

Skill Inference with Personal and Skill Connections

Zhongqing Wang[†], Shoushan Li^{*‡}, Hanxiao Shi[‡], and Guodong Zhou[†]

[†] Natural Language Processing Lab, School of Computer Science and Technology,
Soochow University, China

[‡] School of Computer Science and Information Engineering,
Zhejiang Gongshang University, China

{wangzq.antony, shoushan.li}@gmail.com
hxshi@mail.zjgsu.edu.cn, gdzhou@suda.edu.cn

Abstract

Personal skill information on social media is at the core of many interesting applications. In this paper, we propose a factor graph based approach to automatically infer skills from personal profile incorporated with both personal and skill connections. We first extract personal connections with similar academic and business background (e.g. co-major, co-university, and co-corporation). We then extract skill connections between skills from the same person. To well integrate various kinds of connections, we propose a joint prediction factor graph (JPFG) model to collectively infer personal skills with help of personal connection factor, skill connection factor, besides the normal textual attributes. Evaluation on a large-scale dataset from LinkedIn.com validates the effectiveness of our approach.

1 Introduction

With the large amount of user-generated content (UGC) published online every day in the context of social networks (Tan et al., 2011; Luo et al., 2013), such online social networks (e.g., Twitter, Facebook, and LinkedIn) have significantly enlarged our social circles and much affected our everyday life. One popular and important type of UGC is the personal profile, where people post their detailed information, such as education, experience and other personal information, on online portals. Social websites like Facebook.com and LinkedIn.com have created a viable business as profile portals, with the popularity and success largely attributed to their comprehensive personal profiles.

Obviously, online personal profiles can help people connect with others of similar backgrounds and provide valuable resources for businesses, especially for personnel resource managers to find talents (Yang et al., 2011a; Guy et al., 2010). In the profiles, the personal skill information is the most important aspect to reflect the expertise of a person. However, few social platforms allow users to manually attach such personal skill information into their personal profiles. For example, in our collected dataset, 91.8% skills appear less than 10 times. Even the distribution of the top 10 frequently occurring skills is asymmetric, and only 43.1% people attach skills on their profiles. For this regard, it is highly desirable to develop reliable methods to automatically infer personal skills for personal profiles.

Although it is straightforward to recast skill inference as a standard text classification problem, i.e., predicting the skills with the profile text alone, personal profiles usually are poorly organized, even with critical information missing. Thus, it is challenging to infer skills given the limited information from the profile texts. We propose two assumptions to address above challenges by incorporating additional connection information between persons and skills:

- People are always connected to others with similar academic and business backgrounds (e.g. co-major, co-corporation). For example if there is co-major, co-university, or co-corporation relationship between two persons, it is very likely that they may share similar skills. Therefore, it is reasonable to resort to personal connection information to improve the performance of skill inference.

*corresponding author

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

- One person tends to have some related skills. For example, it is very likely that C++, C, and Python programming languages may co-occur in the one’s profile, i.e., if a person has skill C++, it is highly possible that he would have the skills such as C or Python. Thus, it is useful to integrate skill connection information when inferring personal skills.

Based on these assumptions, we propose a Joint Prediction Factor Graph (JPF_G) model, which collectively predicts personal skills with help of both personal and skill connections. In particular, the JPF_G model provides a general framework to integrate three kinds of knowledge, i.e. local textual attribute functions of an individual person, personal connection factors between persons, and skill connection factors between skills, in collectively inferring personal skills. Specially, we extract personal connections with similar academic and business background (e.g. co-major, co-corporation). We then extract skill connections between skills from same person. Evaluation on a large-scale data set from LinkedIn.com indicates that our JPF_G model can significantly improve the performance of personal skill inference.

The remainder of this paper is structured as follows. We review the related work in Section 2. In Section 3, we introduce the data collection. In Section 4, we give the problem definition and some analysis on the task of personal skill reference. In Section 5, we propose the JPF_G model and corresponding algorithms for parameter estimation and prediction. In Section 6, we present our experimental results. In Section 7, we summarize our work and discuss future directions.

