

Citation: Kastrati, M., Biba, M., "Statistical Relational Learning: A State-of-the-Art Review". Journal of Engineering Technology and Applied Sciences 4 (3) 2019 : 141-156.

STATISTICAL RELATIONAL LEARNING: A STATE-OF-THE-ART REVIEW

Muhamet Kastrati^{a*} , Marenglen Biba^a

^a *Department of Computer Science, Faculty of Informatics, University of New York in Tirana, Albania*

*muhamet.kastrati@gmail.com (*corresponding author), marenglenbiba@unyt.edu.al*

Abstract

The objective of this paper is to review the state-of-the-art of statistical relational learning (SRL) models developed to deal with machine learning and data mining in relational domains in presence of missing, partially observed, and/or noisy data. It starts by giving a general overview of conventional graphical models, first-order logic and inductive logic programming approaches as needed for background. The historical development of each SRL key model is critically reviewed. The study also focuses on the practical application of SRL techniques to a broad variety of areas and their limitations.

Keywords: Statistical relational learning, probabilistic graphical models, inductive logic programming, probabilistic inductive logic programming

1. Introduction

There is a great interest in learning on propositional data which is assumed to be identically and independently distributed (i.i.d.). Very often, this independence assumption fail to comply in real-world applications where data may be characterized by the presence of both uncertainty and complex relational structure. Domains where data is non-i.i.d. are widespread; such as the internet and the world-wide web, scientific citation and collaboration, epidemiology, communication analysis, metabolism, ecosystems, bioinformatics, fraud and terrorist analysis, citation analysis, and robotics [16].

SRL [28] also known as probabilistic inductive logic programming (PILP) [16] is based on the idea that we can build models that can effectively represent, reason and learn in domains with presence of uncertainty and complex relational structure. In doing so, it addresses one of the major concerns of artificial intelligence, where key principle is integration of probabilistic reasoning, first-order logic and machine learning [16]. As shown in Figure 1, SRL combines a logic-based representation with probabilistic modeling and machine learning. SRL models are

usually represented as combination of probabilistic graphical models (PGMs) with first-order logic (FOL) to handle the uncertainty and probabilistic correlations in relational domains.

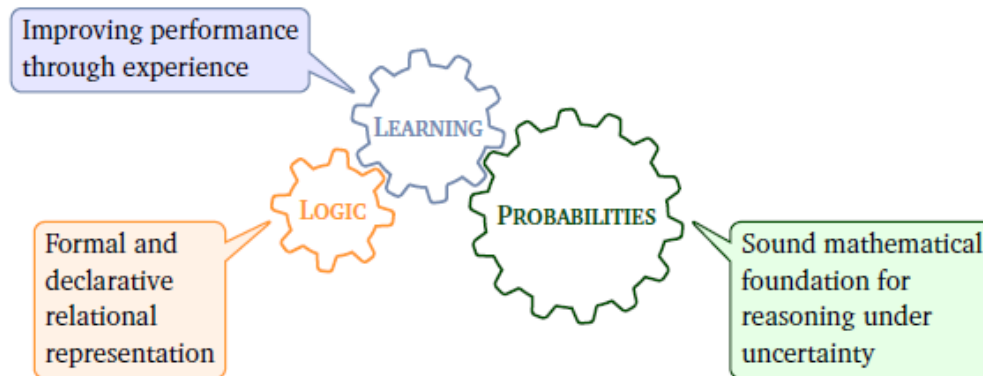


Figure 1: Statistical relational learning aka. probabilistic inductive logic programming combines probability, logic and learning. Adopted from [68].

In recent years, many formalisms and representations have been developed in SRL. Muggleton [42] and Cussens [6] upgraded stochastic grammars towards stochastic logic programs. Friedman [25] combined advantages of relational logic with Bayesian networks (BNs). Kersting et.al [34] combined definite logic programs with BNs. Neville and Jensen [52] extended dependency networks to relational dependency networks. Taskar et al. [71] extended Markov networks (MNs) into relational Markov networks (RMNs), and Domingos and Richardson [18] into Markov logic networks (MLNs).

Recently, many works have been done about further development of new techniques and application of SRL models in science and industry. Therefore, the goal of this paper is to provide a state-of-the-art review of SRL models, in the same time to contribute to research community with a cohesive overview of state-of-the-art results for a wide range of SRL problems (during the five years) and to identify possible opportunities for future research.

The rest of this paper is structured as the following: The second section introduces some background theory and notation, starting by the concepts of PGMs, FOL and Inductive Programming Language. The third section explores recent research into a range to SRL from its origins to the present day, including a discussion of relational models and of the successful role of the SRL problems/limitation and application in real world problem. The last section provides concluding remarks and some recommendations.

2. Background and notation

This part gives some background theory and notation as needed to facilitate understanding of the SRL models and techniques.

2.1 Probabilistic graphical models

PGMs [30] aka. graphical models are tightly integration of probability and graph theory. PGMs are probabilistic models for which a graph represents the joint probability distribution

over a large number of interdependent random variables. In this way, they provide an elegant framework for dealing with uncertainty, independence and complexity.

PGMs are characterized by the following properties:

- they provide simple and useful ways to visualize the structure of a probabilistic model and can be used to design new models
- can go deeper into the properties of the model, that is conditional independence
- complex computations about inference and learning are expressed through graphical manipulation and mathematical formulations.

