Reciprocal and Heterogeneous Link Prediction in Social Networks

Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton, and Ashesh Mahidadia

School of Computer Science and Engineering,
The University of New South Wales, Sydney, NSW 2052, Australia
{xcai,mike,alfredk,wobcke,yskim,compton,ashesh}@cse.unsw.edu.au

Abstract. Link prediction is a key technique in many applications in social networks, where potential links between entities need to be predicted. Conventional link prediction techniques deal with either homogeneous entities, e.g., people to people, item to item links, or non-reciprocal relationships, e.g., people to item links. However, a challenging problem in link prediction is that of heterogeneous and reciprocal link prediction, such as accurate prediction of matches on an online dating site, jobs or workers on employment websites, where the links are reciprocally determined by both entities that heterogeneously belong to disjoint groups. The nature and causes of interactions in these domains makes heterogeneous and reciprocal link prediction significantly different from the conventional version of the problem. In this work, we address these issues by proposing a novel learnable framework called *ReHeLP*, which learns heterogeneous and reciprocal knowledge from collaborative information and demonstrate its impact on link prediction. Evaluation on a large commercial online dating dataset shows the success of the proposed method and its promise for link prediction.

Keywords: Machine Learning, Data Mining, Information Retrieval, Recommender Systems

1 Introduction

Social networks are commonly used to model the interactions among people in communities, which can be represented by graphs where a vertex corresponds to a person in some community and an edge or *link* represents some association between the corresponding people. Understanding the association between two specific vertices by predicting the likelihood of a future but not currently existing association between them is a fundamental problem known as *link prediction* [13].

Social interaction on the Web often involves both positive and negative relationships, e.g., since attempts to establish a relationship may fail due to rejection from the intended target. This generates links that signify rejection of invitations, disapproval of applications, or expression of disagreement with others' opinions. Such social networks are *reciprocal* since the sign of a link indicating

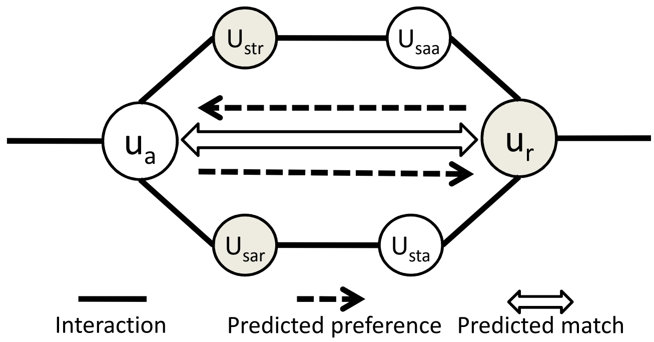

Fig. 1. Collaborative Information for Reciprocal and Heterogeneous Link Prediction. Links as interactions connect nodes (in circles) that belong to different groups (grey shading or no shading). The signs of links represent matches of two way decisions, which are predicted by a learning model using collaborative information.

whether it is positive or negative depends on the attitudes or opinions of both entities forming the link. While the interplay of positive and negative relations is clearly important in many social network settings, the vast majority of online social network research has considered only positive relationships. Moreover, *reciprocal* positive and negative relationships have been even less investigated. Recently, social network analysis has had a variety of applications, such as online dating sites, education admission portals as well as jobs, employment, career and recruitment sites, where people in the networks have different roles, and links between them can only be between people in different roles. Such networks are *heterogeneous*, creating challenges for link prediction since existing approaches focus only on homogeneous networks where nodes in the networks have the same role and any of them may link to any other.

