Research Seminar in Data Science
Paper Reviews

Christoph Schmidl
s4226887
c.schmidl@student.ru.nl

February 21, 2018

Reviewed papers:

- **A Few Useful Things to Know about Machine Learning** by Pedro Domingos (Presented by Bauke Brenninkmeijer)

- **Anomaly Detection in Crowded Scenes** by Vijay Mahadevan et al. (Presented by Joan Ressing)

1 A Few Useful Things to Know about Machine Learning

1.1 Summary

In his paper "A Few Useful Things to Know about Machine Learning", Pedro Domingos describes different guidelines and rules of thumb that he learned through out his career in machine learning. The paper does not focus on a single research question or experiment but tries to give a broad overview to the obstacles which occur often in the domain of machine learning, such as overfitting, underfitting, curse of dimensionality, bias and variance and gives different solutions such as regularization techniques or model ensembles.

1.2 Evidence

The author does not really make claims but goes through the solutions to certain problems which are widely known by machine learning practitioners. The paper is more structured like a literature review as it points to different sources while describing different techniques such as boosting or bagging. Out of 26 references to other papers, 5 of them were written by himself which makes the credibility kind of questionable in my opinion.

1.3 Strengths

Normally you expect a scientific paper to include new knowledge in the form of an novel answer to a certain research question but this paper is more like a summary. This paper is worth reading if you are new to the field of machine learning or want to read a quick recap. That is certainly a strength of this paper and it does its job really well in my opinion.

1.4 Weaknesses

The conclusion was unexpected as it does not include a recap but is pointing to a new book written by Pedro Domingos himself. It made the impression of a badly placed advertisement. The other resources placed in the conclusion seemed more trustworthy to me though.

1.5 Evaluation

I do not know the exact requirements for the acceptance of a scientific paper to a conference but I guess that this paper would not get accepted because it does not add novel knowledge to a certain domain. I also do not see any conference affiliations in the title of the paper, therefore I guess that this paper is not peer reviewed and has not been published by a conference.

1.6 Comments on the quality of the writing

The writing style was informal which made it an easy read but it probably would need some adjustments if it was sent to reviewers of a larger conference.

1.7 Queries for discussion

In section "5. Overfitting has many faces" the author mentions the problem of "multiple testing" which I did not understand entirely. It would be nice if you could incorporate this part into the presentation and explain it a bit more.

2 Anomaly Detection in Crowded Scenes

2.1 Summary

Vijay Mahadevan et al. present a framework for anomaly detection in crowded scenes based on a method using mixtures of dynamic textures. The difficulty of this problem lies in the joint modeling of appearance and dynamics of the scene and to detect temporal and spatial abnormalities. The proposed model is compared to three other approaches like the social force model, the mixture of optical flow and the optical flow monitoring method and later on is shown that the MDT-based anomaly detection model outperforms the others.

2.2 Evidence

The largest part of the paper describes different known approaches to subparts of the problem in isolation while referencing other sources which seem trustworthy. They use training and test sets of different scenes which contain normal and abnormal behaviour while the overall class distribution is a little unbalanced. The authors later on provide ROC courves for five different approaches.

2.3 Strengths

The main strenghts of this paper is probably its novelty and its superior performance with regards to other known solutions. The authors describe the difficulty of combining temporal and spatial abnormality detection and go into more details on already known solutions which concentrate on each of the problem separately. Temporal abnormality detections mostly uses the background subtraction method relying on a Gaussian mixture. Spatial abnormality detection relies on saliency detection. They describe in detail on how they integrated these separated techniques into their solution using MDTs (mixtures of dynamic textures).

2.4 Weaknesses

The Experiments and Results part could have been larger and it is somehow inconvenient that the implementations of the models which get compared with their proposed model were done by themselves.

2.5 Evaluation

I would assume that this paper would get accepted to a conference or journal based on its sophisticated writing style, the level of technical detail and novelty. The only thing which may hinder acceptance is the rather short part on experiments and results and the fact that the proposed model is compared to others which are based on their own implementations. This may not yield full objectivity.

2.6 Comments on the quality of the writing

I think that the quality of writing and overall structure is pretty well and the technical detail is also on a high level. I found only a few typos which can be ignored.

2.7 Queries for discussion

- The paper is pretty dense when it comes to technical details. It would be helpful if you could describe the overall approach and maybe explain the most important formulas.

- How is temporal and spatial abnormality detection combined in MDT?

- Is it a problem that the used dataset is using a somehow unbalanced class distribution? Although the different is not really that big.

- The authors compare five different models in the end. Nearly all of them were implemented by themselves although the underlying ideas originate from other papers. Does this fact disrupt the credibility of the results?