 1. University of Liverpool; 2. Xi'an Jiaotong-Liverpool University



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]:

$$c_t = \sum_i \alpha_i h_i = \sum_i \frac{\exp(v_{wt} \bullet v_i)}{\sum_j \exp(v_{wt} \bullet v_j)} h$$
$$v_i = \tanh(W_t h_i + b_t)$$

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.

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

References

[1] Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11):2673–2681.



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

 $L_{sim} = \frac{1}{2} \sum \sum Sim_{jk} |s_{dj} - s_{dk}|^2$

$$\sum_{sim} = \frac{2}{2} \sum_{d} \sum_{(j,k)|T_j, T_k \in y_d} \sum_{j \in [j,k] \in y_d} \sum_{j \in y_d}$$

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

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 $S_d \ge 0.5$, 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.

| tg | | title: | chinese | culture | and | | an | exploratory | study | |
|--|------------------------|---|--|--|--|---|---|--|--|--|
| | differing | characteristics | of | local | environments | | both | infrastructural | and | socioeconomic |
| | | have | created | а | significant | level | of | variation | in | the |
| | acceptance | and | growth | of | ecommerce | in | different | regions | of | the |
| | this | paper | focuses | on | the | impact | of | these | infrastructural | and |
| | socioeconomic | factors | on | ecommerce | development | in | china | | | |
| | the | findings | provide | insights | into | the | role | of | culture | in |
| | ecommerce | | and | the | factors | that | may | impact | а | broader |
| | acceptance | and | development | of | ecommerce | in | china | | | |
| | in | this | paper | | we | present | and | discuss | our | findings |
| | | and | identify | changes | that | will | be | required | for | broader |
| | acceptance | and | diffusion | of | ecommerce | in | china | | | |
| | cultural | issues | such | as | socializing | effect | of | commerce | 1 | transactional |
| | and | institutional | trust | | and | attitudes | toward | debt | were | determined |
| | to | be | the | major | impediments | to | ecommerce | in | china | |
| | however | | our | research | also | shows | that | | even | though |
| | their | means | for | payment | are | different | | the | most | enlightened |
| | | able | , | and | sophisticated | consumers | in | china | participate | in |
| | prediction: labels: | culture study | e_commerce culture | china e_commerce | chinese china | office | commerce | intercultural_communication | chinese | |
| Dataset | | | X | Y | | V | Ave | | | |
| Bibsonomy (clean) | | | 12,101 5,196 | | 17,619 | 9 11.59 | 101 | ,084 | | |
| aceptance and growth the paper focuses socioeconomic factors on economic factors on acceptance and development in acceptance and development acceptance and development acceptance and development in institutional trust to be the however be the however be the the trust study prediction: culture ecommerce able culture Bibssonomy (clean) Zhihu (sample) | | | | 108,16 | 8 1,9 | 999 | 62,519 | 9 2.45 | 2,65 | 55 |
| | В | acceptance this sociececonomic the economerce acceptance cultural to however their prediction: labels: Dataset Bibsonori | Area acceptance, and this paper socie-accommic factors and insues and and acceptance and in this acceptance and instructional to however their acceptance study Dataset Bibsonomy (cle | Acceptance and growth bis sociaccontrol factors on the findings provide acceptance and development acceptance and development in this paper outural issues such and development acceptance and development outural issues such and diffusion to be the here their acceptance and diffusion acceptance and such acceptance and such acceptance and such acceptance acceptance such and issues such acceptance | acceptance and growth of of base constraints on the growth of acceptance and growth of acceptance and for the acceptance and development of accors and identify changes and identify changes and identify changes and institutional trust to be the major the acceptance and diffusion as the acceptance and diffusion are and institutional trust to be the major be the major be the and a light trust and the acceptance acceptance and diffusion as a such | note note rested a signifiant acceptance and growth of ecommerce scopercomic factors on ecommerce on scopercomic factors on ecommerce insights the findings provide insights into acceptance and development of ecommerce acceptance and identify changes that acceptance institutional trust as socializing and institutional trust and are are abels culture e_commerce chinal chinas prediction culture e_commerce chinas chinas Bibsonomy (clean) 12,101 5,7 | new created a significant level acceptance and growth of ecommerce infinitiant bth paper focuses on the intimate the findings provide insights into the the findings provide insights into the acceptance and development of ecommerce in acceptance and idevelopment of ecommerce in acceptance and institutionat tout as socializing aditioudes out areserch also shows predictor: cuture e.commerce chinate commerce able cuture e.comme | have created a significant level of different study acceptance and growth of ecommerce in the impact of focuses on the impact of different study acceptance and development in the findings provide insights into the forole commerce of a and the factors that may acceptance and development of ecommerce in different in this paper of the major insights with the role commerce and diffusion of ecommerce in different in this paper of the major impediments to acceptance and diffusion as socializing effect of the major impediments to and attruct to be the major impediments to and attruct to pay the study ecommerce chinas office commerce in the diffusion of ecommerce in the diffusion of the diffusi | n. have created a significant level of of variance acceptance and growth of ecommerce inglett offerent offerent offerent offerent these bh findings provide insights insights insights insights offerent officat offer | n have acceptance the scoreaction into acceptance and scoreaction into acceptance and scoreaction into acceptance scoreaction into scoreaction into s |