In this work, we consider the heterogeneous and reciprocal link prediction problem. We propose a framework to address prediction of the sign of a link in heterogeneous and reciprocal networks. We model this problem as a machine learning problem and create structural features for learning, i.e. we construct features for learning based on structural collaborative information. Specifically, motivated by *taste* and *attractiveness* in the Social Collaborative Filter [4], we first define a structural unit called a *tetrad* (to be defined in Section 4.1), i.e. a path crossing four nodes as in Figure 1 [4], in the graph of networks based on a set of variations of collaborative filtering. These represent collaborative information regarding taste and attractiveness of nodes in the graph (people in social networks). The properties of each tetrad are then measured in terms of positive and negative signs through its path. Finally, the properties of a tetrad are used as features in a learning framework for link sign prediction.

The paper is organised as follows. Section 2 presents related work. Section 3 defines the problem. Section 4 develops a learnable framework for the reciprocal and heterogeneous link prediction problem. Experimental evaluation is in Section 5 and we conclude in Section 6.

2 Related Work

Liben-Nowell and Kleinberg [13] developed one of the earliest link prediction models for social networks. They concentrated mostly on the performance of various graph-based similarity metrics for the link prediction problem. This work has since been extended to use supervised learning for link prediction [8, 2, 7], where link prediction was considered as a binary classification task in a supervised learning setup using features extracted from various network properties.

Recent developments in online social networks such as Facebook and Twitter have raised scalability challenges for link prediction. Large scale link prediction was addressed by Acar et al. in [1], where higher-order tensor models based on matrix factorisation were used for link prediction in large social networks.

Recently, work on link prediction has started considering both negative and positive relationships in online websites [12, 3, 10, 11]. Leskovec et al. investigated the problem of link sign prediction to uncover the mechanisms that determine the signs of links in large social networks where interactions can be both positive and negative [12]. Also, learning methods based on multiple sources and multiple path-based features were investigated. In [6], there is a collective link prediction problem where several related link prediction tasks are jointly learned. In [14], a supervised learning framework has been designed based on a rich variety of path-based features using multiple sources to learn the dynamics of social networks.

In reciprocal and heterogeneous link prediction, the reciprocal and heterogeneous nature of networks makes the problem significantly different from traditional link prediction. Therefore, new methods to: 1) model the characteristics of how reciprocal and heterogeneous links form; and 2) that can be used for mining and predicting such links in large social network datasets are essential.

3 Problem Statement

Link prediction is defined as the inference of new interactions among the members of a given social network [13]. More formally, the link prediction problem is defined as: given a snapshot of a social network at time t, seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time $t' = t + \delta t$. The solution to this problem lies in modelling the evolution of the network using intrinsic features derived from the network itself in order to understand which properties of the network lead to the most accurate link predictions.

Edge sign prediction is an important type of link prediction, defined as follows. Suppose we are given a social network with signs on all its edges, but the sign on the edge from node u to node v, denoted $s(u, v)$, has been hidden. How reliably can we infer this sign $s(u, v)$ using the information provided by the rest of the network? This problem is both a concrete formulation of our basic questions about the typical patterns of link signs, and also a way of approaching our motivating application of inferring unobserved attitudes among users of social computing sites. For a given link in a social network, we will define its sign to be positive or negative depending on whether it expresses a positive or negative attitude from the source of the directed link to the target, and *vice versa*.

Heterogeneous and reciprocal link prediction deals with predictions for heterogeneous and reciprocal links in social networks. We define the key concepts as follows. A *heterogeneous and reciprocal link* is a link (u, v) that has its two nodes u and v belonging to different types, or groups (i.e., the nodes are heterogeneous) and its edge sign depends on the attitudes of both nodes (i.e., link establishment is reciprocal). *Heterogeneous and reciprocal social networks* are networks connected only by heterogeneous and reciprocal links. A *heterogeneous and reciprocal link prediction* problem (*ReHeLP*) is the prediction of links in heterogeneous and reciprocal social networks.

Heterogeneous and reciprocal social networks exist in many applications (e.g., online dating sites, education admission sites, as well as jobs, employment, career and recruitment sites). In online dating sites, we have: 1) users belonging to different groups (male or female); 2) links established only between users from different groups; and 3) link signs dependent on the compatibility of user pairs.

