Physica A 482 (2017) 507–523

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Meta-path based heterogeneous combat network link prediction

Jichao Li, Bingfeng Ge *, Kewei Yang, Yingwu Chen, Yuejin Tan

College of Information System and Management, National University of Defense Technology, Changsha, China

HIGHLIGHTS

- We propose a novel integrated methodology framework, called HCNMP (heterogeneous combat networks link prediction based on meta-path), to predict links of multiple types simultaneously in heterogeneous combat networks (HCN).
- HCNMP can exploit not only the observed structural linkage information but also the semantic information of meta-path features in HCNs.
- HCNMP can achieve a very good performance for multiple link prediction with insufficient observed information.

ARTICLE INFO

Article history:
Received 1 November 2016
Received in revised form 17 February 2017
Available online 27 April 2017

Keywords:
Link prediction
Heterogeneous combat network (HCN)
Meta-path

ABSTRACT

The combat system-of-systems in high-tech informative warfare, composed of many interconnected combat systems of different types, can be regarded as a type of complex heterogeneous network. Link prediction for heterogeneous combat networks (HCNs) is of significant military value, as it facilitates reconfiguring combat networks to represent the complex real-world network topology as appropriate with observed information. This paper proposes a novel integrated methodology framework called HCNMP (HCN link prediction based on meta-path) to predict multiple types of links simultaneously for an HCN. More specifically, the concept of HCN meta-paths is introduced, through which the HCNMP can accumulate information by extracting different features of HCN links for all the six defined types. Next, an HCN link prediction model, based on meta-path features, is built to predict all types of links of the HCN simultaneously. Then, the solution algorithm for the HCN link prediction model is proposed, in which the prediction results are obtained by iteratively updating with the newly predicted results until the results in the HCN converge or reach a certain maximum iteration number. Finally, numerical experiments on the dataset of a real HCN are conducted to demonstrate the feasibility and effectiveness of the proposed HCNMP, in comparison with 30 baseline methods. The results show that the performance of the HCNMP is superior to those of the baseline methods.

* Corresponding author.
E-mail address: bingfengge@nudt.edu.cn (B. Ge).

http://dx.doi.org/10.1016/j.physa.2017.04.126

1. Introduction

The rapid development of information technologies has profoundly modified the operational patterns of modern warfare. As a distinct operational concept, network-centric warfare attempts to develop and leverage information superiority through networks of intelligence sensors, command and control systems, and precision strike weapons, thus translating acquired information into competitive advantages [1]. Combat system-of-systems (CSoS) in high-tech informative warfare is composed of different types of distributed combat systems that interact with each other to meet the requirements of combat operations [2]. Therefore, a network-centric CSoS can be considered as a complex heterogeneous combat network (HCN), in which different nodes and edges represent all types of combat systems and their interconnections, respectively.

The military intelligence is vital in conducting military operations, for instance, gathering an accurate intelligence of the network topology of the enemy CSoS guarantees reliability in subsequent critical operational units analysis and military data mining [3]. However, owing to the complexity and uncertainty of the battlefield environment, it is costly, difficult and even impossible to obtain completely accurate intelligence about military networks; there will be some missing information (i.e., missing links) in the acquired enemy network topology. If we could predict or identify them in advance, it would be helpful in both optimizing the military organizational structure of our side and attacking the critical operational components on the other side, which would definitely enhance the accuracy of military decisions and expedite victory [4]. As a special kind of heterogeneous network, the complexity due to structural dependency and heterogeneity of links produces obstacles for link prediction in HCNs. Well-known topological features designed for homogeneous networks are difficult to apply in such complex situations. In this regard, this paper focuses mainly on recovering the missing links to reconfigure combat networks and facilitate the appropriate representation of a real-world network topology according to the observed information, which is also called the "link prediction" problem [5–8].

For traditional link prediction problems, researches generally focus on predicting the links of homogeneous networks with single network architecture [9]. The related methods can generally be divided into two categories: unsupervised and supervised [6]. In [10], the unsupervised methods, such as common neighbors, preferential attachment index, Jaccard's coefficient, Adamic–Adar index, the Katz measure, and random walk with restart, are thoroughly summarized. Most of the unsupervised methods are based on the structural proximity of the nodes in a network, and do not make full use of other information. Many supervised algorithms have been proposed for taking advantage of the predetermined labeled data and the attribute information of nodes for link prediction [11]. A learning algorithm called Supervised Random Walks was developed for link prediction in social networks, by combining the network structural information with the level attributes of nodes and edges [12]. Researches on factorization machines (FM) inference algorithms applied in recommender systems are summarized in [13], and Kim et al. [14] combines the expectation–maximization (EM) algorithm with the Kronecker graphs model to solve the link-inferring problem. Several studies have exploited the temporal and spatial information in supervised link prediction [15,16]. However, the combat network studied in this paper is a heterogeneous network with various types of nodes and links. The foregoing studies mainly focus on predicting a single type of link in homogeneous networks, while this study aims to predict multiple types of HCN links simultaneously.

The heterogeneity of objects and links makes it difficult to use well-known metrics and algorithms for solving link prediction problems in heterogeneous networks. Recently, some studies are conducted on the link prediction in heterogeneous networks [17]. The meta-path based supervised link prediction method, first proposed by Sun et al. [18], employs a two-step process; that is, the first step is to extract the meta path based feature vectors, while the second step is to train a regression or classification model to compute the existence probability of a link [18–22]. The probabilistic models are also widely applied for link prediction tasks in heterogeneous networks. For a given heterogeneous network, the probabilistic model will optimize a built target function to establish a model composed of a group of parameters, which can best fit the observed data of the target network. The probability of the existence of a non-existent link is then estimated by the conditional probability. The probabilistic models include multi-relational influence propagation (MRIP) [23], topical factor graph model (TFGM) [24], and probabilistic matrix factorization (PMF) [25].

However, all these works focused on heterogeneous networks in social or biological domains; few researchers have studied the link prediction problems in the military field. Moreover, for most social networks or biological networks, link prediction usually addresses the fundamental question as to whether a link will form between two nodes in the future, while for HCN, the focus is on predicting the missing links, i.e., existent yet unknown links. In an actual battlefield environment, it is very difficult to obtain sufficient information on HCNs; therefore, the acquired topological information of HCN is usually incomplete and HCNs possess sparse features. A basic assumption in most traditional link prediction methods is that sufficient information is available in heterogeneous networks. Therefore, such an assumption can be invalidated when dealing with HCNs that are extremely sparse [26]. Even though social networks also possess sparse features, few of them can take advantage of the newly inferred results to predict the missing links. In the previous literatures, the target is to predict one type of link rather than to predict multiple types of links simultaneously.

In light of these differences and challenges, this study tries to solve the problem of the simultaneous prediction of multiple types of links for an HCN using an integrated methodology framework, HCNMP. The major contributions of this study are summarized as follows:

(1) We create a meta-path features based HCN link prediction model, HCNMP, to predict all types of links for the HCN, simultaneously. The HCNMP can exploit not only the observed structural linkage information, but also the semantic information of the meta-path features in HCNs.