A High Performing and Scalable Model for   
Computing and Visualizing Urban Transit Accessibility

**Yuxuan Cui**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[cuiyuuxuan@gmail.com](mailto:cuiyuuxuan@gmail.com)

**Luka Vukovic**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[luka.vuko@outlook.com](mailto:luka.vuko@outlook.com)

**Rain Shen**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[rain.ya1213@gmail.com](mailto:rain.ya1213@gmail.com)

**Graham Paul Kerford**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[graham.kerford@gmail.com](mailto:graham.kerford@gmail.com)

June 25th, 2021

**ABSTRACT**

Problem Statement:  
Background and more problems:  
Brief hypothesis/solution statement:  
Brief methodology:  
Results:  
Future work:

**1 INTRODUCTION**

Lorem ipsilum if you know what I mean.

**2 BACKGROUND**

In this section we briefly discuss .. (overview).

**2.1 Literature Review**

Lorem ipsilum if you know what I mean.

**2.2 Specific Terminology Stuff**

Lorem ipsilum if you know what I mean.

**2.3 Research Questions**

Lorem ipsilum if you know what I mean.

**3 METHODOLOGY**

In this section, we detail how the research problem was approached computationally starting from the data to the … .

**3.1 Data**

The data.

**3.2 Creation of Multimodal Networks**

**3.3 Transit Accessibility Scoring**

**3.4 Amenity Accessibility Weights**

**3.5 Software Specification**

**3.6 Experimental Setup**

To summarize the dataset, the produced html files of the score sets, Kepler file and isochrone were summarized in a dashboard using r shiny. R shiny was preferred over other dashboards due to its sleekness and ease of use. The dashboard was inspired by an existing dashboard create by Washington Post's feature ‘Washington: A world apart’ (<http://www.washingtonpost.com/sf/local/2013/11/09/washington-a-world-apart/>) and would comprise four tabs which allows users to switch between accessibility scores.

Originally the dashboard was created using Leaflet choropleth maps that were generated each time a new selection is selected. This however led to the rendering of the newly selected maps inefficient. To improve the efficiency of the dashboard the Leaflet choropleth maps were initial exported as html files before being displayed on the dashboard. The switch to the use of html files significantly improved the efficiency of the dashboard by over 10x. This however led to significant complications in the creation of the dashboard.

Due to the scores and properties of the census blocks being built into the html files, to switch between score sets the dashboard had to select a new html file by editing the file path. This also led to the removal of graphs summarize the data as no active interaction between the html files and the dashboard.

For each new parameter, a new html file had to be created. In total 32 html files were created for the score sets. To limit the number of parameters, only 4 parameters were used when filtering the score sets: ‘Amenity Type’, ‘Weight’, and the number of amenities ‘Nearest n Amenities’. Four htmls were required for the isochrones as they were based on the time to the nearest one amenity rather than a calculated score. The Kepler animation allowed for the inclusion of more parameters than the score sets as the html incorporates a filter requiring only a single html file. However, as the html file incorporates a multitude of maps, the size of the file would grow exponentially for each new parameter introduced. Due to a file size limit of 500 MB in Kepler and to improve performance of the dashboard the number of parameters was limited to the Amenity “Type”, “time”, and the “day” of the week.

The dashboard was originally published using Shinyapps.io cloud. Due to the file size limits the dashboard was published as open sourced on the r shiny server.

**4 EXPERIMENTS AND RESULTS**

***RQ1:*** *?*

.

Table example

|  |  |  |
| --- | --- | --- |
| Model | Precision | Recall |
| Basic LSTM1 | 0.99 | 0.99 |
| LogRobust[3] | 0.962 | 0.962 |
| DeepLog[6] | 0.95 | 0.96 |
| LogCluster[2] | 0.60 | - |

**Table 4: Model performance comparisons to related state-of-the-art log anomaly detection models.**   
1Basic LSTM corresponds to a 2 Layer bidirectional LSTM (see Table 3). 2Averaged from 5 final precision/recall scores reported from the original source [3].