Bayesian networks and MNs are the two most well known types of graphical models [36], Bayesian networks are directed graphical models consisting from links that have a particular directionality denoted by arrows. On the other hand, MNs are the type of undirected graphical models where the links do not own arrows. Key aspects of these two types of graphical models are provided in the following. Further details about these models can be found in [36], [37], [30], [46].

2.2 Bayesian networks

BNs [1] also known as Bayes Nets or Belief Nets, is a directed acyclic graph (DAG). Let \mathcal{G} be BNs graph over the variables X_1, \dots, X_n , than each node in the graph represents a random variable X_i , and edges between the nodes represent conditional probability dependencies among the corresponding random variables. The conditional probability distribution (CPD) for X_i , given its parents in the graph (Pa_{X_i}), is $P(X_i | Pa_{X_i})$.

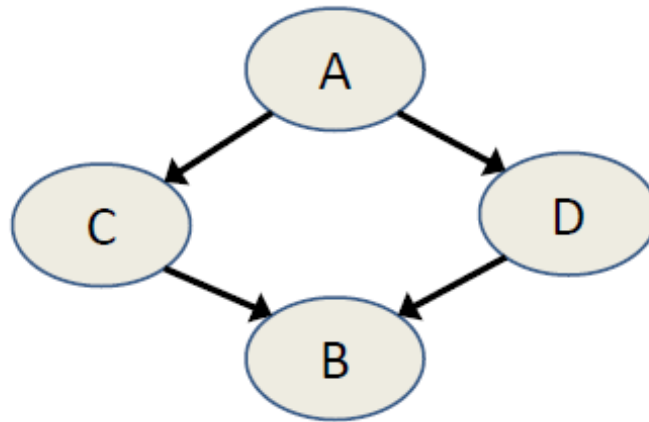


Figure 2: A simple example of a BN graph structure. Adopted from [2].

As defined by [36], let \mathcal{G} be a BN graph over the variables X_1, \dots, X_n . It can be said that a distribution P_B over the same space factorizes according to \mathcal{G} if P_B can be expressed as a product:

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_{X_i}) \quad (1)$$

A BN is a pair $(\mathcal{G}, \theta_{\mathcal{G}})$, where P_B factorizes over \mathcal{G} , and where P_B is specified as set of CPDs associated with \mathcal{G} s nodes, denoted $\theta_{\mathcal{G}}$.

The equation above is known as the chain rule for BN. It provides an efficient method for determining the probability of any complete assignment to the set of random variables. Each factor represents a conditional probability of the corresponding variable given its parents in the network [36].

An example of directed graphs which describes probability distributions is illustrated in Figure 2. Here, $P(A, C, D, B)$ represents an arbitrary joint distribution over four variables A, C, D, B .

According to the chain rule for BNs, the joint probability of all the nodes in the graph above can be expressed using the Equation 2.

$$P(A, C, D, B) = P(A) * P(C|A) * P(D|A, C) * P(B|A, C, D) \quad (2)$$

Equation 2, after using conditional independence relationship can be formulated as in Eq. 3.

$$P(A, C, D, B) = P(A) * P(C|A) * P(D|A) * P(B|C, D) \quad (3)$$

This is achieved because D is independent of C and B is independent of A .

The conditional independence relationships provide a more compact way of representing the joint distribution. If n is a binary node, then the full joint requires $O(2^n)$ space to represent this node. On the other hand, the factored form requires $O(2^{\max|Par(X_i)|})$ space to represent the same node, where, $Par(X_i)$ shows the set of parents of the variable X_i .

2.3 Dependency networks (DNs)

DNs [29] are graphical models for probabilistic relationship, and represent an alternative to the BNs. The graph of a dependency network, not similar to BNs, is potentially cyclic, but probability component of a dependency network, like a BNs, is a set of conditional distributions over multiple random variables, one for each node given its parents. The authors in [29] stated that there exist algorithms capable to learn the structure and the probabilities of a dependency network from data in a straightforward and efficient way in terms of computational complexity. From the graphical perspective, DN integrate the characteristics of directed and undirected models and allow presence of bi-directional links among variables.

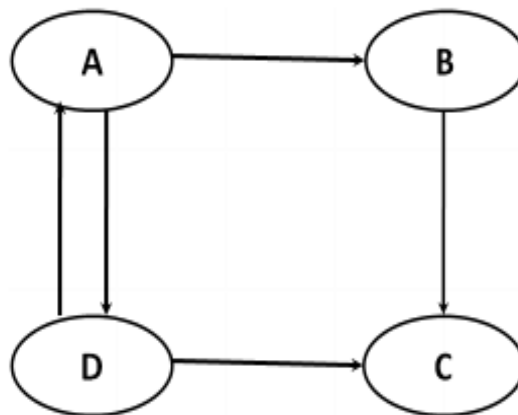


Figure 3: An example of dependency network. Adopted from [50].

2.4 Markov networks (MN)

MN also referred as Markov random field (MRF) [36] is an undirected graphical model which represents the joint probability distribution over events denoted with a set of variables $X = P(X_1, X_2, \dots, X_n) \in \chi$. It contains an undirected graph \mathcal{G} and a set of potential functions ϕ_k for each clique in the graph where the nodes are random variables. MN provides a different and often simpler view compared to directed models, both, in terms of the independence structure and the inference task [36], [2]. The output of potential function is typically a non-negative real-value of the state of the corresponding clique, obtaining the full joint distribution indicated by MN given in Equation 4.

$$P(X = x) = \frac{1}{Z} \prod_k \phi_k(x_{(k)}) \quad (4)$$

where $x_{(k)}$ shows the state of the k_{th} clique.

