Expert Systems with Applications 42 (2015) 4595–4609

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Symbolic-neural rule based reasoning and explanation

Ioannis Hatzilygeroudis [a,1,2], Jim Prentzas [b,*]

[a] University of Patras, School of Engineering, Department of Computer Engineering & Informatics, 26500 Patras, Greece

[b] Democritus University of Thrace, School of Education Sciences, Department of Education Sciences in Early Childhood, Laboratory of Informatics, 68100 Nea Chili, Alexandroupolis, Greece

ARTICLE INFO

Article history:
Available online 4 February 2015

Keywords:
Neuro-symbolic approaches
Integrated intelligent systems
Rule-based reasoning and explanation
Hybrid intelligent systems
Hybrid expert systems

ABSTRACT

In this paper, we present neurule-based inference and explanation mechanisms. A neurule is a kind of integrated rule, which integrates a symbolic rule with neurocomputing: each neurule is considered as an adaline neural unit. Thus, a neurule base consists of a number of autonomous adaline units (neurules), expressed in a symbolic oriented syntax. There are two inference processes for neurules: the connectionism-oriented process, which gives pre-eminence to neurocomputing approach, and the symbolism-oriented process, which gives pre-eminence to a symbolic backwards chaining like approach. Symbolism-oriented process is proved to be more efficient than the connectionism-oriented one, in terms of the number of required computations (56.47–63.88% average reduction) and the mean runtime gain (59.95–64.89% in average), although both require almost the same average number of input values. The neurule-based explanation mechanism provides three types of explanations: 'how' a conclusion was derived, 'why' a value for a specific input variable was asked from the user and 'why-not' a variable has acquired a specific value. As shown by experiments, the neurule-based explanation mechanism is superior to that provided by known connectionist expert systems, another neuro-symbolic integration category. It provides less in number (64.38–69.28% average reduction) and more natural explanation rules, thus increasing efficiency (mean runtime gain of 56.65–56.73% in average) and comprehensibility of explanations.

1. Introduction

Most of the intelligent methods have advantages as well as disadvantages. Research in artificial intelligence (AI) has shown that approaches integrating (or combining) two or more intelligent methods may provide benefits (Hatzilygeroudis & Prentzas, 2011a; Tweedale & Jain, 2014). This is accomplished by exploiting the advantages of the integrated methods to overcome their disadvantages. Complementarity in advantages and disadvantages of the combined methods is usually the basis to the success of such integrations. Example types of popular integrations, among others, involve neuro-symbolic approaches, integrating neural networks with symbolic methods (Garcez D'Avila & Lamb, 2011; Hatzilygeroudis & Prentzas, 2004a), neuro-fuzzy approaches, integrating neural networks with fuzzy methods (Evans & Kennedy, 2014; Lin et al., 2012; Zhang, Ma, & Yang, 2015), approaches combining neural networks and genetic algorithms (Huang, Li, & Xiao, 2015) and approaches combining case-based reasoning with rule-based reasoning (Prentzas & Hatzilygeroudis, 2007) or other intelligent methods (Chuang & Huang, 2011; Prentzas & Hatzilygeroudis, 2009).

A number of neuro-symbolic formalisms have been introduced during last decade (Garcez D'Avila, Broda, & Gabbay, 2002; Garcez D'Avila, Lamb, & Gabbay, 2009; Hatzilygeroudis & Prentzas, 2004a). Combinations of symbolic rules (of propositional type) and neural networks constitute a large proportion of neuro-symbolic approaches (Gallant, 1993; Hatzilygeroudis & Prentzas, 2000, 2001; Holldobler & Kalinke, 1994; Towell & Shavlik, 1994). The efforts integrating rules and neural networks may yield effective formalisms by exploiting the complementary advantages and disadvantages of the integrated components (Hatzilygeroudis & Prentzas, 2004a). Symbolic rule-based systems possess positive aspects such as naturalness and modularity of the rule base, interactive reasoning process and ability to explain reasoning results. Neural networks lack the naturalness and modularity of symbolic rules and it is also difficult (or impossible) to provide explanations. Explanations are crucial in certain domains such as medicine and finance. However, symbolic rules have disadvantages such as, difficulty in acquiring rules from the experts (known as the

* Corresponding author. Tel.: +30 25510 30086.
E-mail addresses: ihatz@ceid.upatras.gr (I. Hatzilygeroudis), dprentza@psed.duth.gr (J. Prentzas).

[1] The names of the authors appear in alphabetic order.

[2] Tel.: +30 2610 996937.

http://dx.doi.org/10.1016/j.eswa.2015.01.068
0957-4174/

'knowledge acquisition bottleneck'), inability to draw conclusions when there are missing values in the input data and possible problems in cases of unexpected input values or combinations of them. On the other hand, neural networks provide generalization, representation of complex and imprecise knowledge and knowledge acquisition from training examples.

Neurules are a type of integrated rules combining symbolic rules and neurocomputing (Hatzilygeroudis & Prentzas, 2000, 2001; Prentzas & Hatzilygeroudis, 2011). Neurules belong to neuro-symbolic representations resulting in a uniform, seamless combination of the two integrated components. Most of the existing such approaches give pre-eminence to connectionism. As a consequence, they do not offer important advantages of symbolic rules, like naturalness and modularity, and also do not provide interactive inference and explanation. Neurules follow a different direction by giving priority to the symbolic than the connectionist framework. Therefore, the knowledge base exhibits characteristics such as naturalness and modularity, to a large degree. Furthermore, neurule-based systems provide interactive inference and explanation.

