New York University

INDIV-UG 1905

Gallatin School of Individualized Study, Fall 2015

Senior Project Proposal

Student:
Peter Mountanos
(*pjm*419)

Mentor:
Panogiotis Adamopoulos
(*padamopo*)

May 1, 2015

1 Introduction

Recommender Systems have been a prevalent area of research since the mid-1990s, beginning with the first papers on collaborative filtering (Adomavicius & Tuzhilin, 2005). Since then, recommender systems have been applied to various dimensions, ranging from marketing, education, social media, financial services, and more. At a high-level, recommendation systems are pieces of software that aim to recommend products or information to users, based on certain preferences.

This research project is centered on two publicly available datasets from Twitter concerning social commerce and "Pay-by-Tweet" services. The project has three main goals: (1) to show a connection between users' personalities and their purchasing decisions; (2) to correlate purchases with a user's influence over their followers; and (3) to develop a recommender system which estimates the probability of purchase for each user and item grouping and then recommends the items with the highest probabilities. This research will culminate in a final paper that will discuss the experimental methodology and intuition behind the recommender system. This proposal will cover my background, the motivation for the project, and the project's expectations.

2 Background

During my time at Gallatin, I have studied the deeply intertwined nature of computer science and mathematics, and how to utilize these two fields to analyze and learn from data. In today's terms, we recognize this as data science. In a truly interdisciplinary nature, data science research has implications way beyond computer science and mathematics alone, as these are just the tools employed to solve some of the most challenging problems in our world today. In fact, data science has been used to solve problems in genomics, finance, public policy, and even high-energy physics. Data science has grown exponentially over the past few years in part due to new developments made in machine learning. Recommender systems are a highly active area of research within the broader field of data science and machine learning. I hope to add valuable research and insight to this research community through my senior project.

The motivation for this project developed out of the work done in my two previous independent studies with Tuzhilin and Adamopoulos: *Recommender Systems* and *Recommender Systems II*. *Recommender Systems*, which took place in the Fall 2014 semester, mainly focused on developing an understanding of the recommender systems literature at a more general level. We covered topics such as content-based recommender systems, collaborative filtering (both memory-based and model-based), dimensionality reduction, online learning, context-aware recommender systems, and the evaluation recommender systems. I took a course, *Data Mining for Business Analytics*, in the Spring 2014 semester with Adamopoulos, which motivated this original independent study. *Recommender Systems II*, which is currently in progress this semester, has focused on understanding ensemble learning theory, and how to apply it to recommender systems. We are currently working on a project to provide more diverse recommendations by modifying a standard k-NN approach to select neighbors that minimize generalization error, as determined by the bias-variance-covariance decomposition as described by Ueda and Nakano.

3 Motivation

Over the past 20 years, there has been much research within the field of recommender systems (RS), and a wide variety of algorithms have been designed to improve recommendations (Adomavicius & Tuzhilin, 2005). There has been much social and business success as a result of the use of recommenders, due to the fact that we live in an age of information overload. In this age of information overload, we use many strategies to make choices about "what to buy, how to spend their leisure time, and even whom to date." We rely on recommender systems to automate some of these decisions with hopes of gaining high-quality, personal recommendations (Jannach, Zanker, Felfernig, & Friedrich, 2010). While recommenders have achieved acceptance, there is still much work to be done to create more useful recommendations.

Social media represents one of the most transformative impacts of information technology on business because it has drastically changed how consumers and firms interact (Aral, Dellarocas, & Godes, 2013). Thus, we see that companies are constantly competing in this space for consumers' attention and brand engagement (Adamopoulos & Todri, 2014). Recently, both American Express and Amazon partnered with Twitter to introduce a novel way of connecting social media and e-commerce. American Express, directly through Twitter's platform, allowed customers to buy products, while Amazon allowed them to add products to their Amazon carts. However, both of these methods allowed the companies to capture the power of social media, by spreading the word about the service and their products.

We hope to use publicly available data from Twitter collected on these two services to help better understand users' purchasing decisions. By leveraging the power of users' past tweets, we can better understand the personalities of these users. Thus, we can then try to correlate personality types to certain product purchases. Further, we can also take advantage of a user's social network by observing their followers. By determining the influence a user has over his or her followers, we can observe how that influence dictates purchasing decisions by a user's network. Through these two factors, we can model a probability of purchase for each user and item pair. Using a hybrid recommender technique, we can then recommend items with high purchasing probabilities to corresponding users.

4 Expectations

The output of this project will be a research paper that discusses the results of the three main project aims as discussed in the introduction. In order to ensure substantial progress is being made throughout the semester, the student and mentor will generally meet once a week in a conference room on the 8th floor of KMEC, which hosts Stern's Department of Information, Operations, and Management Sciences. We hope to submit this paper to several conferences, such as RecSys and WWW, to receive peer feedback on the experimental design and results[1].

