LivingKnowledge: Kernel Methods for Relational Learning and Semantic Modeling

Alessandro Moschitti

Department of Computer Science and Information Engineering
University of Trento
Via Sommarive 14, 38100 POVO (TN) - Italy
moschitti@disi.unitn.it

Abstract. Latest results of statistical learning theory have provided techniques such us pattern analysis and relational learning, which help in modeling system behavior, e.g. the semantics expressed in text, images, speech for information search applications (e.g. as carried out by Google, Yahoo,..) or the semantics encoded in DNA sequences studied in Bioinformatics. These represent distinguished cases of successful use of statistical machine learning. The reason of this success relies on the ability of the latter to overcome the critical limitations of logic/rule-based approaches to semantic modeling: although, from a knowledge engineer perspective, hand-crafted rules are natural methods to encode system semantics, noise, ambiguity and errors, affecting dynamic systems, prevent them from being effective.

One drawback of statistical approaches relates to the complexity of modeling world objects in terms of simple parameters. In this paper, we describe kernel methods (KM), which are one of the most interesting results of statistical learning theory capable to abstract system design and make it simpler. We provide an example of effective use of KM for the design of a natural language application required in the European Project LivingKnowledge[1].

Keywords: Kernel Methods; Structural Kernels; Support Vector Machines; Natural Language Processing.

1 The Data Representation Problem

In recent years, a considerable part of Information Technology research has been addressed to the use of machine learning for the automatic design of critical system components, e.g. automatic recognition/classification of critical data patterns. One of the most important advantages with respect to manually coded system modules is the ability of learning algorithms to automatically extract the salient properties of the target system from training examples. This approach can produce semantic models of system behavior based on a large number of attributes, where the values of the latter can be automatically learned. The

[1] http://livingknowledge-project.eu/

T. Margaria and B. Steffen (Eds.): ISoLA 2010, Part II, LNCS 6416, pp. 15–19, 2010.

statistically acquired parameters make the overall model robust and flexible to unexpected system condition changes. Unfortunately, while attribute values and their relations with other attributes can be learned, the design of attributes suitable for representing the target system properties (e.g. a system state) has to be manually carry out. This requires expertise, intuition and deep knowledge about the expected system behavior. For example, how can system module structures be converted into attribute-value representations?

Previous work on applied machine learning research (see the proceedings of ICML, ECML, ACL, SIGIR and ICDM conferences)[2] has shown that, although the choice of the learning algorithm affects system accuracy, feature (attribute) engineering more critically impacts the latter. Feature design is also considered the most difficult step as it requires expertise, intuition and deep knowledge about the target problem. Kernel methods is a research line towards alleviating the problem above.

2 Data Representation via Kernel Methods

Kernel Methods (KM) are powerful techniques developed within the framework of statistical learning theory [18]. They can replace attributes in learning algorithms simplifying data encoding. More specifically, kernel functions can define structural and/or semantic similarities between data objects at abstract level by replacing the similarity defined in terms of attribute matching.

The theory of KM in pattern analysis is widely discussed in [17] whereas an easier introduction can be grasped from the slides available at `http://disi.unitn.eu/~moschitt/teaching.html`. The main idea of KM is expressed by the following two points:

(a) directly using a similarity function between instances in learning algorithms, thus avoiding explicit feature design; and
(b) such function implicitly corresponds to attribute matching (more precisely scalar product) defined in huge feature spaces (possibly infinite), e.g. similarity between structures can be defined as substructure matching in the substructure space.

The first point states that instead of describing our data, e.g. a data stream, in terms of features (which aim at capturing the most interesting properties or behavior), it is enough to define a function capable to measure the similarity between any pair of different data objects, e.g. pairs of streams.

The second bullet emphasizes the great power of KM as the representations that can be modeled with them are extremely rich and are not computationally limited by the size of feature spaces.

[2] ICML and ECML are the International and European Conferences of Machine Learning, respectively. ACL is the Conference for Association of Computational Linguistics; IR is the most important conference for Information Retrieval and ICDM is the International Conference on Data Mining.

KM effectiveness has been shown in many ICT fields, e.g. in Bioinformatics [16], Speech Processing [1], Image Processing [5], Computational Linguistics [9], Data Mining [4] and so on. In particular, KM have been used to encode syntactic and/or semantic structures in the form of trees and sequences in learning algorithms, e.g. [2,3,7,19,10,20,14,11,13,12].

Given the wide and successful use of KM, they have been applied in the LivingKnowledge project to model several aspects of automatic knowledge acquisition and management, which are basic building blocks required by the project.

3 Using Kernels for Semantic Inference in LivingKnowledge

Judgements, assessments and opinions play a crucial role in many areas of our societies, including politics and economics. They reflect knowledge diversity in perspective and goals. The vision inspiring LivingKnowledge (LK) is to consider diversity as an asset and to make it traceable, understandable and exploitable, with the goal to improve navigation and search in very large multimodal datasets (e.g., the Web itself).

To design systems that are capable of automatically analyzing opinions in *free text*, it is necessary to consider syntactic/semantic structures of natural language expressed in the target documents. Although several sources of information and knowledge are considered in LK, we here illustrate an example focused on text. Given a natural language sentence like for example:

They called him a liar.

the opinion analysis requires to determine: (i) the opinion holder, i.e. *They*, (ii) the direct subjective expressions (DSEs), which are explicit mentions of opinion, i.e. *called*, and (iii) the expressive subjective elements (ESEs), which signal the attitude of the speakers by means of the words they choose, i.e. *liar*.

