Interweaving Knowledge Representation and Adaptive Neural Networks

Ilianna Kollia, Nikolaos Simou,
Giorgos Stamou and Andreas Stafylopatis

Department of Electrical and Computer Engineering,
National Technical University of Athens,
Zographou 15780, Greece
nsimou@image.ntua.gr

Abstract. Both symbolic knowledge representation systems and machine learning techniques, including artificial neural networks, play a significant role in Artificial Intelligence. Interweaving these techniques, in order to achieve adaptation and robustness in classification problems is of great importance. In this paper we present a novel architecture that can provide effective connectionist adaptation of ontological knowledge. The proposed architecture can be used to improve performance of multimedia analysis and man machine interaction systems.

1 Introduction

Intelligent systems based on symbolic knowledge processing, on the one hand, and artificial neural networks, on the other, differ substantially. Nevertheless, they are both standard approaches to artificial intelligence and it is very desirable to combine the robustness provided by neural networks, especially when data are noisy, with the expressivity of symbolic knowledge representation. This has contributed decisively to the growing interest in developing neural-symbolic systems [4, 6, 5]. This integration can be realised by an incremental workflow for knowledge adaptation. Symbolic knowledge bases can be embedded into a connectionist representation, where the knowledge can be adapted and enhanced from raw data. This knowledge may in turn be extracted into symbolic form, where it can be further used.

In this paper we focus on developing connectionist adaptation of ontological knowledge. Section 2 presents the proposed architecture that mainly consists of the formal knowledge and the knowledge adaptation components, which are described in sections 3 and 4 respectively. Conclusions and ongoing research involving semantic multimedia analysis applications are reported in section 6.

2 The Proposed Architecture

Figure 1 summarizes the proposed system architecture, consisting of two main components: the *Formal Knowledge* and the *Knowledge Adaptation*. The *Formal*

Knowledge stores the knowledge base components that describe the problem under analysis. More specifically, the *Ontologies* module formally represents the general knowledge about the problem (TBox) generated during the Development Phase by knowledge engineers and experts.

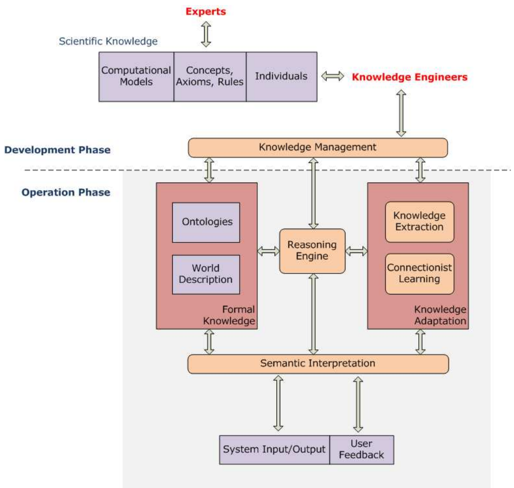

Fig. 1. The semantic adaptation architecture

Moreover, the *Formal Knowledge* contains the *World Description* that is actually a representation of all objects and individuals of the world, as well as their properties and relationships in terms of the Ontology (ABox). It is evident that most of the above data involve different types of uncertain information and, thus, they can be represented as formal (fuzzy) assertions connecting the objects and individuals of the world with the concepts and relationships of the Ontology. This operation is performed by the *Semantic Interpretation* module.

In real environments however, this is a rather optimistic claim. Unfortunately, there may be lot of reasons that cause inconsistencies in the *Formal Knowledge*. For example, it is impossible to model all specific environments and thus, in some cases, conflicting assertions can arise. In such cases, the *Knowledge Adaptation* component of the system tries to resolve the inconsistency through a recursive learning process. Adaptation improves the knowledge of system by changing the world description and to some degree the axioms of the terminology. The new information as represented in a connectionist model and, with the aid of learning algorithms, is adapted and then re-inserted in the knowledge base through the *Knowledge Extraction* and *Semantic Interpretation* modules for system adaptation.

