

Final Report & Data Dashboard

Dr. Fujian Yan

Assistant Teaching Professor

School of Computing, Wichita State University

Office: Jabara Hall 222

Office (student) hours: M 13:00-15:00

fujian.yan@wichita.edu

Outlines

- Presentation
- Data dashboard
- Final report

PRESENTATIONS

Presentations

- Proposal Presentation (20 min)
- Midterm Presentation (20 min)
- Final Presentation (30 min)
- Two check-up presentation More like individual meeting

Proposal Presentation

- Proposal Presentation (15%)
 - 20 min = 15 min + 5 min QA
 - All students should show up during the presentation
 - Preliminary data descriptions
 - Problem formulation
 - Definition of the problem
 - Math model/ Models **plan** to use
 - Tools/ software **plan** to use
 - Tasks Distribution **Plans**
 - Problems if you have any
 - Need to turn in the slides

Proposal Presentation Rubrics

- The presentation will be evaluated based on the following problems:
 1. Presentation quality (5%)
 - a. Time (2%)
 - b. Not directly read the slides (3%)
 2. Presentation context quality (10%)
 - a. Clearly define the data and features in data (5%)
 - i. Dimension of data (1%)
 - ii. Definition of the features (4%)
 - iii. Illustration with plots, figures, or etc. (1%)
 - b. Clearly define/explain the problem, the objectives (2%)
 - c. Clearly define/explain the plans or the outline of the project (2%)
 - d. Clearly define/explain the tasks distributions (1%)

Midterm Presentation

- Midterm Presentation (20%)
 - 20 min = 15 min + 5 min QA
 - All students should show up during the presentation
 - Preliminary results
 - Clearly describe the dataset
 - Progress of current project
 - Related Work (each student should present on paper regarding the topic)
 - Contribution of each group member based on the task distribution
 - Problems if you have any
 - Need to turn in the slides and the **data Dashboard**

Midterm Presentation Rubrics

- Midterm Presentation Rubrics
- The presentation will be evaluated based on the following problems:
 1. Presentation quality: Not directly read the slides (2%)
 2. Presentation context quality (8%)
 - a. Clearly understand the dataset: Can answer any questions about the dataset. (2%)
 - b. Clearly explain the selected related work by (6%):
 - i. What the method, model, or etc is proposed in the selected paper? (1%)
 - ii. How the authors use the method, model, or algorithm to solve the problem? (3%)
 - iii. Relation to your project. How do you want to use it? (2%)
 3. Data Dashboard (10%)
 - a. Informative (2%)
 - i. Please only show the selected data to support your project's objective.
 - b. Clear (4%)
 - i. Clearly display data, constant data format, and clearly define the labels or color.
 - c. Interactive (2%)
 - i. "Click to enlarge" to investigate the details.
 - d. Balance (2%)
 - i. Choose the chart reasonably. An example is that, you don't want to have a dashboard that only have numbers or vice versa.

Final Presentation

- Final Presentation (25%)
 - 30 min = 25 min + 5 min QA
 - All students should show up during the presentation
 - Propose the problem.
 - Propose the methods to solve the problem.
 - Explain what are existing resources (data) and how to apply the proposed method on the existing resources.
 - Show and explain the results.

Final Presentation Rubrics

- Final Presentation Rubrics
- The presentation will be evaluated based on the following problems:
 1. Propose the problem. (4%)
 - 1) Clearly explain to the audience what problem you try to solve. (2%)
 - 2) Why the problem is important; need to have supplement materials? (2%)
 2. Propose the methods to solve the problem. (7%)
 - 1) Clearly explain what the is the proposed method (2%).
 - 2) Why the proposed method is suitable for the current case; need to have supplement materials (3%)
 - 3) Clearly explain the math model of the proposed method (2%)
 3. Explain what are existing resources (data) and how to apply the proposed method on the existing resources. (7%)
 - 1) Clearly explain the dataset (2%).
 - 2) The data pre-processing steps. (3%).
 - 3) How did you apply the proposed model with the pre-processed dataset. (2%).
 4. Show and explain the results. (7%).
 - 1) Demonstrate the results, in plots, table, or charts. (2%)
 - 2) Explain and discuss the results (3%).
 - 3) Conclusion. How did you solve the proposed problem with the given data. (2%).

