COMPATIBILITY FOR ELECTRONIC RECOMMENDER SYSTEMS

An Undergraduate Research Scholars Thesis

by

KEVIN J NGUYEN and VICTORIA WEI

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. James Caverlee

Approved by Graduate Student Advisor: Yin Zhang

May 2018

Major: Computer Science

TABLE OF CONTENTS

ABSTRACT

Compatibility for Electronic Recommender Systems

Kevin J Nguyen and Victoria Wei
Department of Computer Science
Texas A&M University

Research Advisor: Dr. James Caverlee
Graduate Student Advisor: Yin Zhang
Department of Computer Science and Engineering
Texas A&M University

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.

1. 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 [1].

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 equivalence relation between items. This is to ensure that we recommend items that are considered substitutes to that 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 they can be entirely different due to the devices they protect. With this in mind, compatibility should be defined by 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 [2, 3, 4]. Other research has used large datasets for training and provides complex models, but they follow the standard paradigm for machine learning and metric generation:

- Collect a large dataset of related and unrelated items.
- Create a similarity function to provide distance or similarity constant.

- 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 [5]. The metric learning model is very flexible and powerful. However, it 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.

2. RELATED WORK

We relate our work in the context of prior studies and implementations of (1) item-to-item recommendations, e.g. systems that generate item-to-item recommendations by analyzing the relationship between items; and (2) studies involving metric learning including those not for recommendation.

2.1 Item-to-Item Recommendation

The analysis of the relationship among items is fundamental to modern real-world recommender systems, e.g. to generate recommendations of new songs on *Spotify*. Most recommender systems utilize methods based on collaborative filtering, e.g. counting the overlap between users who have liked both songs, as in *Spotify's* own solution [6]. Other approaches include the use of latent-factor approaches that aim to model user-item relationships with low-dimensional factors to find recommendations with close embeddings. The systems that are able to predict the item-to-item relationship based on *content* are most important to our research as the systems are proposed to address specific topics.

Our work correlates with the co-purchase and co-browsing relationships using a dataset provided by *Amazon* [1]. We utilize previous works on these systems to provide quantitative results and compare our work against them. The main contribution of our work is to relax the model assumptions of current co-purchase and co-browsing systems to allow more complex relationships between items.

2.2 Metric Learning

The analysis of the relationship between objects is a vast topic that covers more topics than recommender systems. In modern learning, one is given a collection of relationships between items, and the goal is to identity a function that matches these relationships. The function must be able to generalize the relationship between objects and apply them to new unseen items to predict new relationships. The function is measured against valid data, and the metrics show how accurate the model can identify the relationships. The most developed and advanced learning methods are used to identify hidden variables or factors among items, through matrix-factorization or collaborative filtering. Again, our main goal is to relax the model assumption of current models and allow for more complex notions of 'relatedness'. There are algorithms that work with non-metric learning of relationships, but to the extent of our knowledge do not scale well with larger sets of data.

Bibliography

[1] B. S. G. Linden and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet Computing*, 2003.

[2] R. P. J. J. McAuley and J. Leskovec, “Inferring networks of substitutable and complementary products,” *KDD*, 2015.

[3] Q. S. J. J. McAuley, C. Targett and A. van den Hengel, “Image-based recommendations on styles and substitutes,” *SIGIR*, 2015.

[4] A. K. Menon and C. Elkan, “Link prediction via matrix factorization,” *ECML*, 2011.

[5] M. Der and L. Saul, “Latent coincidence analysis: A hidden variable model for distance metric learning,” *NIPS*, 2012.

[6] M. Mithun, “Music recommendation system spotify - collaborative filtering,” *MUS-17*, 2017.