ARTICLE IN PRESS
ARTINT:2883

Artificial Intelligence ••• (••••) •••–•••

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Semantic-based regularization for learning and inference

Michelangelo Diligenti *, Marco Gori, Claudio Saccà

Department of Information Engineering and Mathematics, University of Siena, Via Roma 56, Siena, Italy

ARTICLE INFO

Article history:
Received in revised form 13 August 2015
Accepted 26 August 2015
Available online xxxx

Keywords:
Learning with constraints
Kernel machines
FOL

ABSTRACT

This paper proposes a unified approach to learning from constraints, which integrates the ability of classical machine learning techniques to learn from continuous feature-based representations with the ability of reasoning using higher-level semantic knowledge typical of Statistical Relational Learning. Learning tasks are modeled in the general framework of multi-objective optimization, where a set of constraints must be satisfied in addition to the traditional smoothness regularization term. The constraints translate First Order Logic formulas, which can express learning-from-example supervisions and general prior knowledge about the environment by using fuzzy logic. By enforcing the constraints also on the test set, this paper presents a natural extension of the framework to perform collective classification. Interestingly, the theory holds for both the case of data represented by feature vectors and the case of data simply expressed by pattern identifiers, thus extending classic kernel machines and graph regularization, respectively. This paper also proposes a probabilistic interpretation of the proposed learning scheme, and highlights intriguing connections with probabilistic approaches like Markov Logic Networks. Experimental results on classic benchmarks provide clear evidence of the remarkable improvements that are obtained with respect to related approaches.

1. Introduction

This paper presents Semantic Based Regularization (SBR), a unified framework for inference and learning that is centered around the notion of a constraint and of the parsimony principle. Semantic Based Regularization bridges the ability of machine learning techniques to learn from continuous feature-based representations with the ability of modeling arbitrary pattern relationships, typically used in Statistical Relational Learning (SRL) to model and learn from high-level semantic knowledge. In order to provide a unified context for manipulating perceptual data and prior knowledge, we propose to use the unifying concept of a *constraint*, which is sufficiently general to represent different kinds of sensorial data along with their relations, as well as to express abstract knowledge on the tasks. We unify continuous and discrete computational mechanisms, so as to accommodate in the same framework very different stimuli. In this paper, we focus on the kernel machine mathematical and algorithmic apparatus to learn from feature-based pattern representations and on constraints resulting from a fuzzy translation of First Order Logic (FOL) formulas, expressing the prior knowledge about the learning task at hand.

More specifically, SBR builds a multi-layer architecture having kernel machines at the input layer. The output of the kernel machines is fed to the higher layers implementing a fuzzy generalization of the FOL knowledge. Thanks to the

* Corresponding author.
E-mail addresses: diligmic@diism.unisi.it (M. Diligenti), marco@diism.unisi.it (M. Gori), sacc@unisi.it (C. Saccà).

http://dx.doi.org/10.1016/j.artint.2015.08.011

basic properties of fuzzy FOL and kernel machines, the resulting model is continuous with respect to the feature values. Therefore, the high-level semantic inference provided by the logic can be back-propagated down to the kernel machines using any gradient-based schema. This process can be iterated during training until convergence. This is an extremely powerful technique to get advantage of the available unsupervised data, as the inference process performed on this data via the logic knowledge can be used to correct the output of the kernel machines.

We substantially extend earlier studies in Diligenti et al. [10] by showing that SBR enables new fundamental tasks of learning and inference that rely on the joint informative evidence coming from real-valued features and simple pattern identifiers, along with the corresponding relations. In particular, the paper gives the following new main results, which are of fundamental importance to gain an overall view of theory and, especially, to enable a large set of applications in statistical relational learning domains:

- VARIABLE DIMENSION DOMAINS AND NULL INPUTS We extend the SBR framework [10] to truly hybrid domains, where real-valued feature pattern representations are integrated with pure symbolic entities (e.g. pattern identifiers). Indeed, in complex relational classification tasks, it is often the case that the entities are naturally representable by pattern spaces of different dimensions, including the remarkable case of “void patterns” in which only relational information is available.
- COLLECTIVE CLASSIFICATION In this paper we propose a novel collective classification method to enforce the constraints on the test set, thus exploiting the full expressiveness of FOL, like in other statistical relational learning (SRL) approaches. Once again, the distinctive feature of the solution proposed in this paper arises when considering that the collective computational scheme also naturally exploits real-valued feature pattern representations.
- PROBABILISTIC LINKS We extend studies on the probabilistic interpretation of regularization networks [38] to our case of learning from constraints. From one side, this highlights connections with Markov Logic Networks (MLNs) [40], while from the other side, this interpretation clearly shows the natural integration of real-valued features and object identifiers in SBR.

Furthermore, the paper presents how plain SVM, Transductive, and Laplacian SVMs can be derived as special cases of the proposed SBR framework. The paper also introduces new heuristics, connected to the ones employed in constraint satisfaction programming, to improve the quality of the found solutions. Finally, we present experimental results to show the effectiveness and generality of the approach.

The paper is organized as follows: in the next section previous work in the field is reviewed. Section 3 introduces First Order Logic and its fuzzy extensions, while Section 4 discusses learning from constraints with kernel machines. Section 5 presents how SBR generalizes several models commonly used in relational and transductive learning. Details on how training is performed in the SBR framework is presented in Section 6. In Section 7 a collective classification approach for SBR is presented and Section 8 presents connections between SBR and probabilistic models like Markov Logic Networks. The experimental evaluation of SBR is presented in Section 9 and, finally, Section 10 draws some conclusions.

2. Previous work

Statistical Relational Learning (SRL) combines robust parameter estimation in the presence of noise with learning complex relational structures. Probabilistic Relational Models (PRMs) [13] are an early SRL approach that learns a statistical model from a relational database. PRMs build a probability distribution over the attributes of the objects as an instance of a schema. A Bayesian network with one node for each attribute is built and parameters are estimated from the data. Relational Dependency Networks [34] learn a (local) conditional probability distribution for each node given its Markov blanket by using a conditional learner (like logistic regression or decision trees).

Markov Logic Networks (MLNs) [40] have received a lot of attention in the SRL community and have been extensively applied in many fields like bioinformatics [28] and computer vision [46]. Markov Logic Networks generalize and combine first-order logic and probabilistic graphical models. Thanks to their flexibility, MLNs have been used to tackle all the SRL main tasks: collective classification, link prediction, link-based clustering, social network modeling, and object identification. Many papers have also studied how to learn the structure of Markov Logic Networks from data without requiring an expert to express the structure in terms of prior knowledge [23,22]. Hybrid Markov Logic Networks (HMLNs) [49] extend MLNs to deal with continuous variables.

Probabilistic Soft Logic (PSL) [5] is another SRL approach, which relaxes MLNs to continuous fuzzy values in the $[0, 1]$ interval and restricts the considered FOL formulas to the ones with conjunctive body and a single literal head. PSL weight training can be solved via a convex optimization problem, but it can face only a small subset of the tasks that are potentially solved by a MLN.

One disadvantage of both MLNs and PSL in real-world applications is how they deal with entities that are associated to complex feature-based representations. Let’s take as an example the common scenario of a multi-class classification task where the patterns are represented by large vectors of numeric features. In order to perform learning and inference in this domain using classical SRL techniques, different approaches are possible:

Please cite this article in press as: M. Diligenti et al., Semantic-based regularization for learning and inference, Artificial Intelligence (2015), http://dx.doi.org/10.1016/j.artint.2015.08.011