Integration in neurules involves all knowledge representation aspects: syntax, semantics and reasoning. Hybridism in syntax and semantics has been presented in most of our past works on neurules and for the sake of completeness is briefly presented here too. Reasoning via neurules can be done via two different inference processes. The one gives pre-eminence to neurocomputing, namely *connectionism-oriented inference*, whereas the other to symbolic reasoning, namely *symbolism-oriented inference*. Both inference processes are integrated in their nature. Connectionism-oriented inference process has been presented in Hatzilygeroudis and Prentzas (2010) and compared to two alternative inference mechanisms used in connectionist expert systems (Gallant, 1993; Ghalwash, 1998), a type of neuro-symbolic systems. An initial version of the symbolism-oriented inference has been presented in Hatzilygeroudis and Prentzas (2000).

In this paper, we present an improved symbolism-oriented inference process. Improvement refers to the number of required computations to produce conclusions, the ability to work with any order of neurule conditions and the ability to work with two different sets of discrete values for representing 'true', 'false' and 'unknown' states. We present experimental results comparing the performance of the new symbolism-oriented process with the connectionism-oriented one.

However, the main contribution of this paper is the introduction of an explanation mechanism for neurule-based inference. The explanation mechanism provides three types of explanations: 'how', 'why' and 'why-not'. We also present experimental results comparing the 'how' explanation mechanism with the corresponding mechanism used in connectionist expert systems.

This paper is structured as follows. Section 2 briefly discusses related work. Section 3 presents neurules. Section 4 discusses the two alternative inference processes. Section 5 presents the explanation mechanism. Section 6 presents explanation examples. Section 7 presents experimental results involving inference and explanation. Section 8 concludes.

2. Related work

The objective of our work is to remain on the symbolic ground and incorporate techniques from the connectionist approach into propositional type symbolic rules to improve their representation capabilities and performance, without significantly reducing features, like naturalness and modularity, or sacrificing functionalities, like interactive inference and explanation. Many attempts based on the connectionist ground, which simulate or translate symbolic processes within a neural network, have been made in the recent years. However, a few of them are able to provide any of the above features in a satisfactory degree.

We can specify two main research trends. The first trend stems from Holldobler and Kalinke (1994), where a way to translate a propositional logic program (i.e. a set of facts and rules) into a neural network, by encoding its associated semantic operators in a connectionist way, is introduced. However, it is not accompanied by any reasoning or explanation process. KBANNs (Knowledge-Based Artificial Neural Networks) (Towell & Shavlik, 1994) use a core of propositional rules to construct an initial neural network and then use empirical knowledge to train the network. This leads to more efficient training and refinement of the initial symbolic knowledge. Such approaches are also not accompanied by integrated reasoning or explanation processes. Outputs are produced from the trained network via neurocomputing (i.e. numeric computation) methods without any explanation.

The other trend concerns connectionist expert systems. In connectionist expert systems, domain concepts are associated with neural network nodes and relationships among concepts are associated with links among nodes. A representative such approach is MACIE (Gallant, 1993). MACIE is accompanied by an inference and an explanation mechanism. The performance of the MACIE inference engine was improved by the introduction of the recency inference engine (Ghalwash, 1998). A comparison of the time performance of these two inference engines, taking into account the required number of computations is presented in Hatzilygeroudis and Prentzas (2010). A disadvantage of connectionist expert systems is that their knowledge bases lack the naturalness and modularity of symbolic rule bases. Meaningless nodes are inserted by the training mechanism, to be able to handle non-linearity of data, thus making representation and provided explanations unnatural (see Hatzilygeroudis & Prentzas, 2001).

3. Neurules: syntax and semantics

Neurules are a kind of integrated rules. The form of a neurule is depicted in Fig. 1a. Each condition C_i is assigned a number sf_i, called its *significance factor*. Moreover, each neurule itself is assigned a number sf_0, called its *bias factor*. Internally, each neurule is considered as an adaline unit neural (Fig. 1b). The *inputs* C_i ($i = 1, \ldots, n$) of the unit are the conditions of the neurule. The weights of the unit are the significance factors of the neurule and its bias is the bias factor of the neurule. The existence of bias (bias factor) helps in training the adaline unit (neurule), more specifically helps the training algorithm in converging more easily and for more cases. Training a neural unit (neurule) means finding values for its weights (significance factors) that satisfy given data, i.e. classify them correctly. Each input takes a value from the following set of discrete values: [1 (true), −1 (false), 0 (unknown)].

The *output D*, which represents the conclusion (decision) of the neurule, is calculated via the standard formulas:

$$D = f(a), \quad a = sf_0 + \sum_{i=1}^{n} sf_i C_i \tag{1}$$

$$f(a) = \begin{cases} 1 & \text{if} \quad a \geqslant 0 \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

where $\boldsymbol{a}$ is the *activation value* and $f(x)$ the *activation function*, which is a threshold function. Hence, the output can take one of two values ('−1', '1') representing failure and success of the rule respectively. The significance factor of a condition represents the significance (weight) of the condition in drawing the conclusion.