Recommendation Systems

Samantha Baker

Gonzaga University

Recommendation Systems

While scouring the internet as I researched this paper, I've encountered a different advertisement on each successive webpage. First, it was a Madewell cardigan on the side of a New York Times article (I like to be warm). Next, it was Amazon trying to sell me dog treats while I read an article about search engines (I just got a new dog). After that, Godiva was pushing hard for me to notice their Christmas-themed chocolates as I paged through data on algorithmic bias (I like chocolate – who doesn't?). Each time I visited a new webpage, I was presented with an ad that had nothing to with what I was searching for but was tailored perfectly to my interests. The only common thread among this diverse set of webpages was that they had ads served by, dictated by, and recommended by Google. How do these advertisements know exactly what to show me? It's simple, they're called recommender systems.

What are Recommender Systems?

Quickly stated, a recommender system mentions various products, services, and other information based on an analysis of a user's data. In more technical terms, a recommender system "uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a particular item" (Sattar et al., 2017, p. 3229). From the music on Spotify and online store products to TV show recommendations and YouTube videos, it's tough to escape recommendation algorithms while online. According to Chua (2019), "The recommendation can derive from a variety of factors such as the history of the user and the behavior of similar users."

Many companies, such as Amazon, are taking full advantage of recommender systems in order to make a larger profit. If deployed correctly, a recommender system could make a significant impact on increased sales, conversions, click-through-rates, and much more (Techlabs, 2018). For example, Amazon's recommendation algorithm generates about 35% of the company's revenue (Morgan, 2018). According to Morgan (2018), Amazon uses "data from individual customer preferences and purchases, browsing history and items that are related and regularly bought together, Amazon can create a personalized list of products that customers actually want to buy." Moreover, the recommended items are typically shown in a ranked list from which are most suitable to the users based on their preferences. These preferences can either be collected from the ratings a user gave on an item or inferred from the actions of the users (Sattar et al., 2017, p. 3230).

According to Techlabs (2018), recommendation systems are becoming a primary source for users to become exposed to new digital work through their behaviors, interests, and preferences. The author states, "In a world of information density and product overload, a recommendation engine provides an efficient way for companies to provide consumers with personalized information and solutions" (Techlabs, 2018). However, like with most things, there are positives and negatives to these recommender systems. On the one hand, you get personalized products with a click of a button, new items you didn't know you needed, and ideas for future purchases. On the other hand, there is the potential for the abuse of your data.

How Recommender Systems Impact E-Commerce Sales

Recommender systems are typically focused solely on the customer and the consumer experience. Moreover, they impact consumer purchases in many ways. “First, as the source of quality-related information, recommendations reduce the uncertainty of the quality of recommended items” (Pathak et al., 2010, p. 165). Secondly, through advertisement and signaling effects, recommender systems can increase cross-selling opportunities (Pathak et al., 2010). For example, purchases made by other consumers could potentially have a strong signaling effect that could influence purchase decisions (Pathak et al., 2010). Finally, according to Pathak et al. (2010), recommender systems have a greater opportunity to build loyalty with their customers, “Retailers accumulate more and more data about customers and products and can provide more and more accurate recommendations, which makes it less appealing for a customer to switch to another seller due to the difficulty in the transfer of this knowledge,”(p.165). With these in mind, recommender systems show a positive impact on e-commerce sales.

Types of Recommender Filtering

In order to make an effective recommendation system, there are different types of filters a system may have. These systems use different methods and sources to predict user needs using an analysis. The most commonly used recommendation filter systems are collaborative, content-based, hybrid, and demographic (Mohamed et al., 2019). Let’s break down what each of these filtering systems means:

Collaborative: Systems collect and analyze a user’s data, including behaviors such as feedback, ratings, and preferences, in order to provide recommendations (Mohamed et al., 2019, p. 151).

