Gottfried Wilhelm Leibniz Universität Hannover
Institut für Verteilte Systeme
Distributed Computing & Security Group

Master thesis
Informatics (M.Sc.)

Anomaly detection in streaming data using autoencoders

Student:	B.Sc. Bin Li
First Supervisor:	Prof. Dr. Eirini Ntoutsi
Second Supervisor:	Prof. Dr. Wolfgang Nejdl
Date:	March 12, 2018

Declaration of Authorship

I hereby certify that this thesis has been composed by me and is based on my own work, unless stated otherwise. No other person's work has been used without due acknowledgement in this thesis. All references and verbatim extracts have been quoted, and all sources of information have been specifically acknowledged.

Hanover, March 12, 2018

B.Sc. Bin Li

Contents

List of Figures

List of Tables

Codeverzeichniss

Chapter 1

Introduction

Der Reiseführer hat eine sehr schöne Theorie zur Komplexität des Universums:

Chapter 2

Related works

2.1 Anomaly detection

There are already planty of algorithms used for anomaly detection.

2.2 Autoencoder for anomaly detection

2.3 Streaming data anomaly detection

Chapter 3

Model

3.1 LSTMs-based autoencoder

The basic anomaly detection model is designed in an autoencoder based architecture. However, the target is to detect abnormal data points from a data stream, or in other words, detect anomalous from a time series in an online fashion, which is actually different from traditional outlier detection from batch data, while data points within a time series always has potencial temporal dependencies between each other, and these information would play important roles. From the perspective of neural networks, Recurrent Neural Networks (RNNs) are designed for the analysis of such dynamic data and sequences of data. But in many practical cases, the RNNs are facing to the vanishing gradient problem. We always want the RNNs remember more knowlegde from the past, therefore the Long-Short Term Memory networks (LSTM networks) becomes a good replacement of the RNNs. In our autoencoder, we construct both the encoder and the decoder with LSTMs, in order to keep the past knowledge and use for future prediction.
Consider the time series X=$\{x^1, x^2\}$,