2 Related Works

In this section, we briefly review related studies in expert finding, social tag suggestion and factor graph model.

2.1 Expert Finding

Expert finding aims to find right persons with appropriate skills or knowledge, i.e. "Who are the experts on topic X?" TREC-2005 and TREC-2006 have provided a common platform for researchers to empirically evaluate methods and techniques on expert finding (Soboroff et al, 2006; Zhang et al., 2007a).

In the literature, expert finding tends to consider each skill individually and seeks the most authority experts for each skill. Thus, expert finding is always considered as a ranking process, i.e., ranking the experts from the candidates who are most suitable for the skill (Balog and Rijke, 2007). For example, Campbell et al. (2003) investigated the issue of expert finding in an email network. They utilized the link between email authors and receivers to improve the expert finding performance.

Besides that link structure-based algorithms, such as PageRank and HITS, are employed to analyze the relationship of the link-relationship graph, social networks are utilized to improve the performance of expert finding. Zhang et al. (2007a) proposed a unified propagation-based approach to address the issue of expert finding in a social network, considering both personal local and network information (e.g. the relationship between persons).

Expert finding is in nature different from skill inference. Our study predicts various skills attachable to a person collectively with both personal and skill connections among people. One distinguishing characteristics of our study is that several skills from a person are simultaneously modeled and the relationship among these skills is fully leveraged in the inference.

2.2 Social Tag Suggestion

Social tag suggestion aims to extract proper tags from social media and can thus help people organize their information in an unconstrained manner (Ohkura et al., 2006; Si et al., 2010). Ohkura et al. (2006) created a multi-tagger to determine whether a particular tag from a candidate tag list should be attached to a weblog. Lappas et al. (2011) proposed a social endorsement-based approach to generate social tags from Twitter.com and Flickr.com where various kinds of information in recommendations and comments are used. Liu et al. (2012) propose a probabilistic model to connect the semantic relations between words and tags of microblog, and takes the social network structure as regularization. Li et al., (2012) propose to model context-aware relations of tags for suggestion by regarding resource content as context of tags.

Different from above researches, our study is forced on skill inference instead of traditional tag suggestion. Basically, the social connections in skill inference are much different from those in social tagging. In our study, we use co-major, co-title and other academic and business relationships to build the social connections. Meanwhile, there are also few researches concern to propose a joint model to leverage both personal and skill connections.

2.3 Factor Graph Model

Among various approaches investigated in social networks in the last several years (Leskovec et al., 2010; Lu et al., 2010; Lampos et al., 2013; Guo et al., 2013), Factor Graph Model (FGM) becomes an effective way to represent and optimize the relationship in social networks (Dong et al., 2012; Yang et al., 2012b) via a graph structure. Tang et al. (2011a) and Zhuang et al. (2012) formalized the problem of social relationship learning as a semi-supervised framework, and proposed Partially-labeled Pairwise Factor Graph Model (PLP-FGM) for inferring the types of social ties. Tang et al. (2013) further proposed a factor graph based distributed learning method to construct a conformity influence model and formalize the effects of social conformity in a probabilistic way.

Different from previous studies, this paper proposes a pairwise factor graph model to collectively infer personal skills with both social connection factor and skill connection factor.

3 Data Construction

We collect our data set from LinkedIn.com. It contains a large number of personal profiles generated by users, containing various kinds of information, such as personal Summary, Experience, Education, and Skills & Expertise. We do not collect personal names in public profiles to protect people’s privacy.

The dataset contains 7,381 personal profiles, among which only 3,182 profiles (43.1% of all the profiles) show the Skills & Expertise field. In this study, we adopt only these profiles in all our experiments. As a result, we get 6,863 skills in total, among which 6,299 skills (91.8% of them) appear less than 10 times. Among the remaining 564 skills, we select top 10 frequently occurring skills as the candidate personal skills in this study (Since the remaining 554 skills only appear less than 250 times in total, it is difficult to build an effective classifier for them). Table 1 illustrates the statistics.