According to Mohamed et al. (2019), they state that with this information, the system “exploits similarities amongst several users or items to predict missing ratings and hence make suitable recommendations. Collaborative filtering methods produce user-specific recommendations of items based on patterns of ratings or usages like purchases without the demand for data about either items or users” (p.151). The authors, Mohamed et al. (2019), uses the following example to explain this process:

> If person P1 likes item 1, item 2, item 3. If person P2 likes item 1, item 3, item 4. If person P3 likes item 1? And so, there is a high chance that person P3 may like item 3 because, from the first two statements we knew that person P1, P2, P3 likes item1 (p.151).

Content-Based: This filtering method is based on the preferences of a user and the description of the item (Mohamed et al., 2019, p. 150). According to Mohamed et al. (2019), “It also recommends items similar to those which a given user has liked in the item rate list. Indeed, the basic operation performed by a content-based system consists of matching the user basic data like age, gender, location, and the rated item list on the site stored in his account with the similar items have a common specification, in order to recommend new items meets his/her interests” (p.150). Additionally, there are process steps, including content analysis, profile learner, and a filtering component.

Hybrid: With a hybrid filtering system, this combines both content-based and collaborative filtering systems. It’s been researched that when combing these filtering methods, it produces better performance and helps alleviate any issues that may occur when different filtering systems are used separately (Khusro et al., 2016, p. 1182). Additionally, this filtering system has a variety of methods, which are called weighted, switching, mixed approach, feature combination, cascade, feature, and meta-level hybridization approach (Khusro et al., 2016, p. 1182).

Demographic: A system collects a user's demographics, which may include their age, gender, location, and even education. The system will then find common users who share these similar qualities and provide recommendations based upon that data (Mohamed et al., 2019, p. 151).

Recommender System Challenges

Like with any Artificial Intelligence system, there comes a large risk of encountering complex challenges. Challenges with recommender systems can range from simple unpredictable items appearing on your Amazon page to the more troubling privacy concerns, such as leaked and shared data. Here are just a handful of challenges recommender systems face:

Cold Start

Recommender systems face a large challenge when a new item or new user steps onto the scene. According to Khusro et al. (2016), "In such cases, neither the taste of the new users can be predicted nor can the new items be rated or purchased by the users leading to less accurate recommendations" (p. 1182). Essentially, it takes time for a recommender system to begin populating recommendations with a new user and new data.

Shilling Attacks

Like with any system, it must be prepared for potential attacks—and recommender systems are no different. It’s possible that a competitor or unhappy user begins giving false ratings and reviews that’ll affect its popularity, whether negatively or positively (Khusro et al., 2016). Khusro et al. (2016) state, “Such attacks can break the trust on the recommender system as well as decrease the performance and quality of recommenders” (p.1183). There are also various attack models, and these attacks can be detected through different approaches (Khusro et al., 2016). It’s critical for a recommender system to be prepared for potential shilling attacks that could cause a disturbance.

Limited Content Analysis

According to Khusro et al. (2016), “Content-based recommenders rely on content about items and users to be processed by information retrieval techniques. The limited availability of content leads to problems” (p. 1184). For example, the authors discuss that many domains have scarce content or content that might be difficult to analyze, e.g. a movie might be trickier to analyze than a page of text (Khusro et al., 2016). In these cases, recommenders are unable to be handle the analyzed content unless it contains enough information that can be used to differentiate the liked and disliked items by the consumer (Khusro et al., 2016).

Ethical Concerns

Privacy is the top concern when it comes to data collection; however, there are additional ethical concerns that recommender systems may face. Due to their popularity, more companies are opting to use recommender systems to keep users engaged. “As the system recommends things that align with the data it has gathered to create a profile of user interests, in reality, the recommendation system domination belies ethical and privacy concerns" (DeLeon, 2019). While a consumer may see their Netflix or Amazon recommendations as a convenience, there are additional ethical concerns present because of the data that is being collected.