4 Methods

Given a directed graph $G = (V, E)$ with a sign (positive or negative) on each edge, we let $s(u, v)$ denote the sign of the edge (u, v) from node u to node v. That is, $s(u, v) = 1$ when the sign of $(u, v))$ is positive, 0 when negative. For different formulations of our task, we suppose that for a particular edge (u, v), the sign $s(u, v)$ is hidden and that we trying to infer it.

4.1 Feature Construction

The first step towards our heterogeneous and reciprocal link prediction is feature construction, which defines a collection of features for learning a model. The features are divided into two categories, according to their relationships to the entities in the networks.

Monadic Features In social networks, the activity and popularity of an entity have impact on the behaviour of the entity. Therefore, the first category of characteristics of entities to be measured for link prediction is the activity and popularity of entities, which are the aggregated local relations of an entity to the rest of the world. This type of information represents the baseline, quantifying how many ingoing and outgoing edges a node could have.

The number of outgoing actions of a node measures how active an entity in the networks is, represented by its out-degree, the number of outgoing edges of a node in graph. We define the first monadic features based on the degree of the outgoing edges, as follows. An *outgoing edge* e of a node v is an edge that directs from v to another node. The *degree of outgoing edges* of a node $d_o(v)$ is the number of outgoing edges from that node v. We also separate the outgoing edges according to their sign and define the degree of positive outgoing edges and the degree of negative outgoing edges, which represents not only the activity but also the general attitude of an entity to the world. The *degree of positive outgoing edges* of a node $d_o^+(v)$ is the number of outgoing edges from that node v and with positive sign. The *degree of negative outgoing edges* of a node $d_o^-(v)$, is the number of outgoing edges from that node v and with negative sign.

Similarly, the number of incoming actions of a node measures how popular an entity is in the network, represented by the degree of incoming edges. Therefore, we define the degree of positive incoming edges and the degree of negative incoming edges to model the popularity and again general attitude of an entity as follows. The *degree of positive incoming edges* of a node $d_i^+(v)$ is the number of incoming edges to that node v with positive sign. The *degree of negative incoming edges* of a node $d_i^-(v)$, is the number of incoming edges to that node v with negative sign.

The four monadic features $(d_o^+(v), d_o^-(v), d_i^+(v), d_i^+(v))$ will be used in our method to represent the activity and popularity as well as the general attitude of an entity in a network.

Dyadic Features We also define dyadic features based on collaborative information. Collaborative information is the information extracted from a community that represents knowledge about the network derived from collaborative efforts. Collaborative information is the basis of collaborative filtering for recommendation, which makes automatic predictions about the interests of a user by collecting preferences or taste information from many other users. We make use of collaborative information for link prediction and extract dyadic features as in collaborative filtering.

The links of interest represent reciprocal relationships between entities, requiring *reciprocal* collaborative information to be considered [4, 5, 9]. Reciprocal collaborative information could be embedded in several different kinds of collaborative filtering frameworks. In [4], a general framework for reciprocal collaborative filtering was developed for recommendation, which was then extended in several variant methods [9]. Here, we contribute to integrating such collaborative information into a learnable framework for link prediction, rather than recommendation. Moreover, we aim at prediction of heterogeneous links, hence both nodes of a link cannot link to the same third node. For example, in people to people dating recommendation, a link only exists between a heterogeneous pair, i.e. a male type and a female type (we do not consider same-sex relationships in this work). In this bipartite representation the sender and recipient cannot both link to the same third person. Therefore, we consider a three step path involving both nodes within a potential link, which is defined as a *tetrad* in Definition 1.