The partition function Z is given in Equation 5.

$$Z = \sum_{x \in \chi} \prod_k \phi_k(x_{(k)}) \quad (5)$$

A clique is a subset of the nodes in a graph, in which, there exists a connection link between all pairs of nodes in the subset. The graphical structure in Figure 4 includes two maximal cliques A, C, B and A, D, B and the rest of the cliques are not maximal A, C, A, D, A, B, D, B, C, B [2].

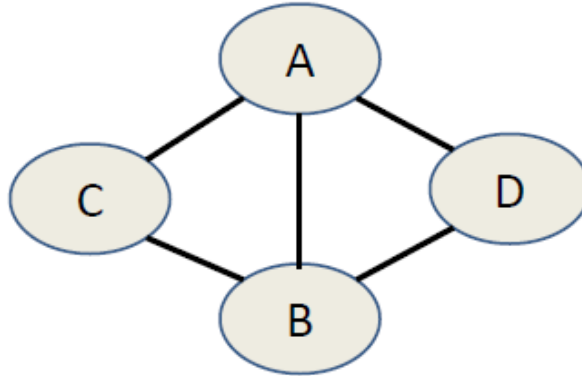


Figure 4: A simple example of a MN. Adopted from [2].

Any MNs parameterized using positive factors can be converted to into log-linear model leading to:

$$P(X = x) = \frac{1}{Z} \exp(\sum_j w_j f_j(x)) \quad (6)$$

where, it is one feature corresponding to each possible state x_k of each clique and w_j is weight of feature j .

2.5 First-order logic

A first-order knowledge base (KB) is defined as a set of expressions or formulas in FOL [26] that is a rich system capable to reasoning about objects and relationships.

Formulas typically consist of four types of symbols:

- Constants: represent objects within the domain of interest
- Variables: represent range over the objects within the domain
- Functions: represent mappings from tuples of objects to objects
- Predicates: represent syntactic symbols, with a given arity, that denotes either properties of a given object or relations among objects

Predicates contain Boolean truth value (*True* or *False*) showing if the predicate holds or not a value. Terms, atoms and formula that have no variables are called *ground*. A literal is an atomic formula or its negation. A formula in conjunctive normal form (CNF) represents a conjunction of *clauses*, in the other side, each clause represents a disjunction of literals.

A Horn clause is a logical formula that contains at most one positive literal. If a Horn clause has exactly one positive literal it is known as definite clause and it is written as in the following expression:

$$head \Leftarrow l_1 \vee \dots \vee l_m$$

The set of literals l_j represents the body of the clause. In the other hand a clause that has no head is known as a fact. A set of Horn clauses with exactly one head each is a definite logic program, and it is often used in ILP to represent the model (hypothesis). A collection of implicitly conjoined formulae is a knowledge base:

$$KB: = \bigwedge_{i=1} (F_i)$$

An "interpretation" is a model for a formula F if it satisfies the formulae. Let F and G be two formulae, then F logically entails G if all models of F belong to models of G . The process of checking whether a formula logically entails another, is known as the satisfiability problem for FOL. Further details of the FOL can be found in [2, 24, 26, 28, 64, 72].

2.6 Inductive logic programming (ILP)

ILP [43] and multi-relational data mining (MRDM) [22] is field of relational learning that integrates learning and logical programming. In the same time ILP serves as a foundation for many SRL methods. It represents a formal framework and offers practical algorithms for inductively learning relational descriptions from examples and background knowledge [16].

The major concern of ILP is to find a deterministic hypothesis H , in the form of a logical program from a set of positive (*Pos*) and negative (*Neg*) examples. For a given dataset D that covers n facts $x_i, i = 1, \dots, n$, denoted with a set of *Pos* and *Neg* examples that are part of class label y_i . For a given background knowledge B of logical formulae, and a language bias L that describes the hypothesis space, the goal of ILP is to induce a logical program H that entails all positive examples P and none of the negatives N . The learned hypothesis should be both complete and consistent with regard to the background knowledge B and the data D .

Overviews of inductive logic programming can be found in [20, 21, 41]. In the following a short introduction on the three main settings for learning in ILP is given. Further details about these three settings can be found in [15].

Learning from entailment: If H represents a causal theory and e denotes a clause, then the coverage relation as shown by [28] is defined as $covers(H, e)$, if and only if $H \models e$. Learning from entailment is probably one of the most used ILP setting and there are many well-known ILP systems such as FOIL [58], PROGOL [44] and ALEPH [70] using it.

Learning from interpretation: In learning from interpretations, if H represents a causal theory and e denotes a Herbrand interpretation as shown by [15] then we have expression as: $covers(H, e)$ if and only if $H \models e$ if and only if e with respect to the background theory B is a model of $B \cup H$. Learning from interpretations is generally simple and computationally easy compared to learning from entailment [17]. Simplicity of this method is related to fact that interpretations hold much more information compared to examples in learning from entailment. The approach used by learning from interpretations is similar to those that learn from entailment. The most important difference is in the generality relationship [2].

Learning from proofs: Learning from proofs was the first learning system applied to perform a kind of the learning in Model Inference System (MIS) [66]. As shown by [15] MIS normally fits within the learned from entailment setting where examples are facts. Inspired by the work of Shapiro on MIS, the authors in [15] presented the learning from proofs setting of ILP. In learning from proofs, if H stands for hypothesis and e for example and B for background theory, then hypothesis H covers example e in regard to B if and only if e is a proof-tree for $H \cup B$.