Privacy

One of the main concerns about recommender systems is privacy—and for a good reason. Many users and consumers aren't aware of how much information service providers and websites are able to collect, what can be inferred from this data, and where this data ends up in cyberworld. Potential data can be collected or shared without a user's consent or knowledge (Milano et al., 2020). Additionally, once a user's data is stored, there is a potential risk of their information being shared or leaked to external sources (Milano et al., 2020). There's also the issue of how secure this data is when it is collected and stored (Milano et al., 2020). "Privacy breaches can involve a variety of parties (fellow users, the service provider, or outsiders) and maybe a deliberate act (snooping, hacking), or accidental (mismanagement, lingering data)" (Jeckmans et al., 2012, pp. 7-8). A user's personal information is vital when it comes to recommender systems. However, there is a fine line that can easily be crossed when it comes to consent, storage, and the sharing of this collected data. Jeckmans et al. (2012) states,

> Because recommender systems typically contain a large amount of information, often about its users, they form an interesting target for attack. Information could end up in the wrong hands or be misused by legitimate data holders. Given the amount and detail of information within recommender systems, the privacy concerns should be taken seriously (p.9).

Individuals now rely heavily on technology, from emails and text messages to virtual school and online shopping—privacy is more important now than ever before.

Opacity

According to Milano et al. (2020), “In theory, explaining how personalized recommendations are generated for individual users could help to mitigate the risk of encroaching on their autonomy, giving them access to the reasons why the system ‘thinks’ that some options are relevant to them,” (p.962). Additionally, the authors mention that this could potentially help guard against bias by increase the transparency of any algorithm decisions (Milano et al., 2020).

Fairness

Fairness has become a large, complicated issue within algorithms. “If a social network under-ranked posts by a given demographic group, that could limit the group’s visibility on the service” (Beutel et al., 2019, p. 1). According to Beutel et al. (2019), “Part of the challenge in studying fairness in recommender systems is that they are complex” (p.1). The authors continue stating that evaluating recommender systems is incredibly difficult due to their dynamics constantly changing, “What a user was interested in yesterday they may not be interested in tomorrow, and we only know a user’s preferences if we recommend them an item” (Beutel et al., 2019, p. 1).

Social Influence

"A much-discussed effect of some recommender systems is their transformative impact on society" (Milano et al., 2020, p.964). It's no secret that recommender systems are built upon forms of social influence—including potential bribing (Ramos et al., 2020). According to Ramos et al. (2020), due to the influence positive ratings have on a consumer, sellers may bribe their customers in order to change their ratings—which will give the seller a more popular reputation. Ramos et al. (2020) imply that bribing can negatively impact social influence as well as present bias throughout the recommendations.

Lessening Recommender System Issues

Although recommender systems are complex, there are a variety of ways to rectify some of the issues and challenges that are presented. During my research, there was one common issue that had different solutions—that was privacy. From anonymization and privacy-preserving cryptographic protocols to user grouping and laws and regulations (Jeckmans et al., 2012, pp. 10-12), it appears many individuals have their own ideas on how to preserve privacy within a recommender system. According to Jeckmans et al. (2012), who had the ideas of preserving cryptographic protocols, anonymization, and randomization,

None of the research areas mentioned in this section can offer complete user privacy for all recommender systems. Privacy is multi-faceted, as are the domains in which recommender systems are applied. Several areas will likely need to be combined to develop proper privacy-protection techniques for a given application (pp. 14-15).

In terms of recommender systems, I do believe that privacy is an incredibly complex issue and that there is not a concrete way to fully protect a user's privacy. If they did find a solution, it would have to benefit both the consumer as well as the e-commerce site.

With recommender systems having filtering systems and their challenges, I was able to discover that there were a handful of proposed solutions. Below are just a few of these solutions that I had found particularly interesting.

Demographic Clustering

According to Khusro et al. (2016), "Using demographic filtering and clustering, a recommender system may cluster users having similar preferences and demographic features that the system can only look into the appropriate user group rather than the entire dataset" (p. 1186). In addition, the authors state that this could potentially increase performance.