Definition 1. *A tetrad $t(u, s_v, s_u, v)$ or $t(u, v)$ is a three step path among four different nodes ($u \to s_v \to s_u \to v$) in a graph, where the source node u (sender)*

Table 1. Dyadic Features Based on Reciprocal Collaborative Information

Typical		Inverted				Transmissible	
RSR	SRS	RRS	RSS	SRR	SSR	SSS	RRR
$n(r_+s_+r_+)$	$n(s_+r_+s_+)$	$n(r_+r_+s_+)$	$n(r_+s_+s_+)$	$n(s_+r_+r_+)$	$n(s_+s_+r_+)$	$n(s_+s_+s_+)$	$n(r_+r_+r_+)$
$n(r_+s_+r_-)$	$n(s_+r_+s_-)$	$n(r_+r_+s_-)$	$n(r_+s_+s_-)$	$n(s_+r_+r_-)$	$n(s_+s_+r_-)$	$n(s_+s_+s_-)$	$n(r_+r_+r_-)$
$n(r_+s_-r_+)$	$n(s_+r_-s_+)$	$n(r_+r_-s_+)$	$n(r_+s_-s_+)$	$n(s_+r_-r_+)$	$n(s_+s_-r_+)$	$n(s_+s_-s_+)$	$n(r_+r_-r_+)$
$n(r_+s_-r_-)$	$n(s_+r_-s_-)$	$n(r_+r_-s_-)$	$n(r_+s_-s_-)$	$n(s_+r_-r_-)$	$n(s_+s_-r_-)$	$n(s_+s_-s_-)$	$n(r_+r_-r_-)$
$n(r_-s_+r_+)$	$n(s_-r_+s_+)$	$n(r_-r_+s_+)$	$n(r_-s_+s_+)$	$n(s_-r_+r_+)$	$n(s_-s_+r_+)$	$n(s_-s_+s_+)$	$n(r_-r_+r_+)$
$n(r_-s_+r_-)$	$n(s_-r_+s_-)$	$n(r_-r_+s_-)$	$n(r_-s_+s_-)$	$n(s_-r_+r_-)$	$n(s_-s_+r_-)$	$n(s_-s_+s_-)$	$n(r_-r_+r_-)$
$n(r_-s_-r_+)$	$n(s_-r_-s_+)$	$n(r_-r_-s_+)$	$n(r_-s_-s_+)$	$n(s_-r_-r_+)$	$n(s_-s_-r_+)$	$n(s_-s_-s_+)$	$n(r_-r_-r_+)$
$n(r_-s_-r_-)$	$n(s_-r_-s_-)$	$n(r_-r_-s_-)$	$n(r_-s_-s_-)$	$n(s_-r_-r_-)$	$n(s_-s_-r_-)$	$n(s_-s_-s_-)$	$n(r_-r_-r_-)$

and a node similar to it s_u (defined by collaborative filtering) both belong to one of the two types, while the target node v (recipient) and another node similar to it s_v both belong to the other type.

A *tetrad* $t(u, v)$ captures a two step relationship across two types, which is the minimum indirect path between a pair of nodes (u, v). *Tetrad* is a novel type of sub-graph feature for heterogeneous and reciprocal link prediction, which captures collaborative information that cannot be captured by features used in existing link prediction approaches. The following feature sets for each pair of nodes (u, v) are then based on a variety of minimum indirect paths defined on the pair.

The first type of dyadic features is based on *Typical Reciprocal Collaborative Information.* As shown in Figure 2, this is the typical reciprocal collaborative filtering for people to people recommendation with two-way preferences [4, 5]. In the figure, u_a is the source node (corresponding to u in Definition 1), u_r the target node (v in Definition 1), u_{str} and u_{sar} the similar nodes (s_v in Definition 1), u_{saa} and u_{sta} the similar nodes (s_u in Definition 1). From this configuration, we can construct two set of features. One set is (on the top half of the figure) to capture the collaborative information for predicting the recipient's preference. The other set is (on the bottom half of the figure) to capture the collaborative information for predicting the initiator's preference. Since we have positive or negative signs for each interaction and there are 3 interactions in each set as shown in the figure, we can create $2 * 2^3 = 16$ features of this type as in Table 1 indicated by RSR and SRS. In Table 1, a tetrad type is presented by the directions of three edges from u to v in a tetrad $t(u, v)$, where S and s mean a link from a precursor to a successor, R and r to a precursor from a successor, and their signs, where $+$ means a positive link and $-$ negative. n is used to represent the number of links of a tetrad type. To give an example, $n(r_+s_-r_+)$ means the total number of the type of tetrad $t(u, s_v, s_u, v)$ that have a positive link from u to s_v, a negative link to s_v from s_u and a positive link from s_u to v.

