Adaptive Ontologies Through Social Evolution (Doctoral Consortium)

Davide Nunes
University of Lisbon, Portugal
dnunes@fc.ul.pt

ABSTRACT

The Semantic Web movement promotes the embedding of semantic content into Web resources. Unfortunately, the Web entities do not find enough motivation nor reward to do their part in the construction of a semantically richer Web environment. To enable a true semantic Web, each human-readable document should provide a formal knowledge representation for its content. As hand-crafted knowledge acquisition seems infeasible when applied to an environment as vast and dynamic as the Web, efforts have been focused on automatic approaches to extract semantic content from textual documents. Advances in Natural Language Processing, Information Extraction and Ontology Learning provided tools that allow for the extraction and analysis of structured semantic knowledge given existing textual corpus. While these tools seem promising to enable a scenario where Semantic Web can be achieved, they need to be adapted to the scale and complexity of the Web. In this article, I propose that a possible solution lies in harvesting the power of emergent multi-agent societies to create an infrastructure capable of bootstrapping the adoption of Semantic Web technologies. I focus on how to create and adapt distributed on-line evolutionary algorithms to continuously design and improve social agents capable of semantic knowledge retrieval.

I argue that currently existing automatic and semi-automatic structured knowledge acquisition techniques can be adapted to serve not only as building blocks for emerging and evolving knowledge extracting processes but also as one of the driving forces behind its adaptation. I then claim that learning extracting procedures can be done not only using annotated resources but also existing techniques creating dynamic fitness models that can be continuously updated to improve the system.

Categories and Subject Descriptors

I.2.11 [**ARTIFICIAL INTELLIGENCE**]: Distributed Artificial Intelligence; H.0 [**Information Systems**]: GENERAL

Keywords

Multi-agent Systems, Distributed Artificial Intelligence, Evolutionary Computation, On-line learning, Ontology Learning, Ontology Emergence, Semantic Web

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

1. INTRODUCTION

The World Wide Web is a continuously growing source of information and services shared by people and applications. Tapping into this dynamic and open environment to extract usable knowledge in an automatic fashion is a difficult problem for various reasons, the main one being the fact that the available information is displayed mainly in a human-readable format. Within the origins of the Web, Berners-Lee [1] envisions a possible response to this problem, the idea of building a Semantic Web in which the information appearing on a document would also contain additional annotations created using well defined standards such as ontologies [3], which would be understandable by a machine.

Since hand-crafting ontologies not enough to promote a large-scale adoption of these standards and jump-start the emergence of a Semantic Web infrastructure, efforts have been made for the creation of automatic or semi-automatic methods that allow the extraction of ontologies from existing textual corpus. *Ontology Learning* [8] emerges as a new research area, devoted to the development of automatic methods for the extraction of a domain model from relevant (domain-specific) data. This area largely builds on methods previously developed in knowledge acquisition, natural language processing and machine learning, with the specific purpose of automatically deriving an domain ontologies.

Wong et al. [8] analysed the state-of-the-art in ontology learning systems and argue that most of these seem to be designed as a proof of concept capable of addressing the requirements for the construction of small toy ontologies. They also highlight that more research effort is required to make existing techniques operational on cross-domain resources on a Web scale.

In my thesis, I advocate that descriptive ontologies should be created not just from isolated documents, but rather taking into account relevant semantic content generated from neighbour documents. To do this, I consider the usage of a multi-agent system and evolutionary computation to create and adapt agents to extract knowledge from distributed resources.

2. MAS AND SOCIAL EVOLUTION

Using multi-agent systems to extract structured knowledge should not be seen as a "divide and conquer" approach but instead, a "distribute and emerge" one, which can be provided by the autonomy of the agents in the system. The problem itself is very difficult to distribute for a later integration of solutions. Approaches to ontology learning like the one described in [5] are prone to ontology integration bottlenecks. Moreover scalability and modularity are also compromised. If the extraction mechanisms or the resources change, the process needs to be restarted to obtain the new results. Instead, I focus on the creation of continuously adapting agents that change their knowledge extraction behaviour over time. Each agent evolves separately in different resources but should be able to capitalise on social knowledge from neighbours evolving in similar contexts.

In fact, having autonomous agents with various sets of goals and distinct behaviours (which might be cooperative or competitive), may allow for more consensual and robust solutions.

