A Fast Approximation of the Weisfeiler-Lehman Graph Kernel for RDF Data*

Gerben Klaas Dirk de Vries [a]

[a] *System and Network Engineering group, Informatics Institute, University of Amsterdam, The Netherlands, g.k.d.devries@uva.nl*

Abstract

We introduce an approximation of the Weisfeiler-Lehman graph kernel algorithm aimed at improving the computation time of the kernel when applied to Resource Description Framework (RDF) data. RDF is the representation/storarge format of the semantic web and it essentially represents a graph. One direction for learning from the semantic web is using graph kernel methods on RDF. This is a very generic and flexible approach to learning from the semantic web, since it requires no knowledge of the semantics of the dataset and can be applied to nearly all linked data.

Graph kernel computation is in general slow, since it is often based on computing some form of expensive (iso)morphism between graphs. We present an approximation of the Weisfeiler-Lehman (WL) graph kernel [2] to speed up the computation of this kernel on RDF data. Typically, applying graph kernels to RDF is done by extracting subgraphs from a large underlying RDF graph and computing the kernel on this set of subgraphs. Our approximation exploits the fact that the subgraph instances are extracted from the same RDF graph. We adapt the WL algorithm to compute the kernel directly on the underlying graph, while maintaining a subgraph perspective for each instance.

We compare the performance of this kernel to the graph kernels designed for RDF described in [1]. For this comparison we use three property prediction tasks on RDF data from two datasets. In each task we try to predict a property for a certain class of resources. For instance, the first task is predicting the affilition of the people in a research institute, for which the data is modeled as RDF. Furthermore, we compare the computation time of the different kernels.

In all three tasks, our kernel shows performance that is better than the regular Weisfeiler-Lehman kernel applied to RDF. Also it is increasingly more efficient as the number of instances grows by exploiting the fact that the RDF instance subgraphs share vertices and edges in the underlying large RDF graph. Furthermore, the presented kernel is faster and/or shows better classification performance than the intersection subtree and intersection graph kernels for RDF, introduced in [1]. The performance difference between the presented approximation of the WL Subtree kernel and the regular version requires further investigation.

Acknowledgments

This publication was supported by the Dutch national program COMMIT.

References

[1] Uta Lösch, Stephan Bloehdorn, and Achim Rettinger. Graph kernels for RDF data. In Elena Simperl, Philipp Cimiano, Axel Polleres, Óscar Corcho, and Valentina Presutti, editors, *ESWC*, volume 7295 of *Lecture Notes in Computer Science*, pages 134–148. Springer, 2012.

[2] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt. Weisfeiler-Lehman graph kernels. *J. Mach. Learn. Res.*, 12:2539–2561, November 2011.

*The full paper has been published in ECML/PKDD (1), volume 8188 of Lecture Notes in Computer Science, pages 606–621, 2013