Expert Systems With Applications 83 (2017) 114–121

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Ensemble method to joint inference for knowledge extraction

Yongbin Liu[a,b], Chunping Ouyang[a,*], Juanzi Li[b]

[a] College of Computer Science and Technology, University South China, Hengyang 421001, China
[b] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 25 October 2016
Revised 16 March 2017
Accepted 18 April 2017
Available online 22 April 2017

Keywords:
Ensemble learning
Joint inference
Knowledge extraction
Markov logic network

ABSTRACT

Joint inference is a fundamental issue in the field of artificial intelligence. The greatest advantage of the joint inference is demonstrated by its capability of avoiding errors from cascading and accumulating on a pipeline of multiple chained sub-tasks. Markov Logic Network(MLN) is the most common joint inference model that provides a flexible representation and handles uncertainty. It has been applied successfully to joint inference on many natural language processing tasks to avoid error propagation. However, due to the great expressiveness of first-order logic, the representation for it in MLN generates rather complicated graph structures, which makes the learning and inference on large scale data intractably. In this paper, we present an ensemble learning approach to deal with the challenges in MLNs. Firstly, we give a proof within the probably approximately correct (PAC) framework. The proof points out what conditions are necessary for successful applying the ensemble learning approach to MLN. Secondly, the paper explains how to combine the learners. Finally, in order to illustrate the working mechanism of the ensemble joint inference model, we present an Ensemble Markov Logic Networks (EMLNs) method and use it to extract knowledge from a large scale corpus published by Google.[1] Experiments suggest that significant speedup can be gained by the EMLNs. Meanwhile, it show that this approach leads to a higher precision and recall than that of those pipeline approaches.

1. Introduction

Recently, joint inference has been successfully applied to many fields. One of the most famous models is Markov logic network (MLN) (Domingos & Lowd, 2009; Richardson & Domingos, 2006), which is an uncertainty extension of first order logic. Markov logic bridges the gap between first-order logic and probability graphical model. The MLN has been successfully employed in a number of NLP tasks, including joint information extraction (Poon & Domingos, 2007; 2009), jointly identifying senses (Meza-Ruiz & Riedel, 2009), and jointly WSD and SRL (Che & Liu, 2010).

Despite the obvious benefits of the joint inference based on MLNs, a main difficulty which the previous works encountered is the scalability of Markov logic networks. The reason is that the learning and inference of the graph model are slower than that of other methods. The cause of the slow is first order logic formulae connect too many nodes in the graph. The size of the largest clique (factor) and the number of nodes (random variables) in the Markov networks (factor graph) grow quite rapidly with the increase of corpus, even though the formulae are fixed. A Markov network is generated by grounding a Markov logic network. The grounding is the process of creating a network for finding a probable state in this world. In a factor graph representation of a ground MLN, the number of factors will grow exponentially with the number of ground atoms. The ground network will be very large. Even approximate inference in the network involving large evidence datasets remains intractable. For example, Fig. 1 shows the learning time by the MLNs (CPU 2.70 GHz, RAM 32.0 GB). No matter what approximate inference is used,the learning and inference are not easy. So, the scalability of the joint model based on the Markov logic is a severe challenge. Next, a lot of works will be considered to speed up the learning time.

We take a fresh look at this challenge as an ensemble learning problem (Opitz & Maclin, 1999) for the first time. Ensemble learning is a learning paradigm where a finite numbers of learners are collected for the same task (Krogh, 1996). However, The joint model is performed on multiple tasks. Why can we apply ensemble learning method to a joint model? How do we conduct the ensemble joint model? In this paper, we figure out the answers to these problems and extend ensemble learning to the more general scenario of the inference on multiple different tasks. Firstly, we address what conditions can make the ensemble learning apply in joint inference within the probably approximately correct (PAC)

* Corresponding author.
E-mail addresses: yongbinliu03@gmail.com (Y. Liu), ouyangcp@126.com (C. Ouyang), juanzili@tsinghua.edu.cn (J. Li).

[1] code.google.com/p/relation-extraction-corpus/.

