PROPOSAL

Victoria Wei, Kevin J Nguyen

Compatibility for Electronic Recommender Systems

1. PROJECT SUMMARY

Recommending new items to users is an increasingly important research topic and recommender systems are used extensively in different applications varying across domains to recommend items from books to music. e-Commerce systems such as Amazon and Netflix depend on recommender systems to increase their profits by recommending products the consumers are interested in against other products.

Current recommender systems recommend items based on two factors: user and items. For example, if a user buys a certain product, then the recommendation system will recommend similar products or products you have already purchased. For certain categories, the focus of compatibility relationship between products should be analyzed and used to recommend products to offer a complementary product, not just a similar product.

Our research proposes that compatibility can provide more accurate recommendations versus traditional recommender systems. This is especially true for electronics, so we will focus our research on electronics initially, and given time, we will progress to other categories. Through compatibility recommender systems, we will define compatibility for electronics, create a model to identify compatible products in electronics, analyze large product datasets and their relationships, and create a method to provide analytics for our results with recommender systems. Furthermore, our research differs from current market recommendation systems in that we will propose a recommendation system focused on compatibility and efficiency of the systems to provide user results.

Our overall challenge after we have created a new definition of compatibility is to analyze features that give us information about compatibility. This information includes text information, such as descriptions and product names, and image data. Another challenge is we must consider specific distinguishing features such as brand names. Next, we will find relationships between features that will give an appropriate compatible recommendation from our existing definition compared with existing recommendation that utilizes similar substitution methods.

2. INTRODUCTION

Modern recommender systems identify and understand the relationships between the items they recommend. In order to build a recommender system, a key component is that the system must have a clear definition on the relationships of items that are similar, substitutes, or complementary to develop a system that can understand a user's intentions and recommend items [2].

To identify the relationships between items, this would require defining an appropriate distance or similarity measure between items or learning from training data to develop a model. Providing some metric to measure between items is suitable for determining an equivalent

relation between items in order to recommend items that are a substitute to the item. However, a distance or similarity measure will propose issues where the compatibility between items is being considered. For example, two phone cases are similar in that they provide protection for a device and composition material, but can be entirely different due to the devices they protect. With this in mind, compatibility should be defined using how items are systematically similar in some ways but also different in other ways.

Currently other research and industry has been aimed toward analyzing the compatibility relationships between products based on their visual appearance, textual descriptions, and product names [3,4,5]. Other research has used large datasets for training and provides complex models, but follows the standard paradigm for machine learning and metric generation:

1. Collect a large dataset of related and unrelated items.
2. Create a similarity function to provide distance or similarity constant.
3. Train the function to determine related items are more similar than non-related.

In the end, these models provide a significant amount of information for distinguishing items that are similar and can range from topics of electronics to people [1]. The metric learning model is very flexible and powerful however can ignore the details where similarity should be considered. The current models themselves are not perfect and subject to limitations:

- Similarity is either defined through an explicit category tree (e.g. 'find the case nearest to this phone') and this subjects the model to noise and deficiencies in defined relations. Our model and algorithms would aim to solve this by performing recommendations without dependence on explicit relationship information.
- Model approaches are too strict in recommendation such as an item cannot be compatible with itself or do not generate a diverse set of recommendations such as recommending a similar product from a different brand. By analyzing the compatibility and relationships between products, we can handle these issues.

In our research, we will propose new models and algorithms to identify the relationships between items in product recommendation settings. The new models and algorithms will utilize our definition of compatibility to create a more relevant and accurate way to recommend items.

3. OBJECTIVES/GOALS

Our research aims to analyze the effect of using compatibility for electronics to recommend and compare to traditional recommender systems. We aim to utilize our definition of compatibility with our new models and algorithms to create a more relevant and accurate way to recommend items.

Current recommender systems recommend items that are frequently purchased together. However, there are some cases where recommendation based on compatibility is more appropriate. We will be contributing to the ongoing discussion of accurate and efficient recommender systems by analyzing compatibility to create a more efficient and accurate recommender system for electronics. Our research objectives are to define compatibility for

electronics, analyze large electronic datasets, train models to identify compatible products, and devise a method to compare our results with existing recommender systems. Our other goals are listed as follows: As part of an effort in building a research foundation in data science and analytics, we will study recent work on recommendations based on title, descriptions, and images. We will then design and implement new mining approaches for extracting high level features from these text and image data, and we will put this data into a recommender system with a new definition of compatibility to see how our recommender system functions compare to that of the existing recommender system functions.

The challenges that we have to overcome in our project are defining a new definition of compatibility, finding ways to evaluate this new definition, and defining features to use for this new definition of compatibility.