2.7 Probabilistic inductive logic programming

PILP [15] is a field that tries to combine ILP principles with statistical learning and the most natural way to do this is by extending ILP settings with probabilistic semantics. Further details on this issue can be found in [13, 15].

The main difference between PILP and ILP is that the *cover* relation becomes a probabilistic one and clauses will be explained with the help of probability values. If e stands for an example, H for an hypothesis and B stands for background theory, a probabilistic *covers* relation returns a probability P that can be given as:

$$cover(e, H \cup B) = P(e|H, B)$$

The latter is the likelihood of the example e and E is the set of examples. With this *cover* relation, goal of PILP is to find hypothesis H that provides maximum likelihood of the data given by expression as: $P(E|H, B)$.

3. Statistical relational learning models

Most of the work related to SRL can be grouped in two main research streams, one that starts with PGMs extending it with relational aspects, and another research stream in [12, 14, 15, 16], those who follow another approach, they start with inductive logic programming (ILP) and extend it with probabilistic semantics. In the following we present state-of-the-art review of statistical relational learning models.

3.1 Probabilistic relational models (PRMs)

PRMs [25] represent a rich representation language for structured statistical models [27]. PRMs combine advantages of relational logic with BNs. In contrast, from PILP approaches, the authors in [25] start from BNs and extend it with the concepts of objects, their properties, and relations between them, creating in these form a model that deals with both relations and uncertainty. One of the key contribution of authors in [25] was to present that some of the techniques of BNs learning can be further extended to the learning of these more complex models. It was this contribution used to generalize the ideas of [38] on this topic. More details on PRMs can be found at [25, 27, 28].

3.2 Stochastic logic programs (SLPs)

SLPs [42] generalize Hidden Markov Models (HMMs), context-free grammars and directed BNs [45]. According to [28] SLPs provide a simple scheme for representing probability distributions over structured objects. As defined by [45] a pure SLPs is specified by a set of labeled clauses $p:C$, where p is a probability and C is first order that has restricted range by a definite clause. The subset S_p of clauses in S with predicate symbol p in the head is known as the definition of p . The sum of probability labels π_p must be at most 1 for each definition S_p . In this case S will be complete if $\pi_p = 1$ for each p and incomplete otherwise. On the other hand the $P(S)$ denotes the definite program combining by all clauses in S , with labels removed [45].

3.3 Bayesian logic programs (BLPs)

BLPs [34] are a language used to integrate definite logic programs with BNs that results probability distributions over first-order interpretations. The main use of BLPs is to create an one-to-one mapping between ground atoms and random variables, and between the immediate consequence operator and the dependency relation. In this way, BLPs combine the advantages of both definite clause logic and BNs. Authors in [35] have presented results obtained by combining ILP learning from interpretations with BNs to learn from both the types of components (qualitative and the quantitative) of BLPs from data.

3.4 Relational dependency networks (RDNs)

RDNs [52] are another type of graphical models that represents extended DN for relational domains similarly, as RBNs that extend BNs [50]. Authors in [52] described RDNs compared to RBNs and RMNs and outlined the relative strengths of RDNs, the ability that RDNs to denote cyclic dependencies, simple approach for parameter estimation, and efficient structure learning techniques. The robustness of RDNs comes from the use *pseudolikelihood* learning techniques, that estimate an efficient approximation of the full joint distribution. In contrast to the conventional learning, the approach used for RDNs learn a single probability tree per each random variable. Furthermore authors in [50] proposed turning the problem into a number of relational function-approximation problems by applying gradient-based boosting. The experimental results obtained by [50] in several data sets showed that using boosting method has significant improvement on RDNs learning that is more efficient compared to other state-of-the-art statistical relational learning methods.

3.5 Relational markov networks

RMNs [71] are undirected graphical models that extend the framework of MNs to relational domains. Undirected models (MNs) address two limitations of the directed models. First, MNs are undirected models so there are free of the problem of cycles. Second, undirected models are more appropriate for discriminative training, where it is optimized the conditional likelihood of the labels for given features, that generally helps to improve classification accuracy. The authors in [71] showed how these models can be trained to be more efficient, and also they illustrated how to use approximate probabilistic inference over the learned model for collective classification and link prediction. A relational Markov network $M = (C, \Phi)$ is defined by a set of clique templates C and corresponding potentials $\Phi = \{\phi_c\}_{c \in C}$ in order to define a conditional distribution as shown in Equation 7:

$$P(I.y|I.x, I.r) = \frac{1}{Z(I.x, I.r)} \prod_{c \in C} \prod_{c \in C(I)} \phi_c(I.x_c, I.y_c) \quad (7)$$

$Z(I.x, I.r)$ represents the normalizing partition function:

$$Z(I.x, I.r) = \sum_{I.y} \prod_{c \in C} \prod_{c \in C(I)} \phi_c(I.x_c, I.y_c) \quad (8)$$

Authors in [71] provided some experimental results obtained on hypertext and social network domains, which shows that accuracy has been significantly improved by modeling relational dependencies.