Check Up presentation

- Random select one student to present their group's progress.
- 10 mins = 6mins + 4 mins QA
- Show codes/ current results/ etc
- No need to turn in sides.

What are your expectations?



FINAL REPORT FORMAT

Final Report Format

- The final report **must** be submitted in **IEEE Trans** format. (A template has been uploaded to the BB)
- The report should include:
 - Abstract
 - Introduction
 - Related Work
 - Method
 - Results
 - Conclusions
 - Reference

Abstract

- One paragraph about major idea of the report.
 - An abstract is a short summary of your (published or unpublished) research paper, usually about a paragraph (c. 6-7 sentences, 150-250 words) long.
 - Example:

Abstract

Recently, the recommender system has been raised as one of the essential research topics in smart tourism. The massive card transaction data generated in the tourism industry has become an important resource that implies tourist consumption behaviors and patterns. However, there are challenges such as the high absence possibility of explicit feedback, which is the basis of traditional collaborative filtering techniques, and the consideration of auxiliary factors (e.g., temporal, spatial, and demographic information) that could improve the recommendation performances. In this paper, we introduce TPEDTR, a novel approach using card transaction data to recommend tourism services. It consists of two main components: (i) temporal preference embedding (TPE) models tourist groups' interactions with services chronologically to obtain their representation vectors. And (ii) deep neural network-based tourism recommendation (DTR) uses the vectors and auxiliary factors as inputs to provide tourist services. To evaluate the TPEDTR, a dataset of card transactions that happened in Jeju island, one of the most famous attractions in South Korea, over eight years is used. Experimental results demonstrate the efficacy of the proposed method and the positive effectiveness of introducing additional information on recommendation performances.

Abstract

The coronavirus pandemic has been globally impacting the health and prosperity of people. A persistent increase in the number of positive cases has boost the stress among governments across the globe. There is a need of approach which gives more accurate predictions of outbreak. This paper presents a novel approach called diffusion prediction model for prediction of number of coronavirus cases in four countries: India, France, China and Nepal. Diffusion prediction model works on the diffusion process of the human contact. Model considers two forms of spread: when the spread takes time after infecting one person and when the spread is immediate after infecting one person. It makes the proposed model different over other state-of-the art models. It is giving more accurate results than other state-of-the art models. The proposed diffusion prediction model forecasts the number of new cases expected to occur in next 4 weeks. The model has predicted the number of confirmed cases, recovered cases, deaths and active cases. The model can facilitate government to be well prepared for any abrupt rise in this pandemic. The performance is evaluated in terms of accuracy and error rate and compared with the prediction results of support vector machine, logistic regression model and convolution neural network. The results prove the efficiency of the proposed model

Raheja, S., Kasturia, S., Cheng, X., & Kumar, M. (2021). Machine learning-based diffusion model for prediction of coronavirus-19 outbreak. Neural Computing and Applications, 1-20.

Introduction

■ Background

- The tourism industry, approximately having had 10.5% of global gross domestic product (GDP), employed 340 million people, and served 1,460 million tourists globally, is considered one of the world's largest manufacturing sectors.

■ Problem

- However, temporal implicit feedback information is usually ignored because it is not obvious in the dataset and is difficult to be modeled. Implicit feedback and temporal information can promote the recommendation accuracy, and it is important to use such information [30]. The problem is to model the relationship between implicit feedback and temporal information appropriately.

■ Importance

- In the tourism industry, it has been repeatedly verified that the recommendation of tourism products is remarkably different from providing traditional items such as movies or books [23, 24, 32, 46, 55].

■ Method

- The first one is modeling implicit feedback with temporal information by introducing Doc2Vec technique and obtaining representation vectors of tourist groups and services. And, the latter uses MLP (Multi-layer Perceptron) as a Deep Neural Network (DNN) to predict preferences to tourism services by incorporating other auxiliary information with the representation vectors.

■ Contribution/Results

- In particular, extensive experiments show the positive influences of implicit preferences modeling based on temporal information and the utilization of auxiliary factors on the deep recommendation process to its performances.
- If you have numerical results, like prediction accuracy, RMS error, precision show them.
- If you have comparison study, show the comparison result, such as A outperform B in %.