2.1 On-line Social Evolution

My thesis focus on three key mechanisms to address different issues in the continuous creation and adaptation of structured knowledge for the Web. The first problem is how to develop on-line evolutionary algorithms that can lean on social knowledge, already acquired by the agent society, to analyse distinct resources. In this context, a test-bed for on-line evolution was developed. Agents evolve using Cartesian Genetic Programming (CGP) [4] to adapt to unknown test environments in a continuous fashion. The benefits of cultural mechanisms, imitation and competition to continuously search in agent design space are currently being tested using the developed platform.

2.2 Evolutionary Knowledge Extraction

In a second stage the adaptation of the explored mechanisms to knowledge extraction in actual documents will be explored. Instead of testing on-line social learning mechanisms on abstract environments, these are tested with the intent of evolving a target knowledge extraction mechanism such as *term extraction* or *hierarchical ontology structure extraction*. The evaluation is made using existing datasets and state-of-the-art extraction techniques as a baseline to be compared to the evolved mechanisms. Despite the rather unsuccessful attempt to disseminate the usage of ontologies, advances have been made in domain specific tasks and applications. Life sciences and healthcare have been areas acting as major promoters for Semantic Web technology such as ontologies [2] and can provide data to evaluate the results from our MAS evolutionary approaches.

2.3 Continuous Adaptation

Finally, the continuous adaptation capabilities of the developed approaches are tested by creating fitness models based on both datasets and extraction techniques. Complex environments usually require a large set of training data in order not only to characterise the domain, but also to make coherent assessment of the quality of the results. A study of dynamic fitness models based on approaches such as the one developed in [7] will be developed. The idea is to allow for a dynamic adjustment of the evaluation mechanism as more resources, techniques and ontology user feedback are available. Evolving fitness models for agent-based knowledge extraction mechanisms will allow for reduced computational costs for the evaluation process(the model can provide a lightweight fitness approximation for an otherwise costly operation). Moreover, this approach is appropriate for domains where explicit computable fitness assessments are not possible due to domain complexity or subjectivity.

In this case, if our system provides a set of structured knowledge that is intended to be used in real applications, user feedback is not only a possibility but a desirable resource to be explored. Similar applications include, for example, human interactive evolution of music or art. Much like in our case, in these domains, user feedback can be used to define a computable fitness landscape that can be searched by the system [6].

3. CONCLUSION

In this article, I describe my thesis point-of-view on emergent knowledge extraction mechanisms to create a continuously infrastructure that would allow to bootstrap the Semantic Web. Ultimately, I focus on the adaptation of evolutionary approaches to search in agent design space to create knowledge extraction mechanisms. These mechanisms capitalise on social dynamics and learning processes such as imitation, competition and culture to evolve an agent society in an on-line manner. Also, due to the domain complexity, an explicit evaluation function does not exist and might never apart from heuristic approaches. As such, much like what happens with the agent society, the objective is to provide continuously evolving fitness landscapes to be explored by our social evolutionary techniques.

4. REFERENCES

[1] T. Berners-Lee and M. Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1st edition, 1999.

[2] P. Ciccarese, E. Wu, G. Wong, M. Ocana, J. Kinoshita, A. Ruttenberg, and T. Clark. The SWAN biomedical discourse ontology. *J. of Biomedical Informatics*, 41:739–751, Oct. 2008.

[3] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.

[4] J. Miller and P. Thomson. Cartesian genetic programming. In *Genetic Programming*, volume 1802 of *Lecture Notes in Computer Science*, pages 121–132. Springer Berlin Heidelberg, 2000.

[5] D. Sánchez. *Domain ontology learning from the web: an unsupervised, automatic and domain independent approach*. PhD thesis, 2008.

[6] M. D. Schmidt and H. Lipson. Actively probing and modeling users in interactive coevolution. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 385–386. ACM, 2006.

[7] M. D. Schmidt and H. Lipson. Coevolution of fitness predictors. *Evolutionary Computation, IEEE Transactions on*, 12(6):736–749, 2008.

[8] W. Wong, W. Liu, and M. Bennamoun. Ontology learning from text: A look back and into the future. *ACM Comput. Surv.*, 44(4):20:1?20:36, Sept. 2012.