Skill	Number	Ratio
Semiconductors	948	0.298
IC	369	0.116
Thin Films	328	0.103
Characterization	326	0.102
CMOS	311	0.098
Matlab	287	0.090
Microsoft Office	283	0.089
Manufacturing	278	0.087
Design of Experiments	262	0.082
Semiconductor Industry	250	0.079

Table 1: The distribution of the candidate personal skills

From Table 1, we can see that the skill distribution in the personal profiles is asymmetric. For example, the Semiconductor skill occurs about 1,000 times, taking 29.8%, while the Semiconductor Industry skill occurs 250 times only, taking 7.9%.

4 Problem Definition and Analysis

Before presenting our approach for skill inference, we first give the definition of the problem, and convey a series of discoveries we observed from the data.

4.1 Problem Definition

We first introduce some necessary definitions and then formulate of the problem.

Definition 1: Skill inference. In principle, we cast skill inference as a skill prediction problem. Since one person might have several skills, we build several vectors for a person and each vector is designed to determine whether the corresponding skill is appropriate for the person or not ("Positive" means that the person has the target skill, whereas "Negative" stands for the opposite). Note that the number of vectors for a person is equal to the number of candidate skills. For example, suppose we have m persons and n candidate skills in the dataset, we totally build vectors to represent if these skills are attached in these persons' profiles.

Definition 2: Textual information. We use texts of Summary and Experience as the textual information for our research. Texts of Summary and Experience are unstructured information, while texts of Skills & Expertise are structured information. However, some skills in the Skill & Expertise fields may not be mentioned in the Summary and Experience fields.

Definition 3: Personal connections. We can explicitly extract four kinds of personal relationships between two persons from the Education and Experience fields, as follows:

- *co_major*, which denotes that two persons have the same major at school
- *co_univ*, which denotes that two persons graduated from the same university
- *co_title*, which denotes that two persons have the same title in a corporation.
- *co_corp*, which denotes that two persons work in the same corporation.

Definition 4: Skill connections. We extract skill connections from same person. That is, if two vectors are from the same person with different skills, we consider these two vectors share skill connections (e.g. John has IC and Thin Films skills).

Learn task: Given the textual information of each profile, the personal connections between profiles, and skill connections of skill from same persons, the goal is to infer the skill through the above information.

To learn the skill inference model, there are several requirements. First, the skills of persons are related to multiple factors, e.g., network structure, personal connections, and skill connections, it is important to find a unified model which is able to incorporate all the information together. Second, the algorithm to learn the inference model should be efficient. In practice, the scale of the social network might be very large.

4.2 Statistics and Observations

In the following, we give some statistics and observations on personal and skill connections.

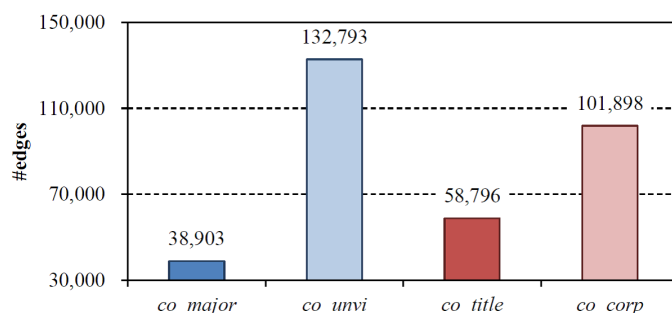


Figure 1: The statistic of personal connection edges in our dataset

Statistics of personal connections: Figure 1 gives the statistics of personal connection edges. It shows that with 3,182 profiles, there exist 332,390 personal connection edges. Besides, among all the

four relations, co_major, co_unvi, co_title, and co_corp occupy 11.7%, 40.0%, 17.7% and 30.6% respectively.

