POPULARITY BIAS IN SEQUENTIAL RECOMMENDATION

COMPARISON AND EVALUATION OF USER PRIVACY RISK IN RECOMMENDATION SYSTEMS

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Shivaen Ramshetty

November 2, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS
Kathryn A. Neeley, Department of Engineering and Society
Hongning Wang, Department of Computer Science

THE CLASH BETWEEN IMPROVED RECOMMENDATIONS AND PRIVACY

The recent impressive growth of the recommender systems has profoundly shaped the online marketplace. Wider application of this system has translated into the fact that "30% of Amazon's page views result from recommendations" and "80% of the content watched by Netflix subscribers comes through personalized recommendations" (Adomavicius et al., 2018, pg. 2). It can be noted from the above that recommendation systems are employed throughout a variety of applications: e-commerce, video streaming, natural language processing (NLP), and other content platforms are some such areas they can be found (Kumar et al., 2014). From Amazon's product recommendations to Netflix's "Top Picks for You," predicting user preferences or offering the next suggestion is growing more and more paramount, since the potential profit gained from each individual user grows when their recommendations more accurately assess their needs/wants. However, the broadening of recommender systems raises questions about how such systems influence and otherwise affect their users.

Firstly, models that build such systems are greedy for data, which means that these models get more accurate with larger sets of data. The problem with more data in regards to recommendation systems is that the data collected is specific to individual users, which may include information such as their purchase history, location, or previous interests (views). Some users may oppose such data practices, but for others the "benefits of getting tailored content outweighed any privacy concerns" (Harley, 2018). However, the aforementioned greedy nature of models could lead to the collection of much more sensitive data, which could put the user at risk in the case of a hack or leak of the system. Another issue surrounding these recommendation models is that they are susceptible to bias. Bias in a model suggests that it is not versatile and chooses to fit a certain set of data better than the rest. This leads to two problems: a shift in the representation of the user's taste over time and the majority dominating the minority (Mansoury et al., 2020). The latter expresses the idea that groups with larger available data or quality of data will steer the model to serve them better than others. On the other hand, the former suggests that bias susceptible models are not effective over time or for a diverse consumer base (Adomavicius et al. 2018). These issues must be reconciled for the success of the recommender system to continue, there must be a balance between reducing bias of models and risk the users are willing to take on.

My project addresses these concerns and proposes methods to alleviate their effects on the system. The technical research will create sequential models that adapt previous models in order to find alternative methods of reducing bias while maintaining or even increasing accuracy. This will first require an understanding of the prior research in question and then a dive into where the drawbacks exist. In my STS research, I will focus on the extent to which recommendation systems pose a risk to its users and whether or not the concern of privacy for the system is well-founded.

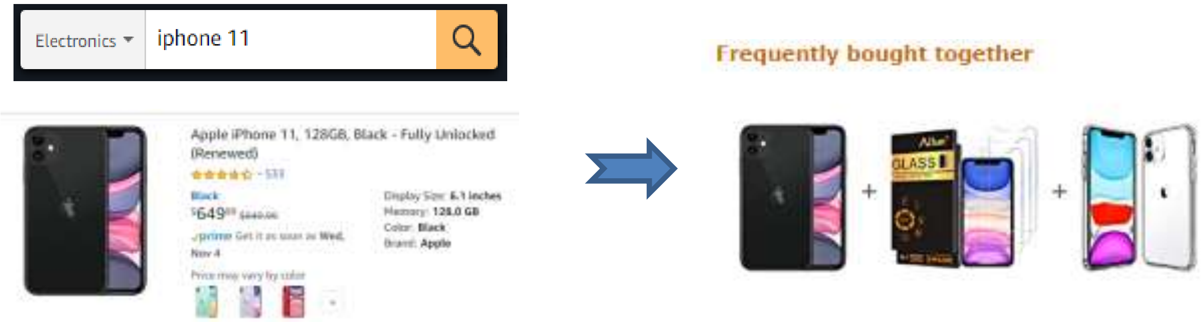


Figure 1. Real-world example of a recommendation following the phone example detailed above. When user looks for phone to buy on Amazon, they are recommended with the package of a phone, case, and screen protector.