3.6 Markov logic networks

MLNs [60] is a state-of-the-art SRL model that integrates FOL representation with MN modeling. MLNs represent a probabilistic extension of FOL with attaching weights to each formula (or clause) and are suitable to deal with complex data which are characterized with uncertainty. As defined by [60] a MLN L represents a set of pairs (F_i, w_i) , in which the expression F_i represents a formula in first-order logic, w_i represent real number and together with finite $C = \{c_1, c_2, \dots, c_{|C|}\}$ set of constants, it defines a MN $M_{L,C}$ as defined by Equation 9:

$$P(X = x) = \frac{1}{Z} \exp(\sum_i w_i n_i(x)) = \frac{1}{Z} \prod_i \phi_i(x_{(i)})^{n_i(x)} \quad (9)$$

Here, w_i denotes weight of formula i , $n_i(x)$ denotes the number of true groundings of formula F_i in x , $x_{\{i\}}$ is the state of the predicates present in F_i , and $\phi_i(x_{(i)}) = e^{w_i}$.

4. SRL limitation and applications

4.1 SRL limitations

Probably the computational complexity of inference is one of the biggest limitation present at most of the SRL methods, followed by the size of the graph G_I that is in proportion to the number of descriptive attributes and objects, which is another relevant factor that limits the scalability for many realistic datasets.

4.2 Application of SRL

Using statistical and relational information gives better results. The SRL community has had a lot of successes with application of SRL methods in a number of areas including link prediction [57] and document mining [56]. Lately there has been a significant interest in SRL methods, with many successful applications in a wide variety of complex problems including information extraction [61], [55], where SRL-based algorithms have been successfully implemented. Some well-known cases where SRLs methods have been successfully applied includes biomedical problems such as breast cancer [9] and drug activity prediction [10]. In addition, there is another learning method known as Relational Functional Gradient Boosting (RFGB) [50] that has been successfully applied by [74] for predicting heart attack, later there was another successful application fo SRL by [49] for predicting of cardiovascular risk and Alzheimer's disease prediction by [51]. SRL was also successfully applied by [11] for the first time in 3D building reconstruction, and to Recognize Textual Entailment by [62], and by [4] for modeling of large data streams using SRL techniques for classification. In the other hand, authors in [53] presented how a statistical models can be *trained* on large knowledge graphs to predict new facts about the world which is similar to predicting new edges in the graph [53]. An approach to Identifying Evidence Based Medicine Categories was presented by [73]. In their paper [75] proposed application of SRL approaches to hybrid recommendation systems. Recently, authors in [54] provide a comprehensive review of employing SRL for collaborative filtering task. At the same time, [54] demonstrated that exist strong evidence of successful application of SRL methods in the recommender system domains.

More recently, statistical relational learning models have demonstrated a great success in classification, handwritten recognition, collaborative filtering, recommendation systems and time series forecasting. Luperto et al. [40] proposed semantic classification by reasoning on the whole structure of buildings using SRL techniques. In this study, authors proposed a method for global reasoning on the whole structure of buildings, considered as single structured objects. They have used SRL algorithm, named kLog and compared it with another classifier, Extra-Trees, which resembles classical approaches, in three tasks: classification of rooms, classification of entire floors of building and validation of simulated worlds. The results obtained showed that their global approach outperforms local approaches when the classification task involves reasoning on the regularities of buildings and when available about rooms is coarse-grained.

Dai et al. [7] presented a new approach for detecting visual relationships with deep relational networks. In this study, authors proposed an integrated framework to tackle the difficulties caused by the high diversity of visual appearance for each kind of relationships or the larger number of distinct values phrases. This framework integrates deep relational network, which is designed specifically for exploiting the statistical dependencies between objects and their relationship. Results obtained by proposed method when used on two large data set, have reached significant improvements compared to state-of-the-art.

Yang et al. [76] proposed combining content-based and collaborative filtering for job recommendation system. In this paper, authors proposed a way to adapt the state-of-the-art in SRL approaches to construct a real hybrid job recommendation system. Furthermore, in order to satisfy a common requirement in recommendation systems, the proposed approach can also allow tuning the trade-off between the precision and recall of the system in a principled way. Through experimental results authors demonstrated the efficiency of their proposed approach as well as its improved performance on recommendation precision.

Natarajan et al. [48] proposed MLNs for adverse drug event extraction from text. Thus, in this work, authors addressed the question of whether they can quantitatively estimate relationships between drugs and conditions from the medical literature. This paper proposed and studied a state-of-the-art NLP-based extraction of ADEs from text.

Cohen et al. [5] proposed relational restricted Boltzmann machines (RBMs): A probabilistic logic learning approach. In this study, authors considered the problem of learning Boltzmann machine classifiers from relational data. The author's goal was to extend the deep belief framework of RBMs to statistical relational models. Empirical results showed that this approach of designing an RBM is comparable or even better than the state-of-the-art probabilistic relational learning algorithms on six relational domains.

Embar et al. [23] proposed scalable structure learning for probabilistic soft logic (PSL). Zhang et al. [77] proposed a model-based asset deterioration assessment framework represented by PRMs. The authors in this study have illustrated the use of the framework with multiple variants of deterioration models.

Kazemi and Poole [33] introduced Relnn: A deep neural model for relational learning. In this study, authors developed relational neural networks (RelNNs) by integrating hidden layers to relational logistic regression (the relational counterpart of logistic regression). In this study authors conducted some initial experiments on several tasks over three real-world datasets and results obtained showed that RelNNs are promising models for relational learning.

Sileo et al. [67] showed improving composition of sentence embeddings through the lens of statistical relational learning. In this study, authors were based on recent SRL models to address textual relational problems, showing that they are more expressive, and can alleviate issues from simpler compositions. The resulting models significantly improve the state of the art in both transferable sentence representation learning and relation prediction.