Observations of skills connections: To validate the tendency of a person sharing similar skills, we use PMI (Point-wise Mutual Information) to measure the co-occurrence between two skills. As a popular way to measure the co-occurrence between a pair (Turney, 2002), PMI is calculated as follows:

$$PMI(i, j) = \log \left(N \frac{P(i \& j)}{P(i)P(j)} \right) \quad (1)$$

N is the number of profiles, $P(i \& j)$ denotes the probability of the skills (i.e., i and j) co-occurrence in a person’s profile, while $P(i)$ denotes the probability of the skill i appearing in a person’s profile.

Skill i	Skill j	PMI
C	COMS	1.711
Thin Films	Characterization	1.624
Thin Films	Design of Experiments	1.543
Semiconductor Industry	IC	1.345
Semiconductor Industry	Design of Experiments	1.345
IC	Microsoft Office	-2.390
CMOS	Microsoft Office	-2.627
Semiconductor Industry	Matlab	-3.112
Average PMI score		0.190

Table 2: The top-5 and bottom-3 co-occurred skill pairs with their PMI scores

Table 2 lists the top-5 and bottom-3 co-occurred skill pairs with their PMI scores, together with the average PMI score. From this table, we can see that if two skills are related, e.g., "IC" and "CMOS", these two skills tend to co-occur in one person’s profile, vice versa.

5 Joint Prediction Factor Graph Model

In this section, we propose a Joint Prediction Factor Graph (JPGF) model for learning and predicting the skills with personal and skill connection information besides local textual information.

5.1 Model

We formalize the problem of skill prediction using a pairwise factor graph model, and our basic idea of defining the correlations is to use different types of factor functions (i.e., personal connection factor, and skill connection factor). Here, the objective function $P_\theta(Y|X, G)$ is defined based on the joint probability of the factor functions, and the problem of collective skill inference model learning is cast as learning model parameters θ that maximizes the joint probability of skills based on the input continuous dynamic network.

Since directly maximizing the conditional probability $P_\theta(Y|X, G)$ is often intractable, we factorize the "global" probability as a product of "local" factor functions, each of which depends on a subset of the variables in the graph (Tang et al., 2013). In particular, we use three kinds of functions to represent the local textual information of the vector (local textual attribute function), personal connection information between vectors (personal connection factor) and skill connection information between skills (skill connection factor), respectively. We now briefly introduce the ways to define the above three functions.

Local textual attribute functions $f(x_{ij}, y_i)_j$: It denotes the attribute value associated with each person i . Here, we define the local textual attribute as a feature (Lafferty et al., 2001) and accumulate all the attribute functions to obtain local entropy for a person:

$$\frac{1}{Z_1} \exp \left(\sum_i \sum_k \alpha_k f_k(x_{ik}, y_i) \right) \quad (2)$$

Where α_k is the function weight, representing the influence degree of the attribute k . For simplicity, we use word unigrams of a text as the basic textual attributes.

Personal connection factor function $g(y_i, y_j)$: For the personal correlation factor function, we define it through the pairwise network structure. That is, if a person i and another person j have a personal relationship, we define a personal connection factor function as follows:

$$g(y_i, y_j) = \exp \left\{ \beta_{ij} (y_i - y_j)^2 \right\} \quad (3)$$

The personal connections are defined Section 4, i.e., co_major, co_univ, co_title, and co_corp. We define that if two persons have at least one personal connection edge, they have a personal relationship. In addition, β_{ij} is the weight of the function, representing the influence degree of i on j .

Skill connection factor function $h(y_i, y_j)$: For the skill connection factor function, we define it through the pairwise network structure. That is, if vector i and vector j are from the same person with different skills, we define their skill connection influence factor function as follows:

$$h(y_i, y_j) = \exp \left\{ \gamma_{ij} (y_i - y_j)^2 \right\} \quad (4)$$

Where γ_{ij} is the function weight, representing the influence degree of i on j .

