Mitacs-Accelerate Final Report

Print Form

Privacy Statement

Please note that Mitacs may share this information with government funding agencies or other groups to assess or promote the Mitacs-Accelerate program. Do not include confidential, proprietary, or sensitive corporate information within this Final Report.

Instructions

The fields in this form will expand as required. Although the form will allow most users to save the form during entries, we recommend producing your content in another document, then pasting your complete answers into this form. Electronic signatures are preferred, but this form may be printed to collect manual signatures if required.

Once all signatures have been obtained, please submit the final copy via email to accelerate@mitacs.ca.

If you are submitting signatures manually, please first submit your unsigned form, then send a separate email with a scanned copy containing manual signatures. Should you experience any difficulty, please email accelerate@mitacs.ca.

Checklist

☒ All sections are completed
☒ All participants have signed and agreed to report content

Further Program Information

Are you interested in making further use of Mitacs-Accelerate and wish to speak with a Mitacs staff member?

Yes

Project Information

Project Title: Recurrent Neural Networks for Credit Card Fraud Detection

Internship Start Date: 6-Jan-20

Internship End Date: 29-May-20

Internship Reference Number (eg. IT1234): IT16464

Participants

Intern

First Name: Ruben

Last Name: Chevez-Guardado

Academic Institution: Memorial University

Department: Computer Science

Degree Program: Masters

Academic Institution Address: 230 Elizabeth Ave, St. John's, NL A1C 5S7

Primary Email: rubencg@mun.ca

Alternate Email: rubencg195@hotmail.com

Phone: 709-764-7777

Academic Supervisor

First Name: Lourdes

Last Name: Peña-Castillo

Department: Computer Science/Biology

Academic Institution: Memorial University

Academic Institution Address: 230 Elizabeth Ave, St. John's, NL A1C 5S7

Primary Email: lourdes@mun.ca

Phone: (709) 864-6769

Organization Partner

First Name: Terrence

Last Name: Tricco

Position: Research

Department: Data Science

Organization: Verafin

Address: 18 Hebron Way, St. John's, NL A1A 0L9

Phone:

Primary Email: terrence.tricco@verafin.com

Background Information

Fraudulent activities are very hard to detect due to the low amounts of incidents per sample of historical data. Still, they can cost millions of dollars in monetary losses and legal costs to financial institutions. Machine learning models are slowly being adopted in the fraud detection industry. Some significant roadblocks are the typical unbalanced data that is traditionally not publicly available and the black-box model behaviour that makes it difficult to explain the outcome of an algorithm's prediction. Recurrent Neural Networks (RNNs) are a promising machine learning approach for fraud detection due to their ability to learn patterns from time series. Financial transactions are arranged in a time-ordered manner, making it a perfect environment to test RNNs.

Describe the research goals and whether or not they were achieved

The primary goal of the study is to perform empirical research to evaluate the performance of RNN for fraudulent behaviour detection. The goal was achieved by identifying publicly available credit card fraud datasets, conducting a complete literature review on previous strategies for training Recurrent Neural Networks, and providing a practical implementation of RNN variants for fraud detection.

Summarize the research outcome(s)

The empirical study and results collected during the research have been compiled in a research paper and a programming source code repository, which has been delivered to the partner organization. The empirical study included the literature review, multiple RNN architecture designs that were trained and tested on two publicly available fraud datasets.

Describe the methods and techniques applied

The study first required us to narrow the research scope. The focus was on credit card fraud detection since there was more publicly available data about it. After identifying the available data sets, the data analysis of such was done to try to find possible patterns. At the same time, a complete literature review was done covering previous related work on recurrent neural networks for fraud detection. All remaining inquiries about the selected dataset were cleared during an interview with the author and creator of the data set. Multiple proposed RNN architecture designs were trained in different configurations until the best performing models were achieved. All results are then benchmarked with pre-existent machine learning algorithms.

Describe how the time you spent at the partner company's site impacted the direction of your research

The partner organization's feedback, the equipment, and the workplace were instrumental to the development of this research. The equipment made our testing environment efficient and faster. The workplace provided a peaceful place to focus, and the feedback helped shape the study and the goals' completion.

What were the benefits to the Organization Partner as a result of this internship?

The organization partner can benefit from one additional tool in their arsenal against fraudsters. The algorithm is described in detail in our complete empirical analysis of Recurrent Neural Networks (RNNs) and its implementation for fraud detection. Every monetary transaction received and managed by the partner organization's clients is stored and ordered by time. The exact kind of data that RNNs are designed to handle.

Discuss future research plans

The experience learned during the internship will be of great benefit for my upcoming Master's thesis. The same algorithms are applied but for a different domain of study. In this case, the study of the bacterial genome using machine learning techniques.

List any publications, patents and/or licenses which resulted from this internship

No publications, patents, or licenses were produced.

Include any additional comments

Executive Summary (300 words)

Please describe using language understandable by a non-expert:
- The nature of the problem the project addressed
- Goals achieved
- Any positive outcomes from the project, including:

papers published, patents obtained, any benefits to the company (improvements, cost-savings, new / improved technologies or processes developed) and whether the intern was hired afterwards

In the digital age or fourth industrial revolution, fraud has become a significant problem for financial institutions. Fraudsters continue to find new ways to exploit financial system vulnerabilities, leaving previous detection methods obsolete. Each security breach can cost millions of dollars in legal and stolen goods for financial institutions. The partner organization's primary service is to provide a financial fraud identification service. Any improvement can save their clients significant amounts of costs. The focus of this study is to provide additional tools in their arsenal against fraudsters. These tools are called Recurrent Neural Networks (RNNs) and are well-suited for handling data that is ordered and can have patterns over time. For example, a bank's user utilizes its credit card daily, spend more on weekends, and pays its bills on time. These actions are normal customer-specific behavior, but a sudden purchase of a product in a different location, a new merchant, or costly item outside the user's routine can be a result of fraud by an external agent. RNNs are perfect in this environment, and with additional tweaks done in its internal architecture, it is possible to generate an algorithm capable of dealing with highly unbalanced datasets. During the study, we created multiple RNN variation, tested them in two datasets and various configurations. After many trial and error, a well-performing model was obtained capable of detecting fraud. The complete empirical analysis containing all the lessons learned was compiled into a research paper and delivered to the partner organization.

By signing below I agree that the information contained in this Final Report is an accurate reflection of this Mitacs-Accelerate Internship. I understand that Mitacs may share the Executive Summary portion of this document with government funding agencies, partners, and prospective partners of the Mitacs-Accelerate Program.

Intern Signature

Supervisor Signature

Organization Partner Signature