Katzouris et al. [31] presented online learning of weighted relational rules for complex event recognition. In this study, authors advanced the state-of-the-art by integrating an existing online algorithm for learning crisp relational structure with an online method for weight learning in MLNs. The authors here have evaluated this approach on a complex real-world application for activity recognition and result obtained showed that it performs better its crisp predecessor and competing online MLNs learners in terms of predictive performance.

Li et al. [39] proposed an unsupervised automatic data cleaning approach based on statistical relational learning. In this study, authors focused on cleaning dirty data without involving data quality patterns or human expert's interaction. This approach follows a series of steps, firstly, they learnt a model of data in the form of BN that represents the dependency relationships among different attributes of database table. After that, they translated the dependency relationship into first-order logic formulas, then form first-order logic formulas into MLNs by giving weigh for each formula. Secondly, they transformed MLNs into DeepDive inference rules and execute this steps in DeepDive framework. Here, the results of inference have been used to estimate the most likely repairs of dirty in data.

Dong et al. [19] presented second-order Markov assumption based Bayes classifier for networked data with heterophily. The authors here in this study describe the problems faced in case of classification of networked data. According to [19] most of the traditional relational classifiers which are based on the principle of homophily are characterized with

unsatisfactory classification performance, this is due to the fact that these methods consider inhomogeneous networks homogeneously. To address this problem, authors in this study proposed a progression of a network-only Bayes classifier based on second order Markov assumption for heterophilous networks. The experimental results obtained showed that the proposed method have better performance when the networks are heterophilous.

Das et al. [8] proposed fast relational probabilistic inference and learning by using approximate counting via hypergraphs. Their experimental results showed that the efficiency of these approximations allows these models to perform significantly faster than state-of-the-art, which can be successfully applied to several complex statistical relational models, without sacrificing effectiveness.

Speichert and Belle [69] proposed learning probabilistic logic programs in continuous and mixed discrete-continuous data. Rossi in [63] formulated the relational time series forecasting, Kazemi and Poole [32] introduced combining weighted rules with graph random walks for statistical relational models, Schlichtkrull et al. [65] introduced relational graph convolutional networks (R-GCNs) and applied them for link prediction (recovery of missing facts) and entity classification (recovery of missing entity attributes), Ravkic et al. [59] presented the graph sampling approach with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model. Mutlu et al. [47] presented a comprehensive review on graph feature learning and feature extraction techniques for link prediction. Bozcan and Kalkan [3] introduced COSMO: Contextualized scene modeling with Boltzmann machines.

5. Conclusions

Statistical relational models (SRL) methods have in most of the cases been very successful and significant progress is made over the last two decades. Many different problems have been defined using SRL models and good results have already been achieved; nevertheless, despite all these advances and successful application, the current SRL models surveyed in this paper still require further improvements to make SRL models more effective in a broader range of application. In this paper, we reviewed various limitation of the current SRL models and discussed possible extension that can be provide better learning capabilities. The computational complexity of inference is the most important limitation present among most of the SRL methods. The size of the graph G_I , that is directly linked (proportional) to the number of describing attributes and objects and in this way limits the scalability for many real world datasets. We hope that the issues presented in this paper will advance the discussion in the statistical relational learning community about the next generation of statistical relational learning techniques.

References

- [1] Ben-Gal, I., “Bayesian networks”, Encyclopedia of statistics in quality and reliability (2008).
- [2] Biba, M., “Integrating Logic and Probability: Algorithmic Improvements in Markov Logic Networks”. PhD thesis, University of Bari, Italy (2009).
- [3] Bozcan, B., Kalkan, S., “Cosmo: Contextualized scene modeling with boltzmann machines”, Robotics and Autonomous Systems (2019) : 132–148.

- [4] Chandra, S., Sahs, J., Khan, L., Thuraisingham, B., Aggarwal, C., “Stream mining using statistical relational learning”, In Data Mining (ICDM), IEEE International Conference on (2014), IEEE (2014) : 743–748.
- [5] Cohen, W., Natarajan, S., “Relational restricted boltzmann machines: A probabilistic logic learning approach”, In Inductive Logic Programming: 27th International Conference, ILP 2017, Orléans, France, September 4-6, 2017, Revised Selected Papers, volume 10759, Springer (2018) : 94.
- [6] Cussens, J., “Parameter estimation in stochastic logic programs”, Machine Learning 44(3) (2001) : 245–271.
- [7] Dai, B., Zhang, Y., Lin, D., “Detecting visual relationships with deep relational networks”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017) : 3076–3086.
- [8] Das, M., Dhami, D.S., Kunapuli, G., Kersting, K., Natarajan S., “Fast relational probabilistic inference and learning”, Approximate counting via hypergraphs (2019).
- [9] Davis, J., Burnside, E. S., de Castro Dutra, I., Page, D., Ramakrishnan, R., Vitor Santos Costa, V.S., Shavlik, J.W., “View learning for statistical relational learning: With an application to mammography”. In IJCAI, Citeseer (2005) : 677–683.
- [10] Davis, J., Ong, I.M., Struyf, J., Burnside, E.S., Page, D., Costa, V.S., “Change of representation for statistical relational learning”, In IJCAI (2007) : 2719–2726,.
- [11] Dehbi, Y., Hadiji, F., Gröger, G., Kersting, K., Plümer, L., “Statistical relational learning of grammar rules for 3d building reconstruction”, Transactions in GIS (2016).
- [12] De Raedt, L., Dietterich, T., Getoor, L., Muggleton, S.H., “Probabilistic, logical and relational learning-towards a synthesis”, In Dagstuhl Seminar Proceedings (2005).
- [13] De Raedt, L., Kersting, K., “Probabilistic logic learning”, ACM SIGKDD Explorations Newsletter 5(1) (2003) : 31–48.
- [14] De Raedt, L., Kersting, K., “Probabilistic inductive logic programming”, In International Conference on Algorithmic Learning Theory”, Springer (2004) : 19–36.
- [15] De Raedt, L., Kersting, K., “Probabilistic inductive logic programming”, In Probabilistic Inductive Logic Programming, Springer (2008) : 1–27.
- [16] De Raedt, L., Kersting, K., “Statistical relational learning”, In Encyclopedia of Machine Learning, Springer (2011) : 916–924.
- [17] De Raedt, L., “Logical settings for concept-learning”, Artificial Intelligence 95(1) (1997) : 187–201.
- [18] Domingos, P., Lowd, D., “Markov logic: An interface layer for artificial intelligence”, Synthesis Lectures on Artificial Intelligence and Machine Learning 3(1) (2009) : 1–155.
- [19] Dong, S., Liu, D., Ouyang, R., Zhu, Y., Li, L., Li, T., Liu, J., “Second-order markov assumption based bayes classifier for networked data with heterophily”, IEEE Access (2019).
- [20] Džeroski, S., Lavrač, N., “An introduction to inductive logic programming”, In Relational data mining, Springer (2001) : 48–73.
- [21] Džeroski, S., “Inductive logic programming in a nutshell”. Introduction to Statistical Relational Learning”, (2007).
- [22] Džeroski, S., “Relational data mining. Data Mining and Knowledge Discovery”, Handbook (2010) : 887–911.