By the above defined correlations, we can construct the graphical structure in the factor model. According to the Hammersley-Clifford theorem (Hammersley and Clifford, 1971), we integrate all the factor functions and obtain the following log-likelihood objective function:

$$\begin{aligned} L(\theta) &= \log_{\theta} P(Y|X, G) \\ &= \frac{1}{Z_1} \sum_i \sum_k \alpha_k f_k(x_{ik}, y_i) \\ &+ \frac{1}{Z_2} \sum_i \sum_{j \in NB(i)} \exp \left\{ \beta_{ij} (y_i - y_j)^2 \right\} \\ &+ \frac{1}{Z_3} \sum_i \sum_{k \in SAME(i)} \exp \left\{ \gamma_{ik} (y_i - y_k)^2 \right\} - \log Z \end{aligned} \quad (5)$$

Where (i, j) is a pair derived from the input network, $Z = Z_1 Z_2 Z_3$ is a normalization factor and $\theta = (\{\alpha\}, \{\beta\}, \{\gamma\})$ indicates a parameter configuration, $NB(i)$ denotes the set of social relationship neighbors nodes of i (personal connection), and $SAME(i)$ denotes the set of the node with the same person of i (skill connection).

5.2 Learning and Prediction

Model Learning: Learning of the factor model is to find the best configuration for free parameters $\theta = (\{\alpha\}, \{\beta\}, \{\gamma\})$ that maximizes the log likelihood objective function $L(\theta)$.

$$\theta^* = \arg \max L(\theta) \quad (6)$$

As the network structure in a social network can be arbitrary (e.g. possible of containing cycles), we use the Loopy Belief Propagation (LBP) algorithm (Tang et al., 2011a) to approximate the marginal distribution. To explain how we learn the parameters, we can get the gradient of each β_k with regard to the objective function (Eq. 5), taking β (the weight of the personal connection factor function $g(y_i, y_j)$) as an example:

$$\frac{L(\theta)}{\beta_k} = E[g(i, j)] + E_{\beta_k P(Y|X, G)}[g(i, j)] \quad (7)$$

Where $E[g(i, j)]$ is the expectation of factor function $g(i, j)$ given the data distribution in the input network and $E_{\beta_k P(Y|X, G)}[g(i, j)]$ represents the expectation under the distribution learned by the model, i.e., $P(y_i|X, G)$.

With the marginal probabilities, the gradient is obtained by summing up all triads (similar gradients can be derived for parameter α_k and γ_{ij}). It is worth noting that we need to perform the LBP process

twice in each iteration. The first run to estimate the marginal distribution of unknown variables $y_i = ?$ and the second one is to estimate the marginal distribution over all pairs. Finally, with the obtained gradient, we update each parameter with a learning rate η .

Skill Prediction: We can see that in the learning process, additional loopy belief propagation is used to infer the label of unknown relationships. After learning, all unknown skills are assigned with labels that maximize the marginal probabilities (Tang et al., 2011b), i.e.,

$$Y^* = \arg \max L(Y|X, G, \theta) \quad (8)$$

6 Experimentation

In this section, we first introduce the experimental setting, and then evaluate the performance of our proposed JPFG model with both personal and skill connection information.

6.1 Experimental Setting

As described in Section 3, the experimental data are collected from LinkedIn.com. With top 10 frequently used skills as candidate skills in all our experiments, we randomly select 2,000 profiles as training data and 1,000 profiles as testing data.

Though positive and negative samples of each skill are imbalanced (In this paper, the number of the negative samples is much larger than that of the positive samples), we select balanced testing and training samples for each skill. Following models are implemented and compared.

- *Keyword*, for each profile, we consider the profile attached with the skill, only if the text of the skill appears on the profile article with textual information.
- *MaxEnt*, which first uses local textual information as features to train a maximum entropy (ME) classification model, and then employs the classification model to predict the skills in the testing data set. The ME algorithm is implemented with the *mallet* toolkit ¹.
- *JPFG*, exactly our proposed model, which jointly predicts personal skills with local textual information, personal connection and skill connection.

For performance evaluation, we adopt Precision (P.), Recall (R.) and F1-Measure (F1.).