The second type of dyadic feature is based on *Inverted Reciprocal Collaborative Information.* This type of dyadic feature is derived by fitting the inverted

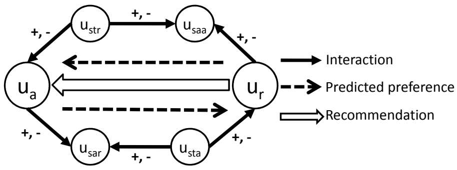

Fig. 2. Typical reciprocal collaborative filtering.

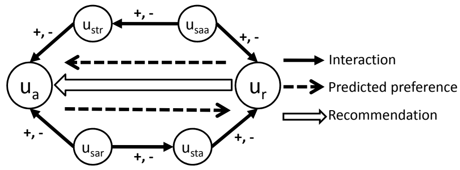

Fig. 3. Inverted reciprocal collaborative filtering with similar recipient.

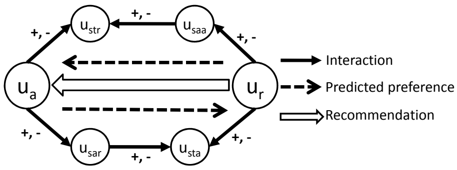

Fig. 4. Inverted reciprocal collaborative filtering with similar sender.

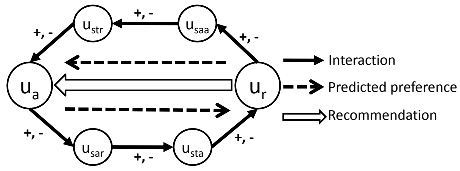

Fig. 5. Reciprocal collaborative filtering by preference transmission.

collaborative filters [9] into the reciprocal collaborative filtering framework [4]. There are two types of inverted collaborative filters: user-based and item-based. In this work, we make use of both of these inverted collaborative filters to generate more features. We first take the recipient-based inverted collaborative filtering and add another novel type of collaborative information (on the top half of the figure) to capture the preference of the recipient as shown in Figure 3. The original inverted collaborative filters only model positive signs for all interactions in the configuration. To allow more collaborative information to be considered by the learning system in our configuration, we use both positive and negative signs in any interaction within the configuration, and rely on the machine learning method to select discriminative features. Similarly, we add one more new type of collaborative information (on the top half of the figure) to the sender-based inverted collaborative filtering to capture the preference of the recipient as shown in Figure 4. We also allow both positive and negative signs for interaction. Features based on inverted collaborative information are summarised in Table 1 indicated by RRS, RSS, SRR and SSR.

The third type of dyadic feature is based on *Transmissible Collaborative Information*. Dyadic features are considered also to capture the transmissible properties of preferences as in Figure 5. Similar to [9], if we only consider positive interactions in creating the collaborative information, we should then have the property of preference transmission. This can be easily validated by the taste and attractiveness concept in [4]. Similarly, we have the first set of features (on the top half of the figure) to capture the recipient's preference and another set of features (on the bottom half of the figure) to capture the sender's preference. We again allow both positive and negative signs in any interaction within the configuration. Features based on preference transmission are illustrated in Table 1 indicated by SSS and RRR.

Each of these 64 tetrad types may provide different evidence about the sign of the edge from the initiator to the recipient, possibly some favouring a negative sign and some favouring a positive sign. We encode this information in a 64-dimensional vector specifying the number of tetrads of each type that both nodes in a link are involved in. Notice that this is the first time a complete set of combinations showing all possible sources of collaborative information derived from the structure of tetrads has been used for link prediction.