- [23] Embar, V., Sridhar, D., Farnadi, G., Getoor, L., “Scalable structure learning for probabilistic soft logic”. arXiv preprint arXiv:1807.00973 (2018).
- [24] Fitting, M., “First-order logic and automated theorem proving”, Springer Science & Business Media (2012).
- [25] Friedman, N., Getoor, L., Koller, D., Pfeffer, A., “Learning probabilistic relational models”, In IJCAI volume 99 (1999) : 1300–1309.
- [26] Genesereth R.M., Nilsson J.N., “Logical foundations of artificial. Intelligence, Morgan Kaufmann 58 (1987).
- [27] Getoor, L., Friedman, N., Koller, D., Pfeffer, A., “Learning probabilistic relational models”, In Relational data mining, Springer (2001) : 307–335.
- [28] Getoor, L., “Introduction to statistical relational learning”, MIT press (2007).
- [29] Heckerman, D., Chickering M.D., Meek, C., Rounthwaite, R., Kadie, C., “Dependency networks for inference, collaborative filtering, and data visualization”, Journal of Machine Learning Research 1(Oct) (2000) : 49–75.
- [30] Jordan I.M., “Learning in graphical models”, volume 89, Springer Science & Business Media (1998).
- [31] Katzouris, N., Michelioudakis, E., Artikis, A., Paliouras, G., “Online learning of weighted relational rules for complex event recognition”, In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer (2018) : 396–413.
- [32] Kazemi, S.M., Poole, D., “Bridging weighted rules and graph random walks for statistical relational models”, Frontiers in Robotics and AI 5 (2018) : 8.
- [33] Kazemi, S.M., Poole, D., “ReINN: A deep neural model for relational learning”, In Thirty-Second AAAI Conference on Artificial Intelligence (2018).
- [34] Kersting, K., De Raedt, L., Kramer, S., “Interpreting Bayesian logic programs”, In Proceedings of the AAAI-2000 workshop on learning statistical models from relational data (2000) : 29–35.
- [35] Kersting, K., De Raedt, L., “Basic principles of learning Bayesian logic programs”, In Probabilistic Inductive Logic Programming Springer, Berlin, Heidelberg (2008) : 189–221.
- [36] Koller, D., Friedman, N., Getoor, L., Taskar, B., “Graphical models in a nutshell”, Introduction to statistical relational learning (2007) : 13–55.
- [37] Koller, D., Friedman, N., “Probabilistic graphical models: principles and techniques”, MIT press (2009).
- [38] Koller, D., Pfeffer, A., “Probabilistic frame-based systems”, In AAAI/IAAI, (1998) : 580–587.
- [39] Li, W., Li, L., Li, Z., Cui, M., “Statistical relational learning based automatic data cleaning”, Frontiers of Computer Science 13(1) (2019) : 215–217.
- [40] Luperto, M., Riva, A., Amigoni, F., “Semantic classification by reasoning on the whole structure of buildings using statistical relational learning techniques”, In 2017 IEEE International Conference on Robotics and Automation (ICRA) IEEE (2017) : 2562–2568.
- [41] Muggleton, S., De Raedt, L., ”Inductive logic programming: Theory and methods”, The Journal of Logic Programming 19 (1994) : 629–679.