6.2 Comparison with Baselines

Our first group of experiments is to investigate whether the JPFG model is able to improve skill inference and whether the personal and skill connections are useful. The experimental results are shown in Table 3. From the table we can find that as some skills may not be mentioned on the Summary and Experience fields directly, the performance of the Keyword approach is far from satisfaction. As incorporating personal and skill connections, the JPFG model yields a much higher F1-measure, which improves the performance with about 6.8% gain than the MaxEnt model.

6.3 Performance of JPFG with Different Training Data Sizes

After we evaluate the effective of the JPFG model with the large-scale training data, we carry out experiments to test the effect of the JPFG model with different training data sizes. Experiment results are shown in Figure 3. It shows that the JPFG model with both personal and skill connections always outperform the two baseline models. Impressively, our JPFG model using 20% training data outperforms MaxEnt using 100% training data.

¹<http://mallet.cs.umass.edu/>

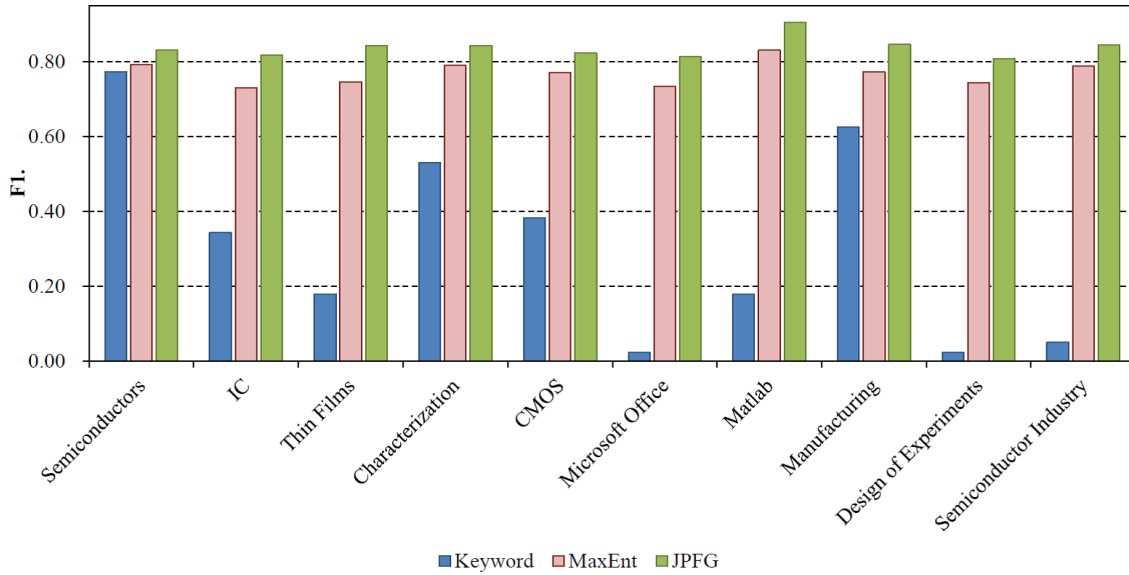


Figure 2: The performance of different methods for skill inference

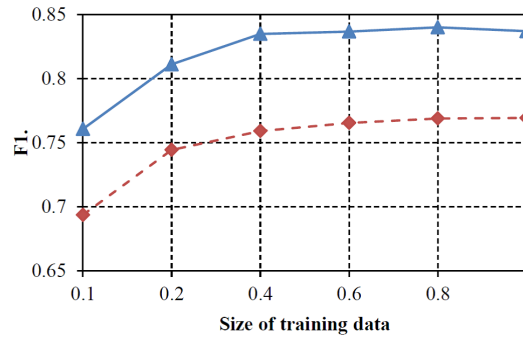


Figure 3: The performance of JPF with different training data sizes

6.4 Connections Contribution Analysis

Personal connections and skill connections can be also used to build the factor graph models to infer the skills. We therefore want to compare our JPF model with the factor graph model with only consider the personal connections or skill connections, and analysis the contribution of each kinds of connection. Specifically, MaxEnt-Personal employs the personal connections as additional features incorporated with textual features to build the maximum entropy classification. FGM-Personal is a simplified version of the JPF model, which only employs textual attribute functions and personal connection factor functions to build the factor graph model. Likewise, FGM-Skill only employs textual attribute functions and skill connection factor functions to build the factor graph model. Table 3 shows the experiment results.