4.2 Learning and Testing

To predict links, we first calculate the feature values and then calculate a measure of combined feature strength as the weighted combination of feature values, as follows:

$$s = \sum_{i=1}^{n} \omega_i x_i + \omega_0 \tag{1}$$

where s is the combined feature strength, x_i the value of the ith feature and ω_i the weight value for x_i. To learn the weights and convert this combined feature strength into an edge sign prediction, we use logistic regression, which will output a value in the range of $(0, 1)$ representing the probability of a positive edge sign:

$$p = \frac{1}{1 + e^{-s}} \tag{2}$$

where p is the predicted probability of an positive edge sign.

We will show in Section 5 that by using logistic regression, we are also able to uncover the contribution of each feature to the prediction by investigating the learned coefficients.

Once we have the learned model, testing is simply to calculate feature values for each test instance (pair of nodes) and input them into the learned model to compute the probability of a positive link between them. The instances are then classified into positive or negative according to the thresholding of the probability value with respect to a threshold.

5 Experiments

5.1 Setup

In these experiments, we aim to evaluate the proposed approach on link prediction of dating social networks in a real world dataset, which is a demanding real-world one. Link prediction on dating social networking is a typical heterogeneous and reciprocal link prediction problem, where the nodes are users and links are interactions. Here users are either of male or female type, links are only estimated between users of a different type and the link sign depends on the decisions of both users. The datasets were collected from a commercial social

Table 2. Dataset Description

	#Interaction	#Positive Link	#Negative Link	#User
Numbers in Dataset	1710332	264142(15%)	1446190(85%)	166699

network (online dating) site containing interactions between users. Specifically, the data contains records, each of which represents a contact (communication) by a tuple containing the identity of the contact's sender, the identity of the contact's recipient, and an indicator showing whether the interaction was successful (with a positive response from the recipient to the sender) or unsuccessful (with a negative response or no response).

The experiments were conducted on a dataset covering a four week period in March, 2010. The dataset contains all users with at least one contact in the specific period. The dataset used for this research is summarised in Table 2. We follow the methodology of Leskovec [12] and created a balanced dataset with equal numbers of positive and negative edges.

To the best of our knowledge, there is no existing work on edge sign prediction in heterogeneous and reciprocal networks. Therefore, we took the recent approach to positive and negative link prediction in online social networks by Leskovec et al. [12] as the baseline to compare to the proposed algorithm, which has been reported to significantly improve on previous approaches [12]. The baseline method uses all valid features except those based on two-step paths that are not valid for heterogeneous and reciprocal link prediction since the latter has a tetrad structure. Notice that although the baseline method has some utility in heterogeneous and reciprocal link sign prediction, it is not designed for that problem. To the best of our knowledge, we are the first to consider the heterogeneous and reciprocal link sign prediction problem.

We use accuracy, precision and recall as evaluation metrics for evaluation of the proposed algorithm. We also use the receiver operating characteristic (ROC) and the area under the ROC curve (AUC) for our evaluation.

Algorithms for feature extraction were implemented using SQL in Oracle 11. Learning and testing algorithms were implemented in Matlab. For the large scale dataset in Table 2, feature extraction required less than 1 hour. Training on 90% of the balanced dataset and testing on the remaining 10% of the dataset took about 1 minute on a workstation with 64-bit Windows 7 Professional, 2 processors of Intel(R) Xenon(R) CPU x5660@2.80GHz and 32GB RAM.

5.2 Results

To compare the proposed method to the baseline method, we conducted experiments to generate values for the evaluation metrics from two methods and statistically tested significance of differences using a paired t-test.