BUILDING ROBUST PRODUCT RECOMMENDATION MODELS THAT ACCOUNT FOR SEQUENTIAL PATTERNS

Currently, market conditions have set the stage for the "recommendation engine market to grow from USD 588.9 Million in 2016 to USD 4414.8 Million by 2022," which has drawn considerable interest to what they are made of (Markets, 2018). The algorithms that construct these systems involve machine learning models that take into account user data to predict the next preference of an individual. For example, suppose a user visits Amazon looking to buy a phone. What may they look to buy next based simply on this action? Taking a look at Figure 1 under the header "Frequently bought together," it is shown that Amazon believes this user would likely also want to buy a case as well as a screen protector. This series of purchases seems obvious and trivial, but what makes these models so powerful is that they find such patterns over a vast number of products and people. Many of these patterns have sequential like constructions, where one action is related to the next and so forth. Yet, past systems focused solely on grouping users based on similarities, known as collaborative filtering, thereby not taking full advantage of the sequential activity of their users (Techlabs, 2017). However, recently there has been interest in sequential models due to "conventional approaches always ignor[ing] to consider the sequential dependencies among the user's interactions, leading to inaccurate modeling of the user's preferences" (Fang et al., 2020, pg. 1).

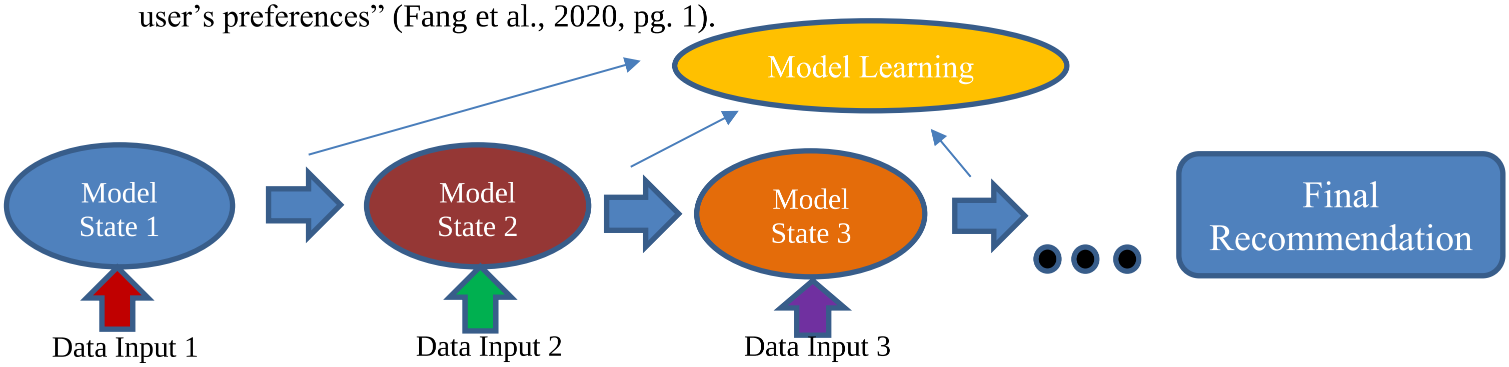


Figure 2: A simple sequential model. The model takes in input after input and updates the final recommendation when it finds a pattern it recognizes. The order and type of the input can alter what the recommendation is since the model has learnt the pattern of inputs/actions that map to particular preferences.