Another table example.

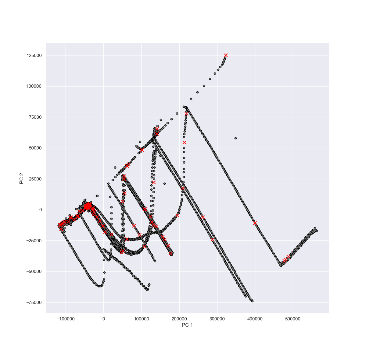
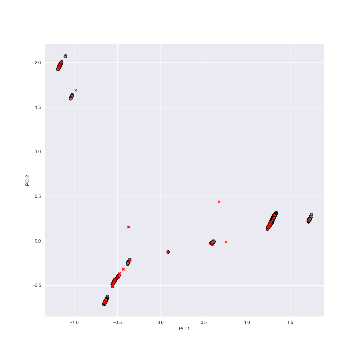
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model |  | | Precision | Recall |
| ***One Sequential Event Feature*** | | | | |
| LSTM | Bidirectional | 2 Layer | 0.99 | 0.99 |
|  | 1 Layer | 0.99 | 0.98 |
|  | Single direction | 2 Layer | 0.98 | 0.98 |
|  |  | 1 Layer | 0.98 | 0.98 |
| ***First 32 Principal Components from 128 Non-Sequential Features*** | | | | |
| Autoencoder | By event | - | 0.12 | 0.24 |
| Convolutional Autoencoder | By Window | - | 0.05 | 0.25 |

**Table 5: Comparison of model performances based on trained feature types and architectures.**

***RQ2:*** *?*

***RQ3:*** *?*

Figure example

****

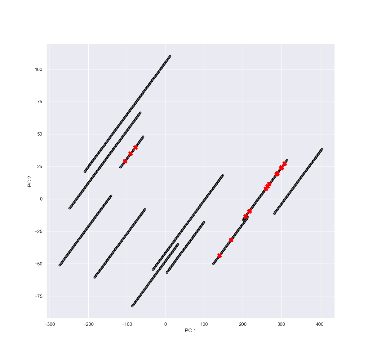
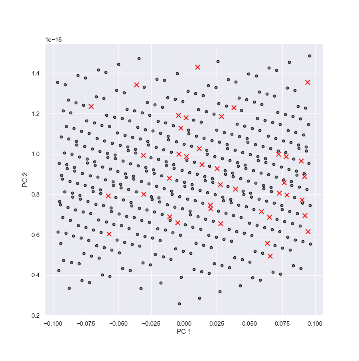
**PC2**

*Magnified*

*Magnified*

**Set B**

**Set A**

****

**PC1**

**Figure 3: First two principal components (PCs) from uninformative feature engineering.** (**A**) data includes structural, time, and message keyword features while   
(**B**) data only includes structural and time features.

**5 DISCUSSION**

Lorem ipsilum if you know what I mean

**6 CONCLUSIONS**

Lorem ipsilum if you know what I mean

**ACKNOWLEDGMENT**

We would like to thank Joseph Kuchar for his ongoing support throughout the project.

**REFERENCES (example)**

[1] Hadoop. <http://hadoop.apache.org/core>

[2] Qingwei Lin, Hongyu Zhang, Jian-Guang Lou, Yu Zhang, Xuewei Chen. Log Clustering Based Problem Identification for Online Service Systems, in Proc. of International Conference on Software Engineering (ICSE), 2016.

[3] Xu Zhang, Yong Xu, Qingwei Lin, et al. 2019. Robust log-based anomaly detection on unstable log data. In Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2019). Association for Computing Machinery, New York, NY, USA, 807–817. DOI:https://doi.org/10.1145/3338906.3338931

[4] Q. Fu, J. Lou, Y. Wang and J. Li, "Execution Anomaly Detection in Distributed Systems through Unstructured Log Analysis," 2009 Ninth IEEE International Conference on Data Mining, 2009, pp. 149-158, doi: 10.1109/ICDM.2009.60.