We used 10 fold cross-validation (CV) repeated 10 times and hold-out to generate results for evaluation. For hold-out, we hold 20% of the data for testing and

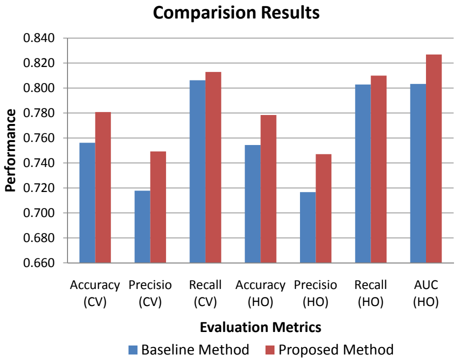

Fig. 6. Comparative Results. CV indicates 10-fold CV and HO holding out.

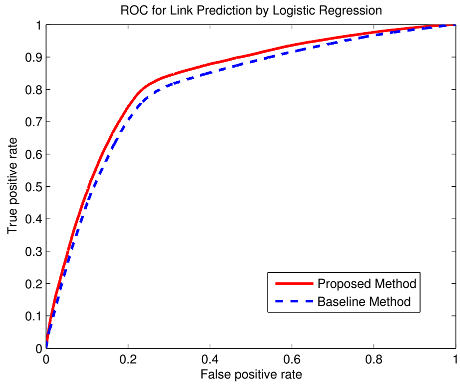

Fig. 7. ROC for comparison of the proposed method and baseline method.

Table 3. Comparative Results. ("*" indicates improvement of the method.)

		Baseline	**ReHeLP**	Improvement
10-fold CV repeated 10 times	Accuracy	0.756	**0.781***	2.46%
	Precision	0.718	**0.749***	3.14%
	Recall	0.806	**0.813***	0.67%
20% hold out for testing	Accuracy	0.754	**0.778***	2.40%
	Precision	0.717	**0.747***	3.04%
	Recall	0.803	**0.810***	0.71%
	AUC	0.803	**0.827***	2.35%

train the model using the remaining 80% of the data. The comparative results are shown in Figure 6 with details in Table 3. The proposed method achieves about 78% predictive accuracy on average on 100 runs by 10 fold CV repeated 10 times while the baseline method only has less than 76% predictive accuracy on average, showing the proposed method outperforms the baseline method by about 2.5% predictive accuracy. For hold-out evaluation, the proposed method similarly outperforms the baseline method by 2.4% predictive accuracy. The proposed method also outperforms the baseline method in terms of precision, recall and AUC as shown in Table 3, where the proposed method achieved 3.14%, 0.67% improvement over the baseline method for precision and recall respectively by 10 fold cross-validation repeated 10 times, and 3.04%, 0.71% improvement over the baseline method for precision and recall respectively by hold-out. The threshold selected for this evaluation is based on the optimal operating point selected using the ROC curve in Figure 7, where the ROC curve of the proposed method remains above that of the baseline method also indicating the improvement by the former. On the AUC of the ROC curve, the proposed method outperforms the baseline method by 2.35%.

Table 4. Paired t-test at the 5% Significance Level.

Hypothesis (h)	p-value	95% Confidence Interval
1	2.74E-98	[0.0241 0.0251]

A paired t-test is used to assess whether the means of the results of our method and the compared method are statistically different from each other. The result of a paired t-test on the corresponding predictive accuracy x of the proposed method and predictive accuracy y of the compared baseline method by 10-fold cross-validation repeated 10 times is shown in Table 4. Here the paired t-test tests the null hypothesis that data in the difference $x - y$ are a random sample from a normal distribution with mean 0 and unknown variance, against the alternative that the mean is not 0, i.e. the null hypothesis that the results come from populations with equal means, against the alternative that the means are unequal. Our experiments show that the test rejects the null hypothesis at the $\alpha = 0.05$ significance level as shown by the hypothesis $h = 1$ in the table. Notice that the 95% confidence interval on the difference mean contains a positive interval that does not contain 0, which indicates that the mean of the predictive accuracy of the proposed method is greater than that of the baseline method. Moreover, the p value has fallen below $\alpha = 0.05$ and in fact even below $\alpha = 0.01$, which can be considered as a very significant difference (significant improvement in accuracy).