As seen in Figure 2, a sequential model is one which takes in data as a stream, one input after another; in doing so, the model can adapt its output to the particular sequence of actions a user takes. This versatility allows for far greater individualized recommendations than clustering models such as the aforementioned collaborative filtering. With a growing number of applications of this model being put in practice, it is clear why sequential learning has grown very rapidly. Furthermore, since the likelihood that those using the system are not a uniform population is very high, there is a need for a system that addresses the bias of past systems. In sequential models each user is not fit into groups of similar preferences; rather, the model learns a wide range of patterns and fits a pattern to the individual user (Fang et al., 2020, pg. 5). This reversal of importance from groups to user allows the model to avoid bias, to a certain degree, by not serving to find a similar majority for the user to fit into. Rather, the model chooses the most accurate next preference for a user from those it has learnt. However, bias is still present within such systems since no model can learn every possible combination of actions; in short, they can only choose from experienced permutations. An additional solution may arise from the following question: is there a way to learn more about the user and their actions from the same data?

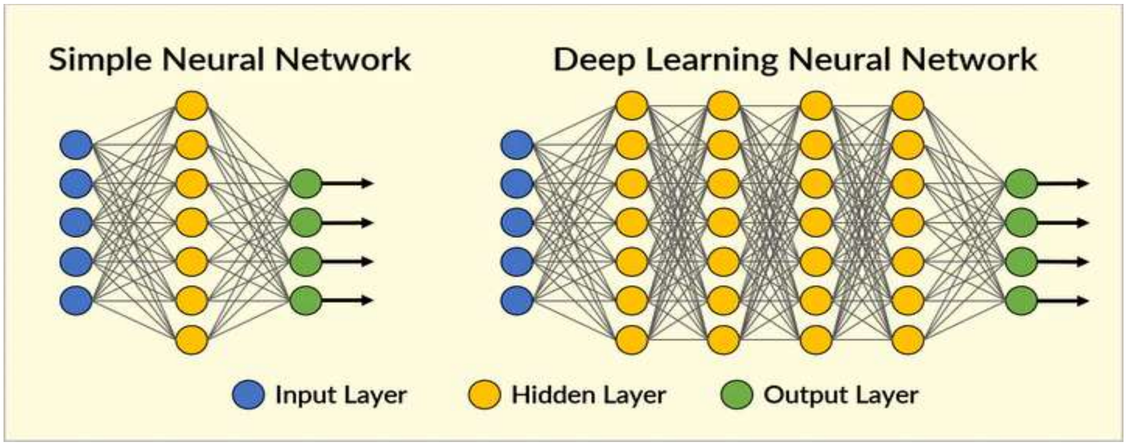


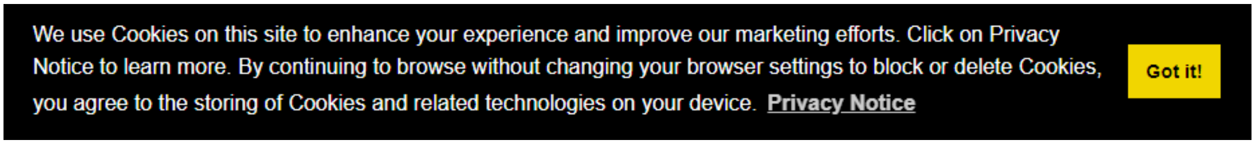
Figure 3. Depicting different types of neural networks. Source: https://www.securityinfowatch.com/video-surveillance/video-analytics/article/21069937/deep-learning-to-the-rescue

To gain more insight from the same information, there needs to be more complexity within the model; this is where deep learning comes into the picture. Deep learning is the concept where the models in question make use of a structure called neural networks. Neural nets are a type of computing system where the data is shared in a web like structure, similar to a brain, hence the term “neural net”. The “deep” in the phrase deep learning, is associated with the greater complexity of the networks, which manifests in the form of layers as seen in Figure 3. Deep learning neural nets consist of two or more hidden layers, which allows the model to learn many more representations of the same data due to the increase in number of shared connections. By integrating deep learning into sequential models, our project takes advantage of the ability to learn more about each user, to then meet the user’s preferences more accurately. In summary, sequential models can help build long/short-term relationships while deep-learning can aid in building the complexity necessary for models that are suitable for a larger more heterogenous audience (Low et al, 2019, pg. 8).

