Cathodic Protection Prediction

Jelle van Wezel: *Cathodic Protection Prediction*, Time-series predictions with GMLVQ, © May 2018

Ohana means family.
Family means nobody gets left behind, or forgotten.

— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.

1939 – 2005

Foreword here …

ABSTRACT

Short summary of the contents in English…a great guide by Kent Beck how to write good abstracts can be found here:

`https://plg.uwaterloo.ca/~migod/research/beckOOPSLA.html`

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— Donald E. Knuth [1]

ACKNOWLEDGMENTS

Put your acknowledgments here.

Many thanks to everybody who already sent me a postcard!

Regarding the typography and other help, many thanks go to Marco Kuhlmann, Philipp Lehman, Lothar Schlesier, Jim Young, Lorenzo Pantieri and Enrico Gregorio[1], Jörg Sommer, Joachim Köstler, Daniel Gottschlag, Denis Aydin, Paride Legovini, Steffen Prochnow, Nicolas Repp, Hinrich Harms, Roland Winkler, Jörg Weber, Henri Menke, Claus Lahiri, Clemens Niederberger, Stefano Bragaglia, Jörn Hees, Scott Lowe, Dave Howcroft, José M. Alcaide, and the whole LaTeX-community for support, ideas and some great software.

Regarding LyX: The LyX port was intially done by *Nicholas Mariette* in March 2009 and continued by *Ivo Pletikosić* in 2011. Thank you very much for your work and for the contributions to the original style.

1 Members of GuIT (Gruppo Italiano Utilizzatori di TeX e LaTeX)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1 INTRODUCTION

PROBLEM DESCRIPTION

Steel gas-pipes are placed in the ground where they form corrosion. The amount of formed corrosion is thought to be depending on multiple environment factors. For example, the composition of the soil, the ground water level, and the depth the pipes are placed. In order to prevent the formation of corrosion on the pipes a method called cathodic protection is applied. Cathodic protection uses an electrical system where anodes (metal rods) are placed in the ground near the cathode (the steel pipes). When current is introduced to the system corrosion forms on the anodes and not on the cathode.

This method is applied on multiple pipe-lines by an energy company called 'Cogas' which is located in Almelo. They have collected the environment factors by hand and know how much current is needed per pipe section. Because of the corrosion buildup over time on the anodes and the pipes, at some point in time the amount of current needed exceeds a threshold. At this point it is no longer possible or economically viable to produce the amount of current needed to prevent the corrosion. When this happens more anodes need to be placed or the steel pipes have to be replaced. Cogas wants to be able to predict when this will happen based on their collected data.

The project will be done for ValueA. ValueA provides ICT solutions for utility companies and has Cogas as a client. One of the software solutions ValueA has provided to Cogas is an interactive dashboard where pipe sections with some of their properties are shown. This project will focus on predicting when a pipe section needs maintenance by replacing the pipes or placing more anodes. These predictions could be incorporated in this dashboard. It is unknown which environment factors (features) leads to the deterioration of the steel pipes. Methods like Supervised Variational Relevance Learning (SUVREL) and Global Metric Learning (GML) could give insight on which features are more important than others in this deterioration process. With the knowledge about the features predictions can be done on the dataset. For this project Learning Vector Quantization (LVQ) will be used to do the predictions. Its performance will be compared to the intuitive and widely used classifier K nearest neighbor (K-nn).

The cathodic protection predictions project is not the only project ValueA is involved with. Another classification problem involves Fraud detection in the energy grid. This problem forms the base of the graduation project of Sebastiaan van Loon. Since both the project of Sebastiaan van Loon and this project will be conducted at ValueA in the same period some collaboration will occur. This collaboration will consist of the development of an abstract module which includes an API that allows the existing software of ValueA to connect to a classifier implementing the abstract module.

This abstract module will be developed to ensure that the ValueA software can connect to a classification module independent of the methods the classification uses. This means that the development of the abstract module does not include the development of the actual classification algorithms and distance measures, which will be developed separately and are not part of the collaboration between Jelle van Wezel and Sebastiaan van Loon.

GOALS

- Research existing techniques to map LVQ and KNN on a regression problem
- Research existing methods for feature reduction/relevance
- Develop a classification module for ValueA which can predict the lifespan of steel gas pipes with LVQ, KNN and a or multiple preprocessing methods like GML and SUVREL.
- Offer a comparison of the results between the used methods
- Describe the research, used methods, developed software, and the results.

WORKING ENVIRONMENT

The student will be partly embedded both in the Intelligent Systems research group within the JBI, and with ValueA at the location of Cogas, a customer of ValueA located in Almelo. It is envisioned that the student will work, on average, two days a week at Cogas and the remaining days at the university and at home. The student will have a meeting at least every fortnight with the first supervisor and will have at least one stand up a week with the third supervisor and/or other employees of ValueA and will attend the monthly meetings of the Intelligent Systems research group.

The first and second supervisor will provide necessary expertise regarding the classification algorithms and dissimilarity measures used in the project. The third supervisor will provide necessary expertise about both the ValueA software solution and the energy utility market, provide data sets for the training and testing of the classifiers, and evaluate and offer feedback on the software developed.

The student will work on his own laptop and PC, and in case of failure work on a university managed system. Deployment of software will be done on a cloud environment provided by ValueA. Code revisions will be stored in a GIT repository managed by ValueA. Data which ownership belongs to ValueA will remain in the ValueA cloud environment and will not be copied to the student's laptops unless explicit permission is granted from ValueA.

