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

Referee report

Emilio Cecchini
emilio.cecchini@stud.unifi.it

Lorenzo Palloni
lorenzo.palloni@stud.unifi.it

June 2, 2019

1 Summary

The goal of this paper is to introduce a faster version of the Weisfeiler-Lehman graph kernel algorithm when applied to Resource Description Framework (RDF) data.

The *Resource Description Framework* (RDF) is the foundation for knowledge representation on the semantic web. A resource is described by a set of *triples* which are of the form *subject-predicate-object*. The entire collection of triples form a graph where the subjects and the objects are the nodes and the predicates are the edges.

The *Weisfeiler-Lehman test* is an algorithm that is used to compute graph isomorphism. The test proceeds in iterations where the key idea is to augment the node labels by the sorted set of node labels of neighbouring nodes, and compress these augmented labels into new, short labels. These steps are then repeated until the node label sets of the two graphs differ, or the number of iterations reaches the prefixed maximum.

The *Weisfeiler-Lehman kernel* is the state-of-the-art for graph kernels. It computes the number of subtrees shared between two graphs by using the Weisfeiler-Lehman test of graph isomorphism.

This paper introduces an approximation of the Weisfeiler-Lehman kernel, which first extracts a set of subgraphs from the entire RDF graph and then the kernels are computed. For each instance a subgraph up to a certain depth is extracted from the RDF dataset and this subgraph is added to a total graph that the extraction algorithm is building. Thus, vertices and edges are only added if they have not been added to the graph already. For each node and edge, together with their labels, their extraction depth is stored. The relabeling process is the same of the standard Weisfeiler-Lehman test with the extension of the labels on the edges. Finally the kernel is computed by counting the number of common labels at each depth.

References

[1] Vries Gerben Klaas Dirk, A Fast Approximation of the Weisfeiler-Lehman Graph
    Kernel for RDF Data, 2013