The newfound capacity of deep learning models brings a new frontier to the industry, where we can gain more quality insight on individuals unlike collaborative filtering models (Kumar et al., 2014, pg. 5256). For this project, my focus will be building/improving a robust deep learning model that also utilizes the principles of a sequential learning. With the use of prior research, I will first evaluate the strengths of various approaches and in which scenarios they perform the best. Finally, the research will evaluate the constraints of the implemented models to establish possible future work; in doing so, the project will test what may come to the market in the near future and produce alternate directions on which research could continue.

RISK OF USER PRIVACY IN RECOMMENDER SYSTEMS

For any part of the recommender system to work, there must be access to one common resource: data. Data has been thrust into the forefront of the online ecosystem in recent years due to the monetary value it holds for large companies. Provided by users willingly or not, entities all over the world use it to improve user experience within an application, or buff profits. The real problem is whether or not the system poses an inherent risk to users through its data collection practices.



We use Cookies on this site to enhance your experience and improve our marketing efforts. Click on Privacy Notice to learn more. By continuing to browse without changing your browser settings to block or delete Cookies, you agree to the storing of Cookies and related technologies on your device. Privacy Notice

Figure 4: Example of a privacy consent form from https://owl.purdue.edu/owl/

Currently, the growing push towards data regulation and protection threatens the freedom and innovation the recommendation system arena has seen, as Wang states, “privacy is … [a] serious security concern of recommender systems, and it has gained enormous attention” (Wang et al., 2015, pg.17). Firstly, most consumers acknowledge that their data is being stored but “feel various dimensions of control over personal information collection are ‘very important’ to them” (Madden et al., 2019). In other words, some users would like to manage the risks that their data faces by having some sort of control on what is collected. However, there is very little the average individual can do to protect themselves and 59% of the them are unaware of the way their experience is shaped around them (Auxier et al., 2020). To close this gap, regulators at the government level have pursued legislation that “introduce explicit guidelines and sanctions to regulate data collection, use, and storage” (Milano et al., 2020). Lately, online mediums have followed suit and instituted privacy forms alerting users of what they are collecting as seen in Figure 4. This gives users clarity on what is being collected and the ability to decide if they want to use the full functionality of the online content at some risk. Recommender systems recent success portrays the notion that users don’t consider it a large enough risk to avoid participating in, but with ever changing policies this current trust or indifference of consumers may flip.

As mentioned before, recommender systems keep track of user clicks, views, purchases, and many other actions to try and predict what the user may want to see. The conflict that arises is best expressed by the Scientific Foresight Unit; they say that “this may lead not only to a massive collection of personal data about individuals, to the detriment of privacy, but also to a pervasive influence on their behaviour, to the detriment of both individual autonomy and collective interests” (Sartor, 2020, pg.19). However, Harley’s study suggests that users themselves share the sentiment that data collection is “‘just something you expect with the world’s technology’” (Harley, 2018). The products and views collected of each individual does not seem to be too much of a risk to most users, and many are willing to give such information for better recommendations. But, as the Scientific Foresight Unit mentioned, further collection may be the tipping point of risk for users, especially if the data collected starts to include “personal” data or tracking.

The difficulty with assessing risk for users in recommendation systems is that there is no consensus on what data should and shouldn’t be collected. For instance, some may view a server storing the location of a user for an e-commerce site to be reasonable, while others would only want the user to input the destination address themselves. My STS research will delineate what constitutes a risk for consumers, while also establishing what risks a user faces in recommendation systems. I will study the aspects of data collection within the system to find whether the fears are substantial or if the risks the system poses to the user are like any other modern-day application.

CONCLUSION

The future of the recommendation system market depends on the ability of the engineers to find new advancements of the science and for users to not feel at risk when using the system. Through this project I will create models that account for user activity in a sequential manner, thereby being able to adapt to a multifaceted user base while maintaining long-term patterns. On the STS front, the project will bring clarity to the line between the importance of data privacy and whether or not users face a real risk within the recommender system. By doing so, the project will set the stage for improvements in both technical and social spheres by creating a system that better aligns to user needs and wants, while also respecting their privacy. Establishing such shared understanding between the users and providers should help further these goals.

