# 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.1̇40595 Düsseldorf, Germany

† test

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

# Contents

1 Introduction 3

2 Conceptual Frameworks 5

2.1 Economic Model . . . . . 5

2.1.1 Uncertainty as Risk . . . . . 6

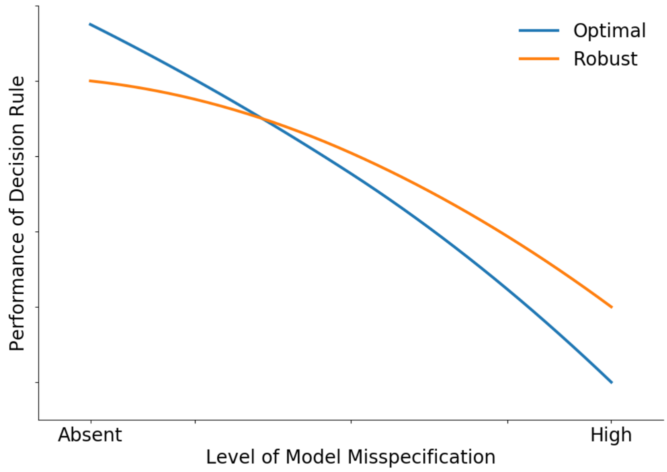
2.1.2 Uncertainty as Risk and Ambiguity . . . . . 7

2.2 Computational Model . . . . . 8

2.2.1 Uncertainty as Risk . . . . . 9

2.2.2 Uncertainty as Risk and Ambiguity . . . . . 9

**Figure 1:** Robust Decisions




# 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:

A kernel is XXX.

. 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.

# 2 Conceptual Frameworks

I now present my dynamic life cycle model for robust human capital investment under risk and ambiguity. I start with the presentation of the economic model, discuss the basic economic environment and alternative models of decision making. Then then turn to the computational model and outline its solution approach.

## 2.1 Economic Model

At each point in time $t \in T$ an individual observes the the economic environment $s \in S$. Observing $s$, the individual chooses an action $a \in \mathcal{A}_s$ from the set of admissible actions in state $s$. The action choice has two results, the individual receives an immediate reward $r_t(s, a)$ and the system evolves to a new state. A decision rule $d_t$ specifies the action to be chosen at a particular time $t$ for any possible state. A policy $\pi = (d_1, \ldots, d_T)$ provides the decision maker with a prescription for choosing an action in any possible future state, it is a sequence of decision rules and its implementation generates a sequence of rewards. The evolution of states over time is uncertain. Let $X_t$ and $Y_t$ denote the random variables for the state and action at time $t$ respectively. The history $h_t$ is the sequence of previous states and decisions up to time $t$. Let $R_t \equiv r_t(X_t, d_t(X_t))$ denote the random reward received in period $t$, $R \equiv (R_1, \ldots, R_T)$ denote the random sequence of rewards, and $\mathcal{R}$ the set of all possible reward sequences. A policy $\pi$ induces a probability distribution $P^{\pi}_{\mathcal{R}}(R)$ on $\mathcal{R}$.

Figure 2 depicts the timing of events in the model for two generic time periods. At the beginning of time $t$ an individual fully learns about the immediate rewards of all alternatives, chooses one of them, and receives the reward. The state space is then updated and the process repeats in $t+1$. Individuals are confronted with uncertainty as future rewards are at least partly unknown.

Individuals use models about their economic environment to form beliefs about the future (**??**). If individuals rely on a single model then they are able to assign unique probabilities to all possible futures. They make their human capital investment decision under risk. If instead they examine multiple models then they have to consider a whole set of probabilites for the future make their decisions facing risk and ambiguity.

Economic decision theory offers guidance on desirable decision principles in both cases. Numerous alternative approaches exist and adopting any particular approach requires to assess its behavioral implications as well as the trade-offs to other approaches (**??**).

**Figure 2:** Timing

| Learn | Choose | Receive |
|---|---|---|
| $\{r_t(s,a)\}_{a\in A}$ | $d_t(s)$ | $r_t(s, d_t(s))$ |

$t$

| Learn | Choose | Receive |
|---|---|---|
| $\{r_{t+1}(s',a)\}_{a\in A}$ | $d_{t+1}(s')$ | $r_{t+1}(s', d_{t+1}(s'))$ |

$t+1$

### 2.1.1 Uncertainty as Risk

There exists a clear consensus on how reasonable individuals make decisions in light of risk. Individuals act as to maximize their expected utility. They simply evaluate the total rewards for all possible futures and weigh them by their respective probabilities. They choose the alternative that yields the highest expected utility. A distinguishing feature between expected utility theories is the source of the probabilities used for the calculation. These are either objective or subjective probabilities.

