3HBP A Three-Level Hidden Bayesian Link Prediction Model in Social Networks

**ABSTRACT**

In social networks, link establishment among the users is affected by complex factors. In this paper, we try to investigate the internal and external factors that affect the formation of links and propose a three-level hidden Bayesian link prediction model by integrating the user behavior as well as user relationships to link prediction. First, based on the user multiple interest characteristics, a latent Dirichlet allocation (LDA) traditional text modeling method is applied into user behavior modeling. Taking the advantage of LDA topic model in dealing with the problem of polysemy and synonym, we can mine user latent interest distribution and analyze the effects of internal driving factors. Second, owing to the power-law characteristics of user behavior, LDA is improved by Gaussian weighting. In this way, the negative impact of the interest distribution to the high-frequency users can be reduced and the expression ability of interests can be enhanced. Furthermore, taking the impact of common neighbor dependencies in link establishment, the model can be extended with hidden naive Bayesian algorithm. By quantifying the dependencies between common neighbors, we can analyze the effects of external driving factors and combine internal driving factors to link prediction. Experimental results indicate that the model can not only mine user latent interest distribution but also can improve the performance of link prediction effectively.

**EXISTING SYSTEM**

* Chang and Blei [29] presented a relational topic models with the text data and analyzed the topic distribution of texts to predict links among the texts. Liu et al. [18] presented a simple but effective similarity-based prediction strategy based on the label propagation, which imitated the communication between people naturally. However, the abovementioned methods focused on analyzing the user direct interests generated by key words and labels, none of these approaches leverage user latent interests generated by user behavior.
* Shahmohammadi et al. [20] proposed three new algorithms that employed collaborative filtering methods by weighting activities (e.g., comments, information shared, and forwarded) to existing networks. These algorithms can recommend users with user activities to the target users. Their algorithms apply user behavior to the process of user influence or recommended analysis, and the analysis of user behavior can be applied to link prediction [38]. However, their algorithms cannot directly apply to our scenario in which its hard to extract user relationship features for link prediction.
* Li et al. [40] presented a prediction algorithm by combining the different roles of users with common neighbor similarity index. Martłnez et al. [41] proposed a link prediction method to exploit the existence of a relationship between the best-performing degree of penalization for shared neighbors and the network clustering coefficient. These approaches extract user relationship features to link prediction which are complementary to ours.

**Disadvantages**

* + It is not based on latent Dirichlet allocation Scheme.
  + There is Absence of Joint dependence that represents the joint influence of common neighbors.

**PROPOSED SYSTEM**

* In the proposed system, for the characteristic of user multiple interests in social networks, the LDA topic model is introduced. Considering the internal and external factors that affect link establishment, we optimize LDA and propose a 3-HBP model. The model can mine user latent interest distribution and predict the links among the users effectively.
* Owing to the power-law characteristics of user behavior, LDA is improved by Gaussian weighting. In this way, the negative impact of the topic distribution to the high-frequency users can be reduced and the expression ability of the interest can be enhanced. The effects of internal driving factors can be analyzed.
* Taking the impact of common neighbor dependencies in link establishment, the model can be extended with the hidden naive Bayesian algorithm. By quantifying the dependencies between common neighbors, we can analyze the effects of external driving factors and combine internal driving factors to link prediction.

**Advantages**

* An efficient system simultaneously uses user attributes and relationships to improve performance of both the user latent interest and link prediction.
* A Three-Level Hidden Bayesian Link Prediction is an effective due to latent Dirichlet allocation techniques.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV or later version

➢ RAM - 4 GB (min)

➢ Hard Disk - 40 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**Software Requirements:**

* Operating System - Windows XP or Later Versions
* Coding Language - Java/J2EE(JSP,Servlet)
* Front End - J2EE
* Back End - MySQL