http://dx.doi.org/10.1016/j.eswa.2017.04.036

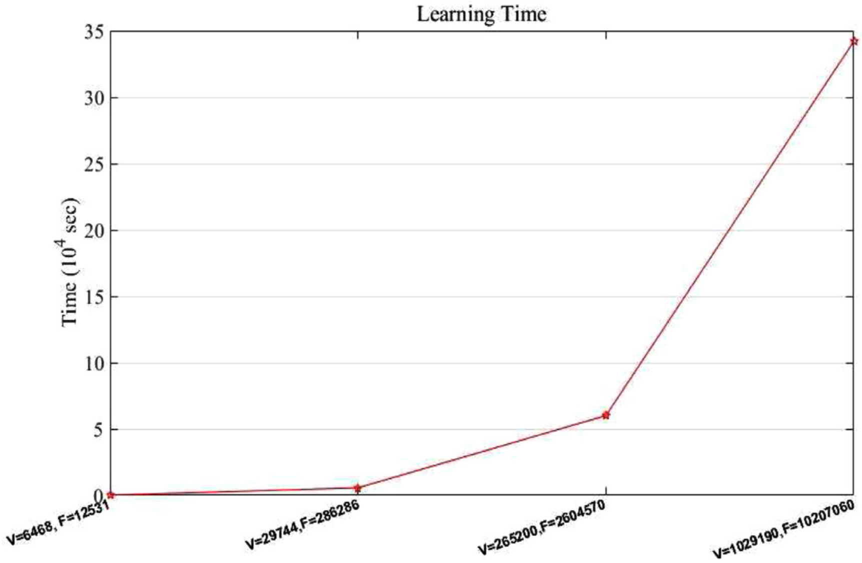

Fig. 1. Learning time: development number of variables and factors.

framework. Then we figure out how to combine the learners. Finally, We choose the Markov Logic Networks (MLNs) to achieve the joint inference and present a novel Ensemble Markov Logic Networks (EMLNs) model. The key of our method is the embedding of the Ensemble Markov Logic Networks (EMLNs). In the EMLNs, we adopt the Markov logic networks to joint inference on multi-tasks, and employ the divide and conquer strategy of ensemble learning to address the intractable inference of the Markov logic networks on large scale data. The experimental results on the large scale corpus published by Google show that our approach leads to a better performance than that of the pipeline systems.

The main contributions of this paper can be summarized as follows:

- Traditionally, the probably approximately correct (PAC) learning refers the single concept class. We discuss the PAC framework of the multiple tasks in the joint inference model. And we extend PAC learning to multi-concept classes.
- We present an ensemble learning approach to joint inference on the three NLP sub-tasks. We explain how to combine those weak learners to a strong ability and present the dynamic weighted combination method in the ensemble joint inference model.
- Our Ensemble Markov Logic Networks (EMLNs) address the problem of the Markov Logic Networks intractable dealing with the large scale data. Experiments show that this approach leads to a higher precision and recall than that of pipeline approaches.

The rest of this paper is organized as follows. Section 2 gives a review of related works and Section 3 figures out joint model ensemble PAC learnability. The Ensemble Markov Logic Networks are proposed in Section 4 and in Section 5, we conduct the experiments on Google corpora, and Section 6 concludes this paper.

2. Related work

2.1. Joint inference

In recent years, there has been an increasing interest in joint inference of multiple natural language processing tasks (McCallum, 2009). The joint inference allows bi-directional information flow to avoid error propagation. McCallum and Jensen (2003) proposed undirected graphical models for joint information extraction. Roth and Yih (2007) employed integer linear programming (ILP) for global inference. Many other papers presented the joint inference models based on Markov logic for NLP tasks. For example, Meza-Ruiz and Riedel (2009) presented jointly identifying predicates, arguments and senses; Che and Liu (2010) explored jointly modeling on semantic role labeling and word sense disambiguation. Markov logic networks can be sufficiently expressive for general AI by first-order logic. It makes the hypothesis space of the learning and inference greater than that of posterior probability model. The learning and inference of Markov logic network are not easily dealt with. So, researchers have presented various approaches to improve learning and inference in the MLNs, e.g. reducing the size of the network (Mihalkova & Richardson, 2009; Shavlik & Natarajan, 2009), and paralleling inference on Markov logic networks(Beedkar, Del Corro, & Gemulla, 2013; Niu, Ré, Doan, & Shavlik, 2011). Although the approaches of reducing size can speed up the learning and inference process to a certain extent, but they still cannot adapt to large evidence dataset. The parallel methods suffer from the problem of the partitioning MLNs, which is too expensive in practice. A new problem arose: how do we find those partitions which are too expensive in practice even for a quite simple MLN (Ahmad, Halawani, & Albidewi, 2012; Zhang et al., 2015). So we hope to address the problem simply.

2.2. Ensemble learning

Ensemble methods which have become a hot topic since the 1990s, are the approaches that train multiple learners and then combine them. These methods had achieved great success in many real-world tasks(Zhou, 2012). The representative approaches of the ensemble learning are Boosting(Schapire, 1990) and Bagging(Breiman, 1996) that are two state-of-the-art learning approaches. Dietterich (2000) generalized the benefit of ensemble method to the three fundamental reasons: statistical issue, computational issue, and representational issue. Statistical reason: A learning algorithm often is viewed as exploring a best hypothesis f in large hypotheses space H. But it is impossible that every algorithm searches the whole space of hypotheses to find the best hypothesis. The reality is that these algorithms search an approximate result. From statistical point of view, the “average” of several different hypotheses can improve the approximation. As