The theory of objective expected utility originally proposed by **?** and subsequently axiomatized by **??** in the context of individual's choosing between different lotteries.

Adopted to a model of human capital investment, individuals consider only a single model of their economic environment and make their decisions using on the objective probabilities derived from a model. There is no role for model misspecification as the model they use is correct. Individuals have rational expectations (**??**), i.e. their behavior is consistent with the model. More formally, the objective of the individual is to maximize the expected discounted lifetime rewards.

$$v^{\pi}_T(s) = \max_{\pi\in\Pi} \mathrm{E}^{\pi}_s \left[ \sum_{t=1}^{T} \delta^{t-1} r_t(X_t, d_t(X_t)) \right] \tag{1}$$

The exponential discount factor $0 < \delta < 1$ captures the individual's preference for immediate over future rewards. Individual maximize equation (1) by choosing the optimal policy $\pi^* = [d_1, \dots, d_T]$ given their initial state $s$.

The individual's preferences induced by this decision rule are time-consistent, i.e. the decision maker's preferences over contingent plans agree with his preferences in the planned-for contingency. Each prior is updated by Bayes' rule and the relative importance of future rewards remains constant due to exponential discounting, this follows from the linearity of the decision rule in the probabilities and the law of iterated expectations.[3]

This is the standard modeling approach in the existing literature. Individuals consider a unique model for their decisions, they have no concerns about model misspecification, and have rational expectations.

### 2.1.2 Uncertainty as Risk and Ambiguity

There exists no clear consensus on how to make reasonable decisions in light of risk within a model and ambiguity about the model. So, I will briefly outline the most common models for decision making under risk and ambiguity and motivate my eventual modeling choice.

A natural extension of the expected utility theory that allows to capture uncertainty across models is offered by **?**. He formulated subjective expected utility theory.[4] In this case, the individual simply assigns each of the models under consideration a subjective probability. In this case, decision making under ambiguity reduces to decision making under risk. Individuals do not distinguish between the probabilities derived from the models and their subjective beliefs about the relative merit of each model.[5]

However, it in my setting it remains unclear what informs the subjective probabilities and numerous experimental and empirical evidence indicates that individuals have a preference for objective over subjective probabilities (**?**). Thus I build on non-expected utility models instead.

The two most popular non-expected utility models are the minimum regret and maxmin utility model.[6] In both cases, individuals strive to make decisions that work well over a whole range of models. For minimum regret, the individual computes the best decision and then evaluates the regret for all alternative choices. In the end, the individual chooses the action to minimize the maximum regret. This criterion was proposed in **?** and axiomatized by **?** and **?**. Maxmin

[3]**?** was the first to note that individuals might choose a plan of action but reconsider later. He also showed the key role of exponential discounting. See **?** and **?** as examples for models with potentially time-inconsistent preferences in a similar model setup.

[4]See **?** and **?** (translation in **?**) for introducing the notion of subjective probability to quantify degrees of beliefs. Alternative models of subjective expected utility theory were proposed by **?** and **?**.

[5]See **?**, **?**, and **?** for alternative approaches that illicit expectations directly.

[6]There are several other ambiguity models, such as Choquet expected utility (**?**), smooth ambiguity (**?**), variational (**?**), and multiplier (**?**) preferences. See **?** and **?** for insightful discussions about the differences between the approaches.

utility advises to choose the alternative that maximizes the expected utility under a worst-case scenario. It was originally proposed by **?** and axiomatized by **?**.[7]

As noted in **?** and **?**, non-expected utility models may lead to time-inconsistent behavior. At the same time, there exists no consensus on how to update ambiguous beliefs (**?**). However, **?** and **?** develop and axiomatize a recursive approach for the maxmin expected utility rule that ensures dynamic consistency. They formulate the requirement that the set of priors is rectangular and is updated by Bayes' Rule applied prior by prior. Rectangularity is a form of an independence assumption and interpreted in an adversarial setting. The choice of a particular distribution in a state does not limit the choices of the adversary in the future. The adversary is able to choose a different probability distribution every time a state is encountered. In effect, the decision makers gains nothing from by having future actions depend explicitly on past realizations of uncertainty. I will provide a more formal definition when discussing the solution technique.