The project that has been described above is very much attainable to complete in one semester. Since the past two semesters have been spent engaging with the field's literature, much of the time can be spent on the experimental process. This, coupled with my previous research experience at Brookhaven National Laboratory[2], gives me full confidence that I will be able to complete this project within one semester. The next page provides a general schedule that will be followed to assure that the project goals can be completely in a timely manner.

---

[1] Some of these conferences will take place after the deadline for this project, but it is still critical to receive such feedback.

[2] I conducted high-energy particle physics research at Brookhaven National Laboratory on Long Island, NY in 2011-2012. There, I worked as a part of the Electric Dipole Moment (EDM) Collaboration, and wrote an awarded paper entitled "The Optimization of Spin Precession and Beam Polarization for the Proton Electric Dipole Moment Experiment".

4.1 Schedule

We provide a general schedule to help display the feasibility of completing the project within one semester. To our advantage, we have begun to consider important factors of the experiment this semester, and will continue to develop these ideas throughout the summer. Thus, we plan to have the Twitter data for users' timelines and their social networks' timelines scrubbed and mined by the end of September. Therefore, by Fall Recess we plan to start utilizing IBM Watson's Personality Insights to determine personality traits of the users (see Section 4.2). We will then spend the rest of October concentrating on building our recommender system and fine tuning our models. This is where we will start to obtain most of our experimental results. Then, we will spend much of the month of November writing and revising the research paper that will discuss our observations made from testing, as well as the theoretical background underpinnings of our recommender. Finally, we plan to submit the paper to be considered for honors by Tuesday, December $1st$, which is two weeks prior to the last day of Fall classes.

4.2 Development Stack

When developing a recommender, much of its roots lies in mathematics and information theory. However, it is absolutely imperative to test and evaluate the algorithm on several datasets. In this case, we will be developing the algorithm using the Python programming language. Python has exceptional libraries for machine learning (`sklearn`[3]) as well as numerical and scientific computing (`numpy`[4] and `scipy`[5]). We will also be utilizing IBM Watson's Personality Insights API[6] for determining users' personality characteristics.

The two datasets we will be using were obtained from publicly available data on Twitter. One dataset involves purchases made by users using the hash tag `#AmazonCart`[7]. The second dataset contains tweets corresponding to all of the offers through American Express[8] (e.g., hash tags like `#AMEXSamsClub`).

[3]See `http://scikit-learn.org/stable/`
[4]See `http://numpy.org/`
[5]See `http://www.scipy.org/`
[6]See: `http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/personality-insights.html`
[7]See: `http://www.amazon.com/gp/socialmedia/amazoncart`
[8]See: `http://people.stern.nyu.edu/padamopo/blog/2013-02-27-AmExTwitter.html`

5 Conclusion

We hope to provide meaningful research to the recommender systems community through this project. The field of recommender systems is extremely interdisciplinary, with research spawning from computer science, information systems, management sciences, psychology, economics, and marketing. In an age of information overload, providing more accurate and diverse recommendations is more important than ever–especially when we know that algorithms can dictate consumer preference (Bettman, Luce, & Payne, 1998). By exploring social commerce data, we hope to gain insights into how users' personalities and social power dictate purchasing decisions. We can then exploit this information to better recommend relevant products to users. Further, we hope to provide a recommender that is more aware to human decision-making processes. Gallatin's mission, as taken from the Gallatin website[9], is:

> A Gallatin education is designed to help students become life-long learners by developing their capacities for creative self-development, for self-reflection about their aspirations, practices, and the worlds they inhabit. Gallatin's approach to student learning is therefore holistic, individualized, and interdisciplinary.

I hope to finalize my holistic, individualized, and interdisciplinary education through this project. If any further information can be provided, please contact Peter Mountanos[10] or Panagiotis Adamopoulos[11].

Please see the following page for a listing of relevant sources. Note, we have provided annotations for eleven of the works. We include several other works that we find to be important and will consider throughout the research process. However, even still this list is not exhaustive, and there will be likely many more citations in the final paper.

[9]See `http://gallatin.nyu.edu/about/mission.html`
[10]Email: `peter.mountanos@nyu.edu`
[11]Email: `padamopo@stern.nyu.edu`

References

Adamopoulos, P., & Todri, V. (2014). Social Commerce: An Empirical Examination of the Antecedents and Consequences of Commerce in Social Network Platforms. In *Thirty Fifth International Conference on Information Systems.*

> This paper covers a previous experiment done on the American Express & Twitter dataset. Our current research hopes to build on this project. It discusses the integrating of e-business with social network platforms, and how certain characteristics of users affect their decision to participate in these services.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey on the State-of-the-Art and Possible Extensions. In *IEEE Transactions on Knowledge and Data Engineering* (Vol. 17, pp. 734–749).

> This is an essential paper to the recommender systems community; it has been cited over 5,600 times. It provides an overview of the field of RSs and provides a framework for thinking about RSs in three ways: content-based, collaborative, and hybrid. Understanding what is discussed in this paper is important for any researcher in recommender systems.

Ansari, A., Essegaier, S., & Kohli. (2000). Internet Recommendation Systems. *Journal of Marketing Research*, *37*(3), 363–375.