3 The Formal Knowledge Component

The focus of the proposed system architecture in Figure 1 is on adaptation of the knowledge base, so as to deal with contextual information and raw data peculiarities obtained from multimodal inputs. In the paper we deal with interweaving of formal knowledge representation and neural-symbolic integration. In particular, we use recent results that extract parameter kernel functions for individuals within ontologies [3, 2, 1]. Exploitation of these kernels permits inducing classifiers for individuals in Semantic Web (OWL) ontologies. In this paper, extraction of kernel functions is the main outcome of the *Formal Knowledge* component - assisted by a reasoning engine - for feeding the connectionist-based *Knowledge Adaptation* task.

The family of kernel functions $k_p^F : Ind(A) \times Ind(A) \to [0,1]$, for a knowledge base $K = \langle T, A \rangle$ consisting of the TBox T (set of terminological axioms of concept descriptions) and the ABox A (assertions on the world state); $Ind(A)$ indicates the set of individuals appeared in A, and $F = \{F_1, F_2, \ldots, F_m\}$ is a set of concept descriptions. These functions are defined as the L_p mean of the, say m, simple concept kernel functions κ_i , $i = 1, \ldots, m$, where, for every two individuals a,b, and $p > 0$,

$$\kappa_i(a,b) = \begin{cases} 1 \ (F_i(a) \in A \wedge F_i(b) \in A) \vee (\neg F_i(a) \in A \wedge \neg F_i(b) \in A) \\ 0 \ (F_i(a) \in A \wedge \neg F_i(b) \in A) \vee (\neg F_i(a) \in A \wedge F_i(b) \in A) \\ \frac{1}{2} \text{ otherwise} \end{cases} \quad (1)$$

$$\forall a, b \in Ind(A) \qquad k_p^F(a,b) := \left[\sum_{i=1}^{m} \left| \frac{\kappa_i(a,b)}{m}^p \right| \right]^{1/p} \quad (2)$$

The rationale of these kernels is that similarity between individuals is determined by their similarity with respect to each concept F_i , i.e, if they both are instances of the concept or of its negation. It has been shown that k_p^F is a valid kernel, based on its composition through simpler matching kernels and on the closure property with respect to sum, multiplication by a constant and kernel multiplication.

It should be stressed that the reasoning engine, included in Figure 1, is of major importance for the whole procedure, because it assists the operation of all knowledge related components. First, during the knowledge development phase, it is responsible for enriching manual generation of concepts and relations. In the operation phase, it interacts with the semantic interpretation layer and the connectionist system for knowledge adaptation to local environments. Both crisp and fuzzy reasoners can form this engine, we use the FIRE engine [11] that is based on the fuzzy extension of the DL $\mathcal{SHIN}$ [7].

We use fuzzy reasoning because a fuzzy assertional component permits more detailed descriptions of a domain. In order to compute (1), (2) the *greatest lower bound* (GLB) reasoning service of FiRE defined in [12] is used, but the resulting greatest lower bound is treated crisply. In other words, if GLB for $F_i(a) > 0$, then $F_i(a) \in A$, while if GLB for $F_i(a) = 0$, then $\neg F_i(a) \in A$. We intend to incorporate the fuzzy element in the estimation of kernel functions using fuzzy operations like fuzzy aggregation and fuzzy weighted norms for the evaluation of the individuals.

4 The Adaptation Mechanism

In the proposed architecture of Figure 1, let us assume that the set of individuals (and corresponding features), that have been used to generate the formal knowledge, is provided, by the *Semantic Interpretation Layer*, to the *Knowledge Adaptation* component. Support Vector Machines constitute a well known method which can be based on kernel functions[2] to efficiently induce classifiers that work by mapping the instances into an embedding feature space, where they can be discriminated by means of a linear classifier. Kernel Perceptrons can be also applied to this linearly separable classification problem.

Let the system be in its -real life- operation phase. Due to local or user oriented characteristics, real life data can be quite different from those of the individuals used in the training phase; thus they may be not well represented by the existing formal knowledge. Whenever a new individual is presented to the system, it should be related, through the kernel function to each individual of the knowledge base w.r.t to a specific concept - category; the input data domain is, thus, transformed to another domain - taking into account the semantics that have been inserted to the kernel function. However, the kernel function in (1), (2) is not continuous w.r.t individuals. Consequently, the values of the kernel functions when relating a new individual to any existing one should be computed. To cope with this problem, it is assumed that the semantic relations, that are expressed through the above kernel functions, also hold for the syntactic relations of the individuals, as expressed by their corresponding low level features, estimated and presented at the system input. Under this assumption, a feature based matching criterion using a k-means algorithm, is used to relate the new individual to each one of the cluster centers w.r.t the low level feature vector. The SVM or Kernel Perceptron can be retrained - including the new individuals in the training data set, while getting the corresponding desired responses by the User or by the Semantic Interpretation Layer - adapting its knowledge to the specific context and use. An hierarchical, multilayer kernel perceptron, the input layer of which is identical to the trained kernel perceptron can also be used [9].