In order to automatically extract such data, the overall sentence semantics must be considered. In turn, this can be derived by representing the syntactic and shallow semantic dependencies between sentence words. Figure 1 shows a graph representation, which can be automatically generated by off-the-shelf syntactic/semantic parsers, e.g. [6], [8]. The oriented arcs, above the sentences, represent syntactic dependencies whereas the arcs below are shallow semantic

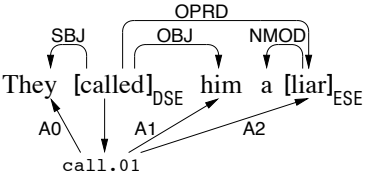

Fig. 1. Syntactic and shallow semantic structure

(or semantic role) annotations. For example, the predicate *called*, which is an instance of the PropBank [15] frame `call.01`, has three semantic arguments: the Agent (A0), the Theme (A1), and a second predicate (A2), which are realized on the surface-syntactic level as a subject, a direct object, and an object predicative complement, respectively.

Once the richer representation above is available, we need to encode it in the learning algorithm, which will be applied to learn the functionality (subjective expression segmentation and recognition) of the target system module, i.e. the opinion recognizer. Since such graphs are essentially trees, we exploit the ability of tree kernels [10] to represent them in terms of subtrees, i.e. each subtree will be generated as an individual feature of the huge space of substructures.

Regarding practical design, kernels for structures such us trees, sequences and sets of them are available in the SVM-Light-TK toolkit (`http://disi.unitn.it/moschitti/Tree-Kernel.htm`). This encodes several structural kernels in Support Vector Machines, which is one of the most accurate learning algorithm [18].

Our initial test on the LivingKnowledge tasks suggests that kernel methods and machine learning are an effective approach to model the complex semantic phenomena of natural language.

4 Conclusion

Recently, Information Technology research has been addressed to the use of machine learning for automatic design of critical system components, e.g. automatic recognition of critical data patterns. The major advantage is that the system behavior can be automatically learned from training examples. The most critical disadvantage is the complexity to model effective system parameters (attributes), especially when they are structured.

Kernel Methods (KM) are powerful techniques that can replace attribute-value representations by defining structural and/or semantic similarities between data objects (e.g. system states) at abstract level. For example, to encode the information in a data stream, we just define a function measuring the similarity between pairs of different streams: such function can be modeled in extremely rich and large feature spaces.

A considerable amount of previous work shows the benefit of employing KM and our initial study in LivingKnowledge, whose application domain requires to model complex textual and image information, further demonstrate their benefits.

Acknowledgements

This research has been supported by the EC project, EternalS: Trustworthy Eternal Systems via Evolving Software, Data and Knowledge, project number FP7 247758.

References

1. Campbell, W.M.: Generalized linear discriminant sequence kernels for speaker recognition. In: International Conference on Acoustics, Speech, and Signal Processing (2002)
2. Collins, M., Duffy, N.: New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron. In: Proceedings of ACL 2002 (2002)
3. Culotta, A., Sorensen, J.: Dependency Tree Kernels for Relation Extraction. In: Proceedings of ACL 2004 (2004)
4. Gärtner, T.: A survey of kernels for structured data. SIGKDD Explor. Newsl. 5(1), 49–58 (2003)
5. Grauman, K., Darrell, T.: The pyramid match kernel: Discriminative classification with sets of image features. In: International Conference on Computer Vision (2005)
6. Johansson, R., Nugues, P.: Dependency-based syntactic–semantic analysis with PropBank and NomBank. In: Proceedings of the Shared Task Session of CoNLL 2008 (2008)
7. Kudo, T., Matsumoto, Y.: Fast methods for kernel-based text analysis. In: Proceedings of ACL 2003 (2003)
8. Moschitti, A., Coppola, B., Giuglea, A., Basili, R.: Hierarchical semantic role labeling. In: CoNLL 2005 shared task (2005)
9. Moschitti, A.: A study on convolution kernels for shallow semantic parsing. In: Proceedings of ACL 2004 (2004)
10. Moschitti, A.: Efficient convolution kernels for dependency and constituent syntactic trees. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) ECML 2006. LNCS (LNAI), vol. 4212, pp. 318–329. Springer, Heidelberg (2006)
11. Moschitti, A.: Making tree kernels practical for natural language learning. In: Proceedings of EACL 2006 (2006)
12. Moschitti, A.: Kernel methods, syntax and semantics for relational text categorization. In: Proceeding of CIKM 2008 (2008)
13. Moschitti, A., Quarteroni, S., Basili, R., Manandhar, S.: Exploiting syntactic and shallow semantic kernels for question/answer classification. In: Proceedings of ACL 2007 (2007)
14. Moschitti, A., Zanzotto, F.M.: Fast and effective kernels for relational learning from texts. In: ICML 2007 (2007)
15. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: An annotated corpus of semantic roles. Comput. Linguist. 31(1), 71–106 (2005)
16. Schšlkopf, B., Guyon, I., Weston, J.: Statistical learning and kernel methods in bioinformatics. In: Artificial Intelligence and Heuristic Methods in Bioinformatics (2003)
17. Taylor, J.S., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge (2004)
18. Vapnik, V.N.: Statistical Learning Theory. Wiley-Interscience, Hoboken (1998)
19. Zhang, D., Lee, W.S.: Question classification using support vector machines. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp. 26–32. ACM Press, New York (2003)
20. Zhang, M., Zhang, J., Su, J.: Exploring Syntactic Features for Relation Extraction using a Convolution tree kernel. In: Proceedings of NAACL (2006), `http://www.aclweb.org/anthology/N/N06/N06-1037`