4. **METHODOLOGY/THEORETICAL FRAMEWORK**

To tackle our research questions, we will use the programming language Python to analyze the data from a public data set that will give us text and image data for electronics. We will also be using Tensorflow, a software that is an open source library for Machine Intelligence, to predict other compatible electronics. Finally, we will be digging into Neural Networks to create models for this prediction and use it across various other datasets in the future. Using several user studies will allow us to figure out which electronics are most compatible with one another. We will carry out an Ethnographic framework to understand user's behavior in predicting the next item they would like to buy. Resources needed to complete our research includes online research papers, public item datasets, and documentation on Tensorflow, Python, and Neural Networks. We will be interacting with people that are interested in being part of our user study. For our user study, we will invite them to buy a certain type of electronic and ask them what other kinds of electronics they would like to buy with the already-bought electronic. This will allow us to decide an accurate definition of what compatibility is dedicated to the general population.

5. **RESEARCH COMPLIANCE**

- Yes, we have marked YES to the following compliance questions in Part 1:
    1. Will you be collecting data by analyzing recordings (voice, video, digital, or images)? We will be collecting image and text data of electronics.
    2. Will you be observing and analyzing human behavior, either of individuals, groups, organizations, or the public? We will be analyzing human behavior of the public. Specifically, we will be observing which electronics users buy with other electronics in one shopping spree. This will determine an existing recommender system that we would like to improve on in terms of compatibility.
    3. Will you be obtaining data from surveys, interviews, focus groups, oral histories, or program evaluations? We will be using Amazon datasets to evaluate electronics and how this program evaluates how electronics are used and other products that may also be bought with certain electronics. We will also tie in a

user study that allows users to share information about what electronics are compatible with one another. This will allow us to figure out a new definition of compatibility for different types of electronics.

- No, we have not worked with Research Compliance & Biosafety to get approval for your project or to be added to an existing protocol. We plan on doing so in the near future.
- No, we have not completed any training.
- Yes, we are working with human subjects. We will be using online content from Amazon datasets to collect the public's opinion about electronic compatibility preferences.

6. **REFERENCES/BIBLIOGRAPHY/WORKS CITED**

1. M. Der and L. Saul. Latent coincidence analysis: A hidden variable model for distance metric learning. In *NIPS*, 2012.
2. G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 2003.
3. J. J. McAuley, R. Pandey, and J. Leskovec. Inferring networks of substitutable and complementary products. In *KDD*, 2015.
4. J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel. Image-based recommendations on styles and substitutes. In *SIGIR*, 2015.
5. A. K. Menon and C. Elkan. Link prediction via matrix factorization. In *ECML*, 2011.

TIMELINE

Weeks 1 and 2

Goals:

- Brainstorm ideas about what it means for something to be compatible.
- Search for literary articles, research papers, and applicable databases to get more understanding on previous research on recommender systems.
- Understand the difference between previous research on recommender systems and how our research to create a new recommender system based on compatibility is different.
- Think more in depth about research question and describe even further what is means to be compatible based on literary articles.
- Start thesis statement for thesis, begin creating a thesis template, and download the thesis template that we want to use throughout the program.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss thesis template, installments, progress reports, thesis statement, and research progress.
- Discuss with Dr. James Caverlee about presenting work publicly in Infolab in the following weeks after we have finished reading research papers on our topic. We will be giving a presentation about previous research papers that we read and what we have learned. This will allow other students in Infolab to understand our current research topic and how our research topic is different from previous research papers.

- Discuss with Dr. James Caverlee about research compliance approval and whether or not we need research compliance approval with our current research project and future plans. If the answer is yes, we will need to attend a Research Compliance Informational and contact the office of Research Compliance & Biosafety and complete training/fill out forms to continue.

Weeks 3 and 4

Goals:

- Update your proposal in preparation for the first installment submission. Submit corrections as requested.
- Focusing on importing, collecting, evaluating, reporting, and processing Amazon datasets into our hands.
- Keep track of notes about key findings and important relevant articles about how to import datasets and where to import them in the future.
- Update our thesis outline based on what methods we used to import data and where we currently store that data for later experiments.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss the content of our thesis, research progress, our public presentation information, installments and progress reports. If they request changes, submit those changes.
- Make sure that by this time, we should have received all research compliance approvals, been added to all of the necessary protocols, and completed all required training necessary to continue with our project. In this way, we can grab the data.
- Submit first installment and progress reports detailing our current progress and our future goals.

Weeks 5, 6 and 7

Goals:

- Importing the already retrieved Amazon dataset into existing recommender systems and understand the environment in different applications we are using. We will also read more existing documentation on how to import the dataset and how to use, experiment, and apply recommender system applications. This will be useful later on when we create our own definition of compatibility to create our own recommender system application.
- Keep track of notes about key findings and important relevant articles on previous recommender systems and how ours is different.
- Update our thesis content to reflect previous recommender systems and how ours will compare in the future.
- Make corrections to first installment if requested.
- Create draft content for second installment and progress report submissions. Include experiments with existing recommender systems and readings on these previous recommender systems. Make changes if applicable.
- Created notes about project presentation information. This should include previous research articles about existing recommender systems, current research proposal, and experiments done

with existing recommender systems and how our research will compare with these existing systems. Created speech for presentation.

- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress and written thesis content, project presentation information, installments, progress reports, and findings from our experiments and readings.

Weeks 8, 9 and 10

Goals:

- Completed public presentation to Infolab.
- Focusing on defining a new definition of compatibility and what it means for electronics to be compatible with one another.
- Read compatibility articles and existing research papers on compatibility in recommendation systems involving different topics.
- Document findings on compatibility for electronics as well as other topics and see if there are differences and similarities between different areas.
- Continue draft content for second installment and progress report submissions. Include compatibility for electronics and compatibility definition with difference across various topics. Make changes if applicable.
- Update our thesis content to reflect this new definition of compatibility and how this compares to existing definitions.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress and written thesis content, installments, progress reports, and compatibility results.
- Begin project submission to Explorations: the Texas A&M Undergraduate Journal (http://explorations.tamu.edu). Incorporate new ideas found during these next few weeks.

Weeks 11, 12, 13, 14

Goals:

- Focusing on this new recommender system to create a new definition of compatibility that will use this definition to predict electronic suggestions.
- Testing this new recommender system with our dataset and analyzing the results retrieved from this new recommender system to see if our results are more efficient and accurate than previous recommender systems.
- Update our thesis content to reflect new recommender system design and how ours compares to previous recommender systems in turns of design, performance, and accuracy.
- Document our findings from our results of using our new recommender system to see if our results are more efficient and accurate than previous recommender systems. This should conclude our literature review.

- Continue draft content for second installment and progress report submissions. Include the new recommender system design and the results from testing the new recommender system with compatibility incorporated.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress, written thesis content, installments, progress reports, new recommendation system design, and the results from this new recommendation system.
- Submit project to Explorations: the Texas A&M Undergraduate Journal (http://explorations.tamu.edu). Incorporate new ideas found during these next few weeks.

2. Spring semester: Starts on January 16

Weeks 1 and 2

Goals:

- Update thesis outline to incorporate difficulties that we had to overcome in our research, data, and experiments and the solutions that we ended up choosing to overcome these difficulties.
- Begin draft Bibliography using our cumulative notes.
- Continue comparing our new recommender system with existing recommender systems, noting more results and conclusions.
- Document any key findings from our experiments from new vs. existing recommender systems. Finish documentation here.
- Continue draft content for second installment and progress report submissions. Include more key findings from our experiments and results from testing the new recommender system vs. existing recommender systems.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress, written thesis content, Student Research Week presentation advice, installments, progress reports, and continued experiments between our new recommendation system vs. existing recommendation systems.
- Begin creating presentation information to present at Student Research Week in March.
- Register to present at Student Research Week in March.

Weeks 3 and 4

Goals:

- Continue creating draft thesis project outline and Bibliography.
- Finish draft content for second installment and progress report submissions. This includes the updated thesis outline and Bibliography.
- Begin drafting content for third installment and progress report submissions.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress, written thesis content, Student Research Week presentation advice, installments, and progress reports.

- Continue creating presentation information to present at Student Research Week in March.

Weeks 5 and 6

Goals:

- Make changes to Bibliography if needed.
- Make changes to thesis outline if needed.
- Continue drafting content for third installment and progress report submissions.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress, Student Research Week in March, written thesis content, installments, and progress reports.
- Continue creating presentation information to present at Student Research Week in March.
- Practice presenting in front of Infolab and note the suggestions that they give and take advice into account.

Weeks 7 and 8

Goals:

- Finish creating presentation information to present at Student Research Week in March.
- Finish practice presenting for Student Research Week in front of peers.
- Attend Student Research Week and present project!
- Continue revising thesis content.
- Finish drafting content for third installment and progress report submissions.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress, Student Research Week in March, written thesis content, installments, and progress reports.

Weeks 10 and 11

Goals:

- Begin drafting the presentation report through eCampus.
- Send the latest draft of our final thesis to Dr. James Caverlee for review before the April 9 deadline.
- Make corrections to third installment if needed.
- Continue to revise final draft of thesis.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress and written thesis content.

Weeks 12 and 13

Goals:

- Submit public presentation report.
- Continue revising our final thesis. Pay particular attention to the Common Mistakes page in the Thesis Manual and make sure your document is uniformly formatted throughout.
- Talk to Dr. James Caverlee about the appropriate embargo selection for your thesis.

- Attend a drop-in thesis tutorial session. Understand the final thesis submission process.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress and written thesis content.

Weeks 14 and 15

Goals:

- Talk to your faculty advisor about the appropriate embargo selection for your thesis.
- Once we are approved in eCampus, complete our final submission in the Scholar's Thesis Submittal System (STSS). This includes uploading a completed abstract, keywords, and embargo selection.
- Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss research progress and written thesis content.
- Finish revising our final draft for our thesis.
- Complete the program exit survey.
- Make corrections to final thesis if needed.
- Reminder for May 2019 to RSVP for the LAUNCH Recognition Ceremony to receive URS medallion.