Knowledge extraction from trained neural networks has been a topic of extensive research [8]. One way is to transfer the - most characteristic - new individuals obtained in the local environment, together with the corresponding desired outputs - concepts, to the knowledge development module of the main system (Figure 1), so that with the assistance of the reasoning engine, the system formal knowledge, i.e., both the T-Box and the A-Box, be updated, w.r.t the specific context or user. A methodology that can be used to adapt a knowledge base $K = \langle T, A\rangle$ for a concept F_i, is the following. Check all related concepts, denoted as $R_{F_i}F_1 \ldots R_{F_i}F_i$ under the specific context, count the number $|R_{F_i}F_i|$ of occurrences of $R_{F_i}F_i \in A$, as well as the axioms defined for the concept F_i in the knowledge base (i.e. $Axiom(F_i) \in T$). Let $R_{F_i}F_i \in Axiom(F_i)$ when the concept $R_{F_i}F_i$ is used in $Axiom(F_i)$ and $R_{F_i}F_i \notin Axiom(F_i)$ when it is not used. The related concepts with the highest occurrence are selected for adaptation of the terminology, while the insignificant ones are removed.

5 Conclusion

In this paper we presented a novel architecture for connectionist adaptation of ontological knowledge. We are currently performing experimentation of the system performance for solving an image/video segmentation and classification problem [9, 10]. Future work, includes the incorporation of fuzzy set theory in the kernel evaluation. Additionally, we intend to further examine the adaptation of the knowledge base using the connectionist architecture, mainly focusing on the selection of the appropriate DL constructors and on inconsistency handling.

References

1. S. Bloehdorn and Y. Sure. Kernel methods for mining instance data in ontologies. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, 2007.
2. N. Fanizzi, C. d Amato, and F. Esposito. Randomised metric induction and evolutionary conceptual clustering for semantic knowledge bases. In *Proceedings of CIKM 07*, 2007.
3. N. Fanizzi, C. d Amato, and F. Esposito. Statistical learning for inductive query answering on owl ontologies. In *Proceedings of the 7th International Semantic Web Conference (ISWC)*, pages 195–212, 2008.
4. Artur S. Avila Garcez, K. Broda, and D. Gabbay. Symbolic knowledge extraction from trained neural networks: A sound approach. *Artificial Intelligence*, 125:155–207, 2001.
5. Barbara Hammer and Pascal Hitzler, editors. *Perspectives of Neural-Symbolic Integration*, volume 77 of *Studies in Computational Intelligence.* Springer, 2007.
6. P. Hitzler, S. Holldobler, and A. Seda. Logic programs and connectionist networks. *Journal of Applied Logic*, page 245272, 2004.
7. I. Horrocks, U. Sattler, and S. Tobies. Reasoning with Individuals for the Description Logic $\mathcal{SHIQ}$. In David MacAllester, editor, *CADE-2000*, number 1831 in LNAI, pages 482–496. Springer-Verlag, 2000.
8. E. Kolman and M. Margaliot. Are artificial neural networks white boxes? *IEEE Trans. on Neural Networks*, 16(4):844–852, 2005.
9. N. Simou, Th. Athanasiadis, S. Kollias, G. Stamou, and A. Stafylopatis. Semantic adaptation of neural network classifiers in image segmentation. 18th ICANN 2008, pp. 907-916, September 2008, Prague, Czech Republic., 2008.
10. G. Stamou and S. Kollias. *Multimedia Content and the Semantic Web: Methods, Standards and Tools.* John Wiley & Sons Ltd, 2005.
11. Giorgos Stoilos, Nikos Simou, Giorgos Stamou, and Stefanos Kollias. Uncertainty and the semantic web. *IEEE Intelligent Systems*, 21(5):84–87, 2006.
12. U. Straccia. Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.