6 Conclusion

We have presented a learning framework for the heterogeneous and reciprocal link prediction problem. To the best of our knowledge, this the first work to address the link sign prediction problem in heterogeneous and reciprocal social networks. The improvement gained by the proposed approach has been clearly demonstrated by a set of extensive experiments. The experiments were conducted in demanding real world datasets collected from a commercial social network site, which shows that the proposed method is able to make heterogeneous and reciprocal link predictions and outperforms the use of existing link prediction techniques.

Future work will include investigating methods to use better collaborative information and understand the way that the collaborative information contributes to the prediction in order to design and make use of improved features.

Acknowlegement

This project is funded by the Smart Services CRC under the Australian Government's Cooperative Research Centre program. We would also like to thank our industry partners for providing the datasets.

References

1. Acar, E., Dunlavy, D.M., Kolda, T.G.: Link prediction on evolving data using matrix and tensor factorizations. In: Proceedings of the 9th IEEE International Conference on Data Mining Workshops. pp. 262–269 (2009)
2. Bilgic, M., Namata, G.M., Getoor, L.: Combining collective classification and link prediction. In: Proceedings of the 7th IEEE International Conference on Data Mining Workshops. pp. 381–386 (2007)
3. Brzozowski, M.J., Hogg, T., Szabo, G.: Friends and foes: ideological social networking. In: Proceeding of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems. pp. 817–820 (2008)
4. Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y., Compton, P., Mahidadia, A.: Collaborative filtering for people to people recommendation in social networks. In: Li, J. (ed.) AI 2010: Advances in Artificial Intelligence, pp. 476–485. Springer-Verlag, Berlin (2010)
5. Cai, X., Bain, M., Krzywicki, A., Wobcke, W., Kim, Y.S., Compton, P., Mahidadia, A.: Learning collaborative filtering and its application to people to people recommendation in social networks. In: Proceedings of the 10th IEEE International Conference on Data Mining. pp. 743–748 (2010)
6. Cao, B., Liu, N.N., Yang, Q.: Transfer learning for collective link prediction in multiple heterogenous domains. In: Proceedings of the 27th International Conference on Machine Learning. pp. 159–166 (2010)
7. Chao, W., Satuluri, V., Parthasarathy, S.: Local probabilistic models for link prediction. In: Proceedings of the 7th IEEE International Conference on Data Mining. pp. 322–331 (2007)
8. Hasan, M.A., Chaoji, V., Salem, S., Zaki, M.: Link prediction using supervised learning. In: Workshop on Link Analysis, Counter-terrorism and Security (at SIAM Data Mining Conference) (2006)
9. Krzywicki, A., Wobcke, W., Cai, X., Mahidadia, A., Bain, M., Compton, P., Kim, Y.: Interaction-based collaborative filtering methods for recommendation in online dating. In: Chen, L., Triantafillou, P., Suel, T. (eds.) Web Information Systems Engineering – WISE 2010, pp. 342–356. Springer-Verlag, Berlin (2010)
10. Kunegis, J., Lommatzsch, A., Bauckhage, C.: The slashdot zoo: mining a social network with negative edges. In: Proceedings of the 18th International Conference on World Wide Web. pp. 741–750 (2009)
11. Lampe, C.A., Johnston, E., Resnick, P.: Follow the reader: filtering comments on slashdot. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1253–1262 (2007)
12. Leskovec, J., Huttenlocher, D., Kleinberg, J.: Predicting positive and negative links in online social networks. In: Proceedings of the 19th International Conference on World Wide Web. pp. 641–650 (2010)
13. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: Proceedings of the 12th International Conference on Information and Knowledge Management. pp. 556–559 (2003)
14. Lu, Z., Savas, B., Tang, W., Dhillon, I.S.: Supervised link prediction using multiple sources. In: Proceedings of the 10th IEEE International Conference on Data Mining. pp. 923–928 (2010)