System	P.	R.	F1.
MaxEnt	0.744	0.797	0.769
MaxEnt-Personal	0.758	0.812	0.783
FGM-Personal	0.765	0.817	0.790
FGM-Skill	0.704	0.967	0.815
JPF	0.780	0.905	0.837

Table 3: The contribution of connections

From Table 3, we can observe that, 1) Both FGM-Personal and FGM-Skill outperform the baseline

MaxEnt approach. It shows that both personal connections and skill connections are helpful for skill inference; 2) MaxEnt-Personal and FGM-Personal outperform the baseline MaxEnt approach, it show that personal connections are helpful for inferring skills, and as considering the global optimization, FGM-Personal is more effective; 3) FGM-Skill built on the skill connections is more effective than MaxEnt-Personal and FGM-Personal, it show that skill connections are more useful than personal connections; 4) JPF model outperforms both FGM-Personal and FGM-Skill, it suggests that we should incorporate both personal and skill connections to the factor graph model when we infer the skills from profile.

7 Conclusion

In this study, we propose a novel task named personal skill inference, which aims to determine whether a person takes a specific skill or not. To address this task, we propose a joint prediction factor graph model with help of both personal and skill connections besides local textual information. Evaluation on a large-scale dataset shows that our joint model performs much better than several baselines. In particular, it shows that the performance on personal skill inference can be greatly improved by incorporating skill connection information.

The general idea of exploring personal and skill connections to help predict people's skills represents an interesting research direction in social networking, which has many potential applications. Besides, as skill information of a person is normally incomplete and fuzzy, how to better infer personal skills with weakly labeled information is challenging.

Acknowledgements

This research work is supported by the National Natural Science Foundation of China (No. 61273320, No. 61331011, and No. 61375073), National High-tech Research and Development Program of China (No. 2012AA011102), Zhejiang Provincial Natural Science Foundation of China (No. LY13F020007), the Humanity and Social Science on Young Fund of the Ministry of Education (No. 12YJC630170).

We thank Dr. Jie Tang and Honglei Zhuang for providing their software and useful suggestions about PGM. We thank Prof. Deyi Xiong for helpful discussions, and we acknowledge Dr. Xinfang Liu, and Yunxia Xue for corpus construction and insightful comments. We also thank anonymous reviewers for their valuable suggestions and comments.

References

- Balog K and M. Rijke. 2007. Determining Expert Profiles (With an Application to Expert Finding). In *Proceedings of IJCAI-07*.
- Campbell C, P. Maglio, A. Cozzi, and B. Dom. 2003. Expertise Identification Using Email Communications. In *Proceedings of CIKM-03*.
- Dong Y., J. Tang, S. Wu, J. Tian, N. Chawla, J. Rao, and H. Cao. 2012. Link Prediction and Recommendation across Heterogeneous Social Networks. In *Proceedings of ICDM-12*.
- Guo W., H. Li, H. Ji, and M. Diab. 2013. Linking Tweets to News: A Framework to Enrich Short Text Data in Social Media. In *Proceedings of ACL-13*.
- Guy I., N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel. 2010. Social Media Recommendation based on People and Tags. In *Proceedings of SIGIR-10*.
- Hammersley J. and P. Clifford. 1971. Markov Field on Finite Graphs and Lattices, *Unpublished manuscript*.
- Helic D. and M. Strohmaier. 2011. Building Directories for Social Tagging Systems. In *Proceedings of CIKM-2011*.
- Lafferty J, A. McCallum, and F. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of ICML-01*.