|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 | $.522 \pm .020^{*}$ | $.217 \pm .016^{*}$ | $.306 \pm .019^{*}$ | 1480±92s |
| HAN | $.572 \pm .008^{*}$ | $.246 \pm .012^{*}$ | $.344 \pm .013^{*}$ | 1164±52s |
| JMAN-s-tg | $.591 {\pm} .010$ | $.269 {\pm} .006^{*}$ | $.370 {\pm} .007^{*}$ | 1075±87s |
| JMAN-s-att | $.586 {\pm} .009$ | $.269 {\pm} .005^{*}$ | $.369 {\pm} .006^{*}$ | 968±81s |
| JMAN-s | $.586 {\pm} .004$ | $.282 {\pm} .005$ | $.380 {\pm} .005$ | 894±55s |
| JMAN | $.592 {\pm} .009$ | .284±.006 | .384±.007 | 1044±73s |

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

| r - G | | | | | | | |
|------------|----------------------|-----------------------|-----------------------|-----------|--|--|--|
| Zhihu | Precision | Recall | F_1 Score | Time/Fold | | | |
| Bi-GRU | $.238 \pm .011^{*}$ | $.154 \pm .009^{*}$ | $.187 {\pm} .010^{*}$ | 1455±69s | | | |
| HAN | $.257 {\pm} .012$ | $.167 {\pm} .010^{*}$ | $.203 {\pm} .011^{*}$ | 1387±78s | | | |
| JMAN-s-tg | $.257 {\pm} .005$ | $.175 \pm .003^{*}$ | .208±.006** | 1220±81s | | | |
| JMAN-s-att | $.254 \pm .007^{**}$ | $.174 \pm .005^{*}$ | $.207 {\pm} .005^{*}$ | 1275±99s | | | |
| JMAN-s | $.257 {\pm} .008$ | $.177 {\pm} .005$ | $.210 {\pm} .007$ | 1147±44s | | | |
| JMAN | $.260 {\pm} .006$ | $.179 {\pm} .003$ | $.212 \pm .004$ | 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: <u>without</u> semantic-based loss regularisers. and the titleguided sentence-level attention mechanism.

JMAN-s-att: <u>without</u> semantic-based loss regularisers and the original sentence-level attention mechanism.

[2] Kyunghyun Cho, Bart van Merrienbeer, 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] Zihao Yang, Divj Yang, Chris Dyer, Xlaodong He, Alex Smola, and Eduard Howy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference on Remain Conference o*