Aral, S., Dellarocas, C., & Godes, D. (2013). Introduction to the Special Issue-Social Media and Business Transformation: A Framework for Research. *Information Systems Research*, *24*(1), 3–13.

> This paper investigates the relationship between social media and business transformation. It provides useful conclusions on how social media, business, and society interact. They provide a framework for this relationship that is essential to understanding the context of the "Pay-by-Tweet" services.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998, Dec). Constructive Consumer Choice Processes. *Journal of Consumer Research*, *25*(3), 187–217.

Bruce, N., Foutz, N., & Kolsarici, C. (2012). Dynamic Effectiveness of Advertising and Word of Mouth in Sequential Distribution of New Products. *Journal of Marketing Research*, *49*(4), 469–486.

Brynjolfsson, E., Hu, Y., & Simester, D. (2011). Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science*, *57*(8), 1373–1386.

> This paper explores the Long Tail phenomenon that is unique to the Internet and information technology. It provides another source to view the implications of recommender systems through an economic lens. We wish to take an inter-disciplinary approach to this algorithmic development, rather than only consider the mathematical underpinnings.

Cooke, A. D., Sujan, H., Sujan, M., & Weitz, B. A. (2002). Marketing the Unfamiliar: The Role of Context and Item-Specific Information in Electronic Agent Recommendations. *Journal of Marketing Research*, *39*(4), 488–497.

Dellarocas, C. (2006). Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms. *Management Science*, *52*(10), 1577–1593.

Fleder, D., & Hosanagar, K. (2009, May). Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science*, *55*(5), 697–712.

> Fleder and Hosangar explore issues that collaborative filtering algorithms have with diversifying recommendations. They describe the rich-get-richer effect, and how recommenders can decrease aggregate diversity. Further, they discuss important design choices that can increase concentration bias.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., . . . Verlegh, P. (2005). The Firm's Management of Social Interactions. *Marketing Letters*, *16*(3–4), 415–428.

> This paper discusses social interactions, and how face-to-face communications and word-of-mouth (WOM) communications should both fit under this term. They show that due to the effectiveness of WOM at driving sales and generating leads, marketers have been attempting to engineer WOM, rather than expect the discussions to occur naturally. This is an important implication to consider for our project.

Häubl, G., & Murray, K. B. (2003). Preference Construction and Persistence in Digital Marketplaces: The Role of Electronic Recommendation Agents. *Journal of Consumer Psychology*, *13*(1-2), 75–91.

Hosanagar, K., Fleder, D., Lee, D., & Buja, A. (2013). Will the Global Village Fracture into Tribes? Recommender Systems and their Effects on Consumer Fragmentation. *Management Science*, *60*(4), 805–823.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2010). *Recommender Systems: An Introduction*. Cambridge University Press.

> This book, like the *Recommender Systems Handbook*, provides approaches for developing state-of-the-art recommender systems. It has case studies, and several algorithmic approaches. Understanding past approaches is necessary for developing better algorithms.

Oestreicher-Singer, G., & Sundararajan, A. (2012a). Recommendation Networks and the Long Tail of Electronic Commerce. *MIS Quarterly*, *36*(1), 65–83.

> This paper provides a discussion of the Long Tail, and the conjecture that recommendation agents in e-commerce will redistribute the demand of popular products to less popular, “niche”, products. It allows us to think of the recommender systems through an economic lens, with discussion of the Lorenz curve and the Gini coefficient.

Oestreicher-Singer, G., & Sundararajan, A. (2012b). The Visible Hand? Demand Effects of Recommendation Networks in Electronic Markets. *Management Science*, *58*(11), 1963–1981.

Pariser, E. (2012). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Books.

> This is an extremely thought-provoking discussion on the implications of recommender systems. Pariser proposes that due to recommenders, we will increasingly live in a world of filter bubbles, which is when algorithms selectively guess what information we’d like to see based on our preferences, which essentially isolates us from content that disagrees with our own viewpoints. This confirms the necessity to algorithms which promote diversity and eliminate concentration bias (our goal).

Park, Y.-J., & Tuzhilin, A. (2008). The Long Tail of Recommender Systems and How to Leverage It. In *Proceedings of the 2008 ACM Conference on Recommender systems* (pp. 11–18).

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender Systems Handbook*. Springer.

This handbook is an essential reference piece for any researcher studying recommender systems. It provides an multi-disciplinary view on recommender systems, with experts from artificial intelligence, human computer interaction, information technology, data mining, statistics, marketing, consumer behavior and more. This book provides useful information when making design decisions for the algorithm.

Xiao, B., & Benbasat, I. (2007, Mar). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly*, *31*(1), 137–209.

Xiao and Benbasat provide an analysis of recommender systems beyond trying to just develop better algorithms. They consider other essential factors such as user-recommender interaction, recommender characteristics, recommender use, and provider characteristics. Further, they provide an outline of empirical studies within the recommender systems community, which is extremely useful when needing to reference other papers.