- [42] Muggleton, S., “Stochastic logic programs”, *Advances in inductive logic programming*, 32 (1996) : 254-264.
- [43] Muggleton, S., “Inductive logic programming”, *New generation computing* 8(4) (1991) : 295–318.
- [44] Muggleton, S., “Inverse entailment and prolog”, *New generation computing* 13(3-4) (1995) : 245–286.
- [45] Muggleton, S., “Learning stochastic logic programs”, *Electron. Trans. Artif. Intell.*, 4(B) (2000) : 141–153.
- [46] Murphy, K., “A brief introduction to graphical models and bayesian networks”, (1998).
- [47] Mutlu, E.C., Oghaz, T.A., “Review on graph feature learning and feature extraction techniques for link prediction”, *arXiv preprint arXiv:1901.03425*, (2019).
- [48] Natarajan, S., Bangera, V., Khot, T., Picado, J., Wazalwar, A., Costa, V. S., Caldwell, M., “Markov logic networks for adverse drug event extraction from text”, *Knowledge and information systems* 51(2) (2017) : 435–457.
- [49] Natarajan, S., Kersting, K., Ip, E., Jacobs, D. R., Carr, J., “Early prediction of coronary artery calcification levels using machine learning”, In *Twenty-Fifth IAAI Conference* (2013).
- [50] Natarajan, S., Khot, T., Kersting, K., Gutmann, B., Shavlik, J., “Gradient-based boosting for statistical relational learning”, *The relational dependency network case*, *Machine Learning* 86(1) (2012) : 25–56.
- [51] Natarajan, S., et al. “Relational learning helps in three-way classification of Alzheimer patients from structural magnetic resonance images of the brain”, *International Journal of Machine Learning and Cybernetics* 5(5) (2014) : 659–669.
- [52] Neville, J., Jensen, D., “Relational dependency networks”, *Journal of Machine Learning Research* 8(Mar) (2007) : 653–692.
- [53] Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E., “A review of relational machine learning for knowledge graphs”, *Proceedings of the IEEE* 104(1) (2015) : 11–33.
- [54] Nishani, L., Biba, M., “Statistical relational learning for collaborative filtering a State-of-the-Art Review”, In *Natural Language Processing: Concepts, Methodologies, Tools, and Applications IGI Global* (2020) : 688-707.
- [55] Poon, H., Vanderwende, L., “Joint inference for knowledge extraction from biomedical literature”, In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics (2010) : 813–821.
- [56] Popescul, A., Ungar, L.H., Lawrence, S., Pennock, D.M., “Statistical relational learning for document mining”, In *Third IEEE International Conference on Data Mining*, IEEE (2003) : 275–282.
- [57] Popescul, A., Ungar H. L., “Statistical relational learning for link prediction”, In *IJCAI workshop on learning statistical models from relational data*, volume 2003 (2003).
- [58] Quinlan, J.R., “Learning logical definitions from relations”, *Machine learning* 5(3) (1990) : 239–266.
- [59] Ravkic, I., Žnidaršič, M., Ramon, J., Davis, J., “Graph sampling with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model”, *Data Mining and Knowledge Discovery* 32(4) (2018) : 913–948.

- [60] Richardson, M., Domingos, P., “Markov logic networks”, *Machine learning* 62(1-2) (2006) : 107–136.
- [61] Riedel, S., Chun, H.W., Takagi, T., Tsujii, J.I., “A markov logic approach to bio-molecular event extraction”, In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, Association for Computational Linguistics (2009) : 41–49.
- [62] Rios, M., Specia, L., Gelbukh, A., Mitkov, R., “Statistical relational learning to recognise textual entailment”, In *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Berlin, Heidelberg (2014) : 330–339.
- [63] Rossi, A.R., “Relational time series forecasting”, *The Knowledge Engineering Review* 33 (2018).
- [64] Russell, S.J., Norvig, P., “Artificial intelligence: a modern approach”, (1995).
- [65] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M., “Modeling relational data with graph convolutional networks”, In *European Semantic Web Conference*, Springer, Cham (2018) : 593–607.
- [66] Shapiro Y. E., “Algorithmic program debugging”, MIT press (1983).
- [67] Sileo, D., Van de Cruys, T., Pradel, C., Muller, P., “Improving composition of sentence embeddings through the lens of statistical relational learning”, (2018).
- [68] Skarlatidis, A., “Event recognition under uncertainty and incomplete data”, PhD thesis, Institute of Informatics (2014).
- [69] Speichert, S., Belle, V., “Learning probabilistic logic programs in continuous domains”, arXiv preprint arXiv:1807.05527 (2018).
- [70] Srinivasan, A., “The aleph manual”, (2001).
- [71] Taskar, B., Abbeel, P., Wong, M.F., Koller, D., “Relational markov networks”, *Introduction to statistical relational learning* (2007) : 175–200.
- [72] Teso, S., “Statistical Relational Learning for Proteomics: Function, Interactions and Evolution”, PhD thesis, University of Trento (2013).
- [73] Verbeke, M., Van Asch, V., Morante, R., Frasconi, P., Daelemans, W., De Raedt, L., “A statistical relational learning approach to identifying evidence based medicine categories”, In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics (2012) : 579–589.
- [74] Weiss, J.C., Natarajan, S., Peissig, P.L., McCarty, C.A., Page, D., “Statistical relational learning to predict primary myocardial infarction from electronic health records”, In *Twenty-Fourth IAAI Conference* (2012).
- [75] Yang, S., Korayem, M., AlJadda, K., Grainger, T., Natarajan, S., “Application of statistical relational learning to hybrid recommendation systems”, arXiv preprint arXiv:1607.01050 (2016).
- [76] Yang, S., Korayem, M., AlJadda, K., Grainger, T., Natarajan, S., “Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive statistical relational learning approach”, *Knowledge-Based Systems* 136 (2017) : 37–45.
- [77] Zhang, H., Marsh, D.W.R., “Towards a model-based asset deterioration framework represented by probabilistic relational models”, (2018).