Related Work

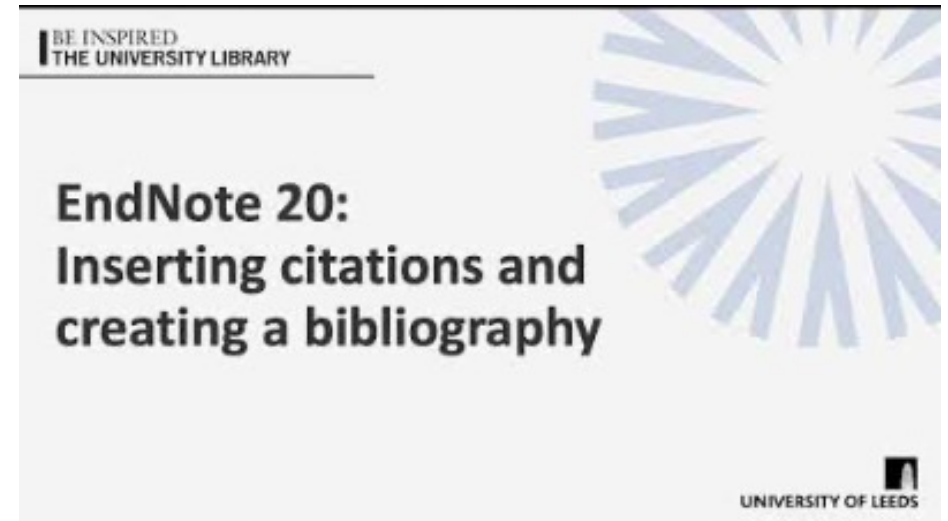
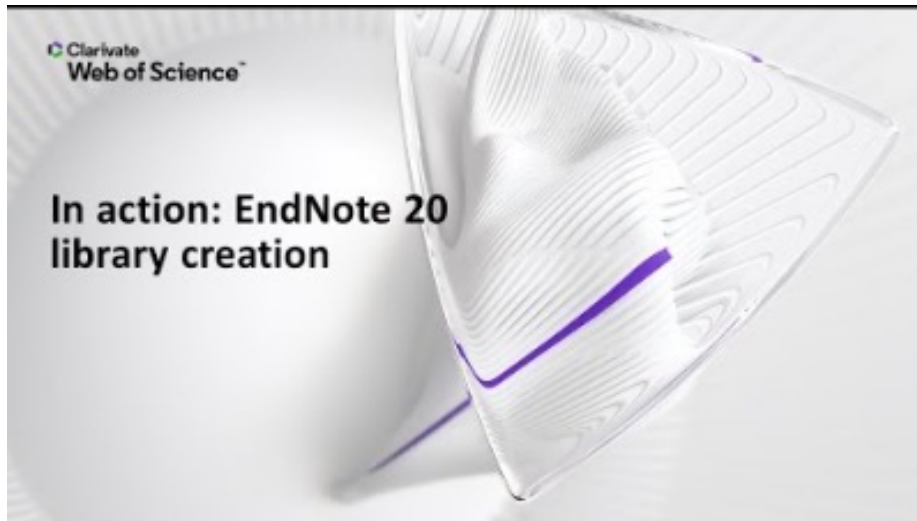
- Survey of other methods.
 - Papers regarding the methods you have used in the report.
 - One paper for one member
- Organized in logics
 - In time frame
 - In methods perspective
 - Traditional statistics methods
 - Machine Learning methods
 - Deep learning methods
- Do not directly copy
 - Try answer the following question
 - What is the problem the paper try to solve?
 - What is the method the paper proposed to solve this problem?
 - What is the results? Prediction accuracy...

Related Work

- Endnote

- Download:

- <https://www.wichita.edu/services/its/userservices/endnote.php>



Method

- Dataset description.
 - What is the dataset about?
 - Features definitions
 - Dataset insights
 - Number of features
 - ...
- Method part
 - How you formulate your objective in math model?
 - What method you have applied?
 - Method's expression
 - Explain how you use the expression in your application.

Results

- Should include:
 - Plots
 - Curve
 - Table
 - ...
- Not only show the results, but also discussion.
 - Does the results explain the objective problem?
 - If so, why?
 - If not, why?

What are your expectations?



Data Dashboard

- A data dashboard is a tool to **track, analyze, and display** data.

What are your expectations?



Summary

- Review the plan of this semester
- Dataset Preview
- Form a team next week by Tuesday.
- Project Selection