I adopt the maxmin expected utility model. The objective of the individual is to maximize the expected lifetime rewards under a worst-case scenario among all models considered. More formally, the individual's objective is to maximize the expected discounted lifetime rewards under the worst-case.

$$v_T^\pi(s) = \max_{\pi \in \Pi} \left\{ \min_{\mathbf{P} \in F^\pi} \mathrm{E}_s^{\mathbf{P}} \left[ \sum_{t=1}^{T} \delta^{t-1} r_t(X_t, d_t(X_t)) \right] \right\} \tag{2}$$

$F^\pi$ clarifies that the expectation is taken with respect to the set of measures $\mathcal{F}^\pi$ that is associated with policy $\pi$. As the true model is included in the set of admissible models, so the assumption of rational expectations is only weakened. If $\mathcal{P}_t(s, a)$ is a singleton, we are back to the standard model of decision making under risk.

## 2.2 Computational Model

The human capital model under risk and ambiguity is set up as an ambiguous Markov decision problem (AMDP). As it nests the standard Markov decision problem in the special case of decision making under risk only, I briefly review the basic element for the standard case before turning to the new issues arising due to the introduction of ambiguity. A deterministic and Markovian decision rule is always optimal for an MDP and AMDP, thus I restrict the formal presentation to this particular type of decision rule to ease notation.

[7] See **?** for a historical account of the diffusion of the maxmin principle across different disciplines.

### 2.2.1 Uncertainty as Risk

Individuals determine the optimal policy by comparing $v_T^\pi(s)$ for alternative $\pi \in \Pi$ as each induces an alternative reward sequence over the random rewards $\mathcal{R}_T$. The challenge is to compute the optimal policy $\pi^*$ efficiently. We can draw on **?** and solve this multistage problem by analyzing a sequence of simpler inductively defined single-state problems. This avoids the tasks of determining the joint probability distribution of histories under each policy $\pi$. Let $u_t^\pi(h_t)$ denote the expected total discounted reward going forward:

$$u_t^\pi(s_t) = r_t(s_t, d(s_t))) + \delta \mathrm{E}_{s_t}^\pi \left[ u_{t+1}^\pi(X_{t+1}) \right]$$

Then the optimality equations simply read:

$$u_t^*(s_t) = \max_{a \in A} \left\{ r_t(s_t, a) + \delta \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \right\}$$

These can be solved to inductively determine the solution to determine the optimal problem. See Algorithm (1) for a formal description.

**Algorithm 1** Backward Induction Algorithm for MDP

**for** $t = T, \ldots, 1$ **do**
  **if** t == T **then**
    $u_T^*(s_T) = r_T(s, a) \qquad \forall \quad s_T \in S$
  **else**
    Compute $u_t^*(s_t)$ for each $s_t \in S$ by
      $u_t^*(s_t) = \max_{a \in A} \left\{ r_t(s_t, a) + \delta \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \right\}$
    and set
      $d_t^*(s_t) = \arg\max_{a \in A} \left\{ r_t(s_t, a) + \delta \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \right\}.$
  **end if**
**end for**

### 2.2.2 Uncertainty as Risk and Ambiguity

Recent work in operations research (**??**) establishes that, given the assumption of rectangularity, key results for standard Markov decision processes such as the Bellman recursion carry over.[8] This literature tackles the problem of making sequential decisions in light of ambiguous future transition probabilities of the system. It is motivated by the sensitivity of the optimal policy to perturbations in the transition probability resulting in serious degradation of

[8]See **?** for a textbook introduction to standard dynamic programming and **?** for an overview on Markov decision processes in structural estimation.

performance (????). In the model, the uncertainty about the future transition probabilities corresponds to the uncertainty in future choice probabilities due the ambiguity of the distribution of the random reward components.

The set of all conditional measures consistent with a deterministic Markov decision rule $d_t(s)$ is given by:

$$\mathcal{F}^{d_t} = \{\mathbf{p} : S_t \mapsto \mathcal{M}(S_{t+1}) : \forall s \in S_t, \mathbf{p_s} \in P_t(s, d_t(s))\}$$

For every state $s \in S$, the next state can be determined by any $p \in P_t(s, d_t(s))$.

A policy $\pi$ induces a collection of measures on the history space $\mathcal{H}_T$. There is a key assumption associated with the set of measures $\mathcal{F}^\pi$ associated with policy $\pi$ is rectangular.

