Kernels for RDF data using Spark

Dennis Kubitza *

Maximilian Radomsky †

February 11, 2018

Project References on Github

Lab Report [1]

Abstract

Machine learning paradigms strongly depend on the specific structure of the observed and unobserved Data. For the big goal of promoting Machine Learning on Structured Data like Resouce Description Frameworks, with its schema-free structure, one class of Algorithms is naturally well suited: Kernel Based Algorithms. We follow the examinations of Lösch et al. (2012) and implement their proposed Graph-Kernels for the usage in Apache Spark, especially for further usage in the Semantic Analytics Stack (SANSA). Our implementation combines different approaches from Graph Combinatorics, Data-Mining and Big Data Analysis to ensure scalability in storage and computational performance.

*Email: **denn_kubi@freenet.de**, Postal adress: Rudolf-Breitscheid-Str.140595 Düsseldorf, Germany

†test

[1]as Part of the Examination of Modul 4223, Master of Computer Science, University of Bonn

Contents

1 Introduction

While each every Machine Learning Task is defined by its Input Set, it's Set of valid models and the expected behaviour of the learning Agent, some algorithms exist that solve problems under such general assumtpions that almost any Data-dependet Problem can be reduced to fit their requirements. Kernel-based Machine learning methods don't require any specific structure for the Data, solomly that a scalar valued function exists, suitable for summarizing a Observation or Subobservation as a single value. We call such a function a kernel Functions. Before stating a mathematic exact definition, applyable to even the most general settings we are taking a look on some examples where Kernel-Based Machine learning is applied to RDF.

Kernel Examples

- XXX
- XXX
- XXX

A definition of Kernels, that is suitably general, but still mathematically precise can be found in XXX.In short notion we can say:

Definition 1 *A kernel is XXX.*

In the case of RDF-Date Lösch et al. (2012) listed four different kinds of Kernels that are . Risk creates opportunities that can be exploited, while ambiguity cripples the will to invest.

Psychic Costs and Human Capital Investment I also shed new light o
Based on business cycle dynamics Lösch et al. (2012), no trade results (**?**), and the equity premium puzzle (**?**).[2] ternatives. Section **??** outlines

[2]See Lösch et al. (2012) for a recent review and numerous additional references. **?** provide a textbook treatment in the context of macroeconomics.

References

Lösch, U., Bloehdorn, S., and Rettinger, A. (2012). Graph kernels for rdf data. In Simperl, E., Cimiano, P., Polleres, A., Corcho, O., and Presutti, V., editors, *The Semantic Web: Research and Applications*, pages 134–148. Springer Berlin Heidelberg, Berlin, Heidelberg.