Master Thesis

Gaining Customer Insights using Machine Learning on Graphs

University of Basel

Author:
Michael von Siebenthal

Supervisor:
Prof. Dr. Dietmar Maringer

April 10, 2021

Abstract

Abstract goes here

Dedication

To mum and dad

Declaration

I declare that..

Acknowledgements

I want to thank...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Overview

Everyone has surely made the at times inconvenient experience of having been contacted by marketers, worst of all telemarketers. Not only are these marketers more often than not unpleasantly persistent, they in addition often try to sell something which we are not interested in. While we may never get rid off marketers, there are methods which could ensure that we at least receive offers which we are likely to be interested in. This would make a positive outcome for both the customer and the marketer much more likely.

Thanks to the advent of big data, technology companies such as Google and Facebook have long since built their entire business model on providing customer data to marketers. The customer data is often collected through social media, internet searches and by any other ways and means by which customer data can be collected. Of course, many marketers also possess their own databases from their customers which they may use for marketing purposes.

In order for marketers or from now on more generally companies to accurately predict, whether a particular client would be interested in a particular product, different methods for predicting client interests can be applied. For instance, a naive but often time predictive method would be to offer a product to people which have purchased the same or a similar product in the past. While one could apply this method, it has a major shortfall in that it does not allow us to identify new customers for the product. In order to receive better results, more sophisticated methods from machine learning could be used. In machine learning terminology, predicting whether a client would be interested in a product is referred to as a classification task. There are many machine learning methods which can perform

such a task to varying success such as:

1. Linear- or Quadratic Discriminant Analysis
2. Support Vector Machines
3. Ensemble Methods
4. Neural Networks

Of course this list is not exhaustive and there are a myriad of methods available. Further, there is no one size fits all solution, as every machine learning method has its advantages and disadvantages. The methods listed are all powerful methods in their own right and are capable of performing a large range of classification tasks. Traditionally, these methods were designed to perform tasks using "classical" datasets such as cross sectional data. The term "classical" is a reference where an observation in a dataset stands for itself. For instance survey data collected from the respondent "Peter" are not directly linked/connected to the answers of the respondent "Paul". Graphs or networks are very different in this respect. As an illustration, observations in social networks are very much connected (e.g. Peter and Paul could be friends). In the past 10 years there has been an explosion of new research in this area, where data scientist have developed methods to perform machine learning tasks on network data. This is an exciting new frontier in data science, which from a methodological perspective will be the focus of this master thesis for performing client classification tasks.

In the following sections a detailed description of the classification task for this master thesis is given, followed by a literature review of the extant literature.

1.2 Topical Setting

Retail banking is an area in which client advisers typically serve several thousand clients. In addition, advisers typically work in teams which makes personalized advice virtually impossible. For this reason, retail clients are often undeserved and are often dissatisfied with their bank. This is an area where classifying clients according to their needs or interests could tremendously improve the service quality provided.

Another interesting application and the focus for this thesis is to classify bank clients according to their investment preference (e.g. which type of products should be advertised to which client?). This would be especially useful in the retail banking segment where advisers typically cannot know their clients personally due to the

large number of clients being serviced. Investor classification is the intended main focus of this Master Thesis.

1.3 Challenges

This topic faces many different hurdles due to the low availability and mostly absence of available bank client data. To the extent possible, appropriate data sets will be retrieved (thus far 1 dataset found). The main difficulty however is to find a dataset which both includes attribute/feature data and the network structure of the customer data. For this reason, mostly methods to create synthetic data will be used to create the dataset for the subsequent simulation/testing.

Bibliography