**Definition 1** ***Rectangularity*** *The set $\mathcal{F}^\pi$ of measures associated with a policy $\pi$ is given by*

$$\begin{aligned}\mathcal{F}^\pi &= \left\{\mathbf{P} : \forall h_T \in \mathcal{H}_T, \mathbf{P}(h_T) = \prod_{t=1}^{T} \mathbf{p}_s, \mathbf{p}_s \in \mathcal{F}^{d_t}, t \in T\right\} \\ &= \mathcal{F}^{d_1} \times \mathcal{F}^{d_2} \times \ldots \times \mathcal{F}^{d_T}\end{aligned}$$

*where the notation simply denotes that each $p \in \mathcal{F}^\pi$ is a product of $p_t \in \mathcal{F}^{d_t}$ (?).*

As noted earlier, the rectangularity assumption is best interpreted in an adversarial setting. The decision maker chooses $\pi$, an adversary observes $\pi$, and chooses a measure $\mathbf{P} \in \mathcal{F}^\pi$ that minimizes the reward. In this context, rectangularity is a form of an independence assumption. The choice of a particular distribution $\bar{p} \in \mathcal{P}_t(s_{,t}\, a_t)$ in a state action pair $(s_t, a_t)$ at time $t$ does not limit the choices of the adversary in the future. In the finite-horizon setting this can be justified by invoking time inhomogeneity.

Alternative policies are evaluated under a worst-case scenario:

$$u_t^\pi(s_t) = \min_{\mathbf{P} \in F_t^\pi} \mathrm{E}_{s_t}^{\mathbf{P}} \left[\sum_{\tau=t}^{T} \delta^{\tau-t} r_\tau(X_\tau, d_\tau(X_\tau))\right]$$

The modified optimality equations are:

$$u_t^*(s_t) = \max_{a \in A} \left\{ r_t(s_t, a) + \delta \min_{p \in \mathcal{P}_t(s_t, a)} \mathrm{E}_{s_t}^{p} \left[u_{t+1}^*(X_{t+1})\right] \right\} \tag{3}$$

Algorithm (2) outlines the procedure:

**Algorithm 2** Backward Induction Algorithm for AMDP

**for** $t = T, \dots, 1$ **do**
  **if** t == T **then**
    $u_T^*(s_T) = r_T(s, a) \qquad \forall \quad s_T \in S$
  **else**
    Compute $u_t^*(s_t)$ for each $s_t \in S$ by

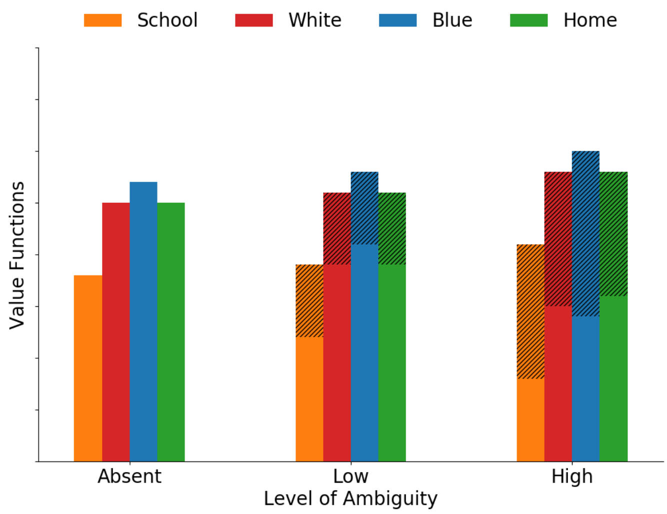
$$u_t^*(s_t) = \max_{a \in A} \left\{ r_t(s_t, a) + \delta \min_{p \in \mathcal{P}_t(s_t, a)} \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \right\}$$

    and set

$$d_t^*(s_t) = \arg\max_{a \in A} \left\{ r_t(s_t, a) + \delta \min_{p \in \mathcal{P}_t(s_t, a)} \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \right\}$$

  **end if**
**end for**

**Figure 3:** Admissible Value Functions




I will now illustrate the individual's decision problem under different specifications of uncertainty for a generic state $s_t$. Let $u_t^\pi(s_t, a)$ denote the alternative-specific value function:

$$u_t^*(s_t, a) = r_t(s_t, a) + \delta \mathrm{E}_{s_t}^p \left[ u_{t+1}^*(X_{t+1}) \right] \qquad \forall a \in \mathcal{A}.$$

Figure 3 plots the alternative-specific value function at $s_t$. As in my empirical application, the individual is considering whether to work in the labor market in a white or blue collar occupation, enroll in school, or remain at home. In the absence of ambiguity, the decision problem is straightforward. The individual clearly prefers to enroll in school. However, now ambiguity is introduced and thus with each alternative a whole host of admissible values is indicated by the dashed area. Given the maxmin utility function, the individual will still decide to enroll in school for a low level of ambiguity. However, further increases now lead the individual to stay at home.

# 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.