- Lamos V., D. Preo?iuc-Pietro, and T. Cohn. 2013. A User-centric Model of Voting Intention from Social Media. In *Proceedings of ACL-13*.
- Lappas T., K. Punera, and T. Sarlos. 2011. Mining Tags Using Social Endorsement Networks. In *Proceedings of SIGIR-11*.
- Li H., Z. Liu, and M. Sun. 2012. Random Walks on Context-Aware Relation Graphs for Ranking Social Tags. In *Proceedings of COLING-12*.
- Liu Z., X. Chen, and M. Sun. 2011. A Simple Word Trigger Method for Social Tag Suggestion. In *Proceedings of EMNLP-2011*.
- Liu Z., C. Tu, and M. Sun. 2012. Tag Dispatch Model with Social Network Regularization for Microblog User Tag Suggestion. In *Proceedings of COLING-12*.
- Lu Y., and P. Tsaparas, A. 2010. Ntoulas and L. Polanyi. 2010. Exploiting Social Context for Review Quality Prediction. In *Proceedings of WWW-10*.
- Luo T., J. Tang, J. Hopcroft, Z. Fang, and X. Ding. 2013. Learning to Predict Reciprocity and Triadic Closure in Social Networks. *ACM Transactions on Knowledge Discovery from Data*. vol.7(2), Article No. 5.
- Murphy K., Y. Weiss, and M. Jordan. 1999. Loopy Belief Propagation for Approximate Inference: An Empirical Study. In *Proceedings of UAI-99*.
- Ohkura T., Y. Kiyota and H. Nakagawa. 2006. Browsing System for Weblog Articles based on Automated Folksonomy. In *Proceedings of WWW-06*.
- Si X., Z. Liu, and M. Sun. 2010. Explore the Structure of Social Tags by Subsumption Relations. In *Proceedings of COLING-10*.
- Soboroff I., A. Vries and N. Craswell. 2006. Overview of the TREC 2006 Enterprise Track In *Proceedings of TREC-06*.
- Turney P. 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of reviews. In *Proceedings of ACL-02*.
- Tan C., L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li. 2011. User-Level Sentiment Analysis Incorporating Social Networks. In *Proceedings of KDD-11*.
- Tang W., H. Zhuang, and J. Tang. 2011a. Learning to Infer Social Ties in Large Networks. In *Proceedings of ECML/PKDD-11*.
- Tang J., Y. Zhang, J. Sun, J. Rao, W. Yu, Y. Chen, and A. Fong. 2011b. Quantitative Study of Individual Emotional States in Social Networks. *IEEE Transactions on Affective Computing*. vol.3(2), Pages 132-144.
- Tang J., S. Wu, J. Sun, and H. Su. 2012. Cross-domain Collaboration Recommendation. In *Proceedings of KDD-12*.
- Tang J., S. Wu, and J. Sun. 2013. Confluence: Conformity Influence in Large Social Networks. In *Proceedings of KDD-13*.
- Xing E, M. Jordan, and S. Russell. 2003. A Generalized Mean Field Algorithm for Variational Inference in Exponential Families. In *Proceedings of UAI-03*.
- Yang S., B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. 2011a. Like like alike - Joint Friendship and Interest Propagation in Social Networks. In *Proceedings of WWW-11*.
- Yang Z., K. Cai, J. Tang, L. Zhang, Z. Su, and J. Li. 2011b. Social Context Summarization. In *Proceedings of SIGIR-11*.
- Zhang J., J. Tang, and J. Li. 2007a. Expert Finding in A Social Network. In *Proceedings of the Twelfth Database Systems for Advanced Applications (DASFAA-2007)*.
- Zhang J., M. Ackerman, and L. Adamic. 2007b. Expertise Networks in Online Communities: Structure and Algorithms. In *Proceedings of TREC-07*.
- Zhuang H, J. Tang, W. Tang, T. Lou, A. Chin, and X. Wang. 2012. Actively Learning to Infer Social Ties. In *Proceedings of Data Mining and Knowledge Discovery (DMKD-12)*, vol.25 (2), pages 270-297.