REFERENCES

Adomavicius, G., Bockstedt, J., Curley, S. P., Zhang, J., & Ransbotham, S. (2018, November 13). The Hidden Side Effects of Recommendation Systems. *MIT Sloan Management Review*. Retrieved from https://sloanreview.mit.edu/article/the-hidden-side-effects-of-recommendation-systems/

Auxier, B., & Rainie, L. (2020, August 17). Key takeaways on Americans' views about privacy, surveillance and data-sharing. *Pew Research Center.* Retrieved from https://www.pewresearch.org/fact-tank/2019/11/15/key-takeaways-on-americans-views-about-privacy-surveillance-and-data-sharing/ [Summary of results from privacy study]

Blue Pi. (2015, November 14). Classifying Different Types of Recommender Systems. Blue Pi. Retrieved from https://www.bluepiit.com/blog/classifying-recommender-systems/ [Article detailing different types of recommender systems]

Fang, H., Zhang, D., Shu, Y., & Guo, G. (2020, January). Deep Learning for Sequential Recommendation: Algorithms, Influential Factors, and Evaluations [PDF File]. ACM Transactions on Information Systems, (1), 1-5. Retrieved https://arxiv.org/pdf/1905.01997.pdf

Gadepally, V. N., Hancock, B. J., Greenfield, K. B., Campbell, J. P., Campbell, W. B., & Reuther, A. I. (2016, November 1). Recommender Systems for the Department of Defense and Intelligence Community [PDF File]. *Lincoln Laboratory Journal*, 22(1), 81-88. Retrieved from https://www.ll.mit.edu/sites/default/files/page/doc/2018-05/22_1_6_Gadepally.pdf

Harley, A. (2018, September 30). Individualized Recommendations: Users' Expectations & Assumptions. Nielsen Norman Group. Retrieved October 12, 2020, from https://www.nngroup.com/articles/recommendation-expectations/

Kumar, P. V., & Reddy, V. R. (2014, August). A Survey on Recommender Systems (RSS) and Its Applications [PDF File]. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(8), 5254-5260. Retrieved from http://www.ijircce.com/upload/2014/august/5_ASurvey.pdf

Low, J., Tan, I. K., & Ting, C. (2019, November). Recent Developments in Recommender Systems. *Multi-disciplinary Trends in Artificial Intelligence,* 1-11. Retrieved from https://www.researchgate.net/publication/337051634_Recent_Developments_in_Recommender_Systems

Madden, M., & Rainie, L. (2019, December 31). Americans' Views About Data Collection and Security. *Pew Research Center*. Retrieved from https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/

Markets, M. (2018, March). Recommendation Engine Market by Type (Collaborative Filtering, Content-Based Filtering, and Hybrid Recommendation), Deployment Mode (Cloud and On-Premises), Technology, Application, End-User, and Region - Global Forecast to 2022. Markets and Markets. Retrieved from https://www.marketsandmarkets.com/Market-Reports/recommendation-engine-market-151385035.html

Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., & Burke, R. (2020, July 25). *Feedback Loop and Bias Amplification in Recommender Systems* (Tech. No. 13019v1). Retrieved https://arxiv.org/pdf/2007.13019.pdf

Milano, S., Taddeo, M. & Floridi, L. (2020, February 27). Recommender systems and their challenges. *AI & Soc*, 35, 957–967. Retrieved October 06, 2020, from https://link.springer.com/article/10.1007/s00146-020-00950-y

Sartor, Giovanni (2020, June). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence [PDF File]. *Scientific Foresight Unit*, 15-79. Retrieved October 14, 2020, from https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

Techlabs, M. (2017, October 02). How Does a Recommendation Engine Really Work? Towards Data Science. Retrieved from https://towardsdatascience.com/how-does-a-recommendation-engine-really-work-656bdf12a5fc

Wang, J., Tang, Q. (2015). Recommender Systems and their Security Concerns [PDF File]. (n.p.), 13-22. Retrieved from https://eprint.iacr.org/2015/1108.pdf