DELIVERABLES

As requested by ValueA the student will develop the software in Python.

- End of December. Classification module description. A document describing the classification module, its main features and its integration with the existing ValueA software.
- Start of January. Literature survey. A document summarizing the main findings of a literature review and the relevant references regarding Cathodic protection predictions, LVQ, KNN, Metric learning.
- End of January. Abstract classification module. An abstract module facilitating the connection of a classifier with the existing software and data of ValueA. (In collaboration with Sebastiaan van Loon.)
- Mid of February. KNN classification module. A KNN classification module for predictions of gas pipe lifespan.
- End of March. LVQ classification module. A LVQ module for predictions of gas pipe lifespan with relevance learning as initialization.
- Mid of April. Test results.
- Start of May. Thesis draft.
- Mid of May. Presentation draft.
- End of May. Presentation to the board of Cogas.
- End of May. Final presentation, code, and thesis.

2 BACKGROUND

CATHODIC PROTECTION

3 DATASET

configuration

The data consisted of 24 CP areas. Each of these areas have one anode point where the current is pushed into the ground. The current is then propagated through the ground and led back to the anode through the steel gas pipes completing the electric system.

The pipes spread from the anode and branch off from each-other, creating a tree like structure. The pipes are divided into segments, Usually where a pipe branches into two pipelines, two new segments form. A pipe can only be part of one segment but a segment can consist of multiple pipes. In between the segments measuring points are placed to measure the current and difference in voltages in the segment. The number of segments an measuring points differ per area and is shown in table

Table 3.1

right	center	left
right	center	left

Table 3.2

right	center	left
right	center	left

missing data

Usually the measurements are conducted once per year at every measuring point. The measurements started as early as 1987 up-until 2016. This means their should be 39 measurements per measuring point. However this is not the case as is shown in figure

Figure [REF] shows a fast drop in measurement points after the 20 measurements mark. At 26 measurements the drop stagnates and the maximum number of measurements is reached at 37 measurements. This means that none of the measuring points have the maximum number of 39 measurements.

interpolation

The method we proposed to use to predict the voltages of the pipelines with LVQ [REF] depends on the data being consistently spaced in time.

Before the interpolation method was applied the data was checked for outliers by taking the mean and two times the standart deveation.

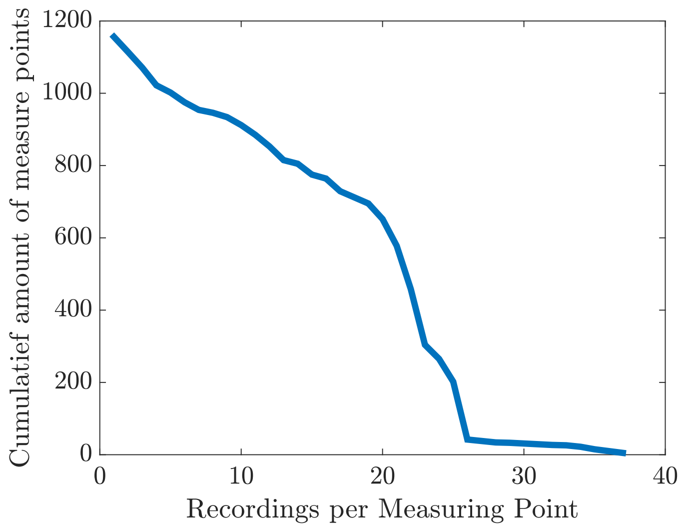

Figure 3.1: The number of measurements per measuring point against the cumulative number of measuring points having that number of measurements.

Then removing the data points that did not fall in the range of $x > 2 \times \sigma - \mu$ and $x < 2 \times \sigma + \mu$.

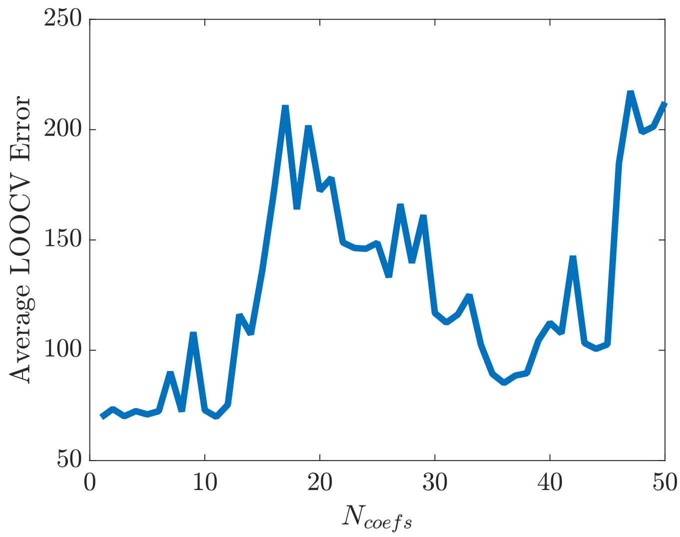

Figure 3.2: The average leave one out cross validation over all areas plotted against the number of N coefficients used for the polynomial interpolation.

4 METHODS

5 RESULTS

6 DISCUSSION

BIBLIOGRAPHY

[1] Donald E. Knuth. "Computer Programming as an Art." In: *Communications of the ACM* 17.12 (1974), pp. 667–673.

DECLARATION

Put your declaration here.

Groningen, May 2018

Jelle van Wezel

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Thank you very much for your feedback and contribution.