Cold-Start Solution

Recommender systems have a difficult time populating recommendations for new users due to insufficient data. However, one solution is registration. "Contextual information such as location, time, etc. can be obtained through their IP address, and those items are recommended that have been mostly viewed, downloaded and purchased by other users having similar contextual information" (Khusro et al., 2016, p.1187). By collecting the data of a user, the system can connect similarities between the new user and established users to provide recommendations.

Two Lists

A change in a user's preference can make the recommender system's job more difficult. In order to combat this issue, a proposed idea is to have two maintained lists—one new and one old. Khusro et al. (2016) state, "The first one should be maintained according to the current preferences of the user, while the second one should keep track of user's long-term preferences so that the system should recommend items that match user previous transaction history," (p.1187). Having both lists could make it easier for the recommender system to offer new recommendations based on a user's past and present list stored in its system.

Final Thoughts

Many individuals are now shopping online more than ever before—especially with the Covid-19 pandemic. For example, Cyber Monday 2020 sales reached more than $10.8 billion, which, according to Adobe, is the biggest e-commerce day ever (Thomas, 2020). I wonder how many of these sales were from e-commerce recommendations? Online shopping is easy, more affordable, and you don't need to leave the comforts of your couch. However, with the recent surge in online shopping, many consumers don't think about how their Amazon recommendations are populated—as well as how much data the company is truly collecting. It's true, recommender systems are complex and benefit e-commerce companies more than consumers. Nevertheless, recommender systems should be more diligent about protecting a user's privacy—especially with so many people depending on the internet. With these recommendation algorithms taking over e-commerce websites, TV show streaming services, and music apps, it's important that these companies put a focus on ethical and privacy concerns to better protect and benefit their consumers.

References

Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E. H., & Goodrow, C. (2019). Fairness in recommendation ranking through pairwise comparisons. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. https://doi.org/10.1145/3292500.3330745

Chua, R. (2019, June 29). *A simple way to explain the recommendation engine in AI*. Medium. https://medium.com/voice-tech-podcast/a-simple-way-to-explain-the-recommendation-engine-in-ai-d1a609f59d97

DeLeon, H. (2019, April 24). *The ethical and privacy issues of recommendation engines on media platforms*. Medium. https://towardsdatascience.com/the-ethical-and-privacy-issues-of-recommendation-engines-on-media-platforms-9bea7bcb0abc

Jeckmans, A. J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R. L., & Tang, Q. (2012). Privacy in recommender systems. *Computer Communications and Networks*, 263-281. https://doi.org/10.1007/978-1-4471-4555-4_12

Khusro, S., Ali, Z., & Ullah, I. (2016). Recommender systems: Issues, challenges, and research opportunities. *Lecture Notes in Electrical Engineering*, 1179-1189. https://doi.org/10.1007/978-981-10-0557-2_112

MacKenzie, I., Meyer, C., & Noble, S. (2013, October 1). *How retailers can keep up with consumers*. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers#

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & SOCIETY*, *35*(4), 957-967. https://doi.org/10.1007/s00146-020-00950-y

Mohamed, M. H., Khafagy, M. H., & Ibrahim, M. H. (2019). Recommender systems challenges and solutions survey. *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)*. https://doi.org/10.1109/itce.2019.8646645

Morgan, B. (2018, July 16). *How Amazon has reorganized around artificial intelligence and machine learning*. Forbes. https://www.forbes.com/sites/blakemorgan/2018/07/16/how-amazon-has-re-organized-around-artificial-intelligence-and-machine-learning/?sh=d2805c873618

Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems*, *27*(2), 159-188. https://doi.org/10.2753/mis0742-1222270205

Ramos, G., Boratto, L., & Caleiro, C. (2020). On the negative impact of social influence in recommender systems: A study of bribery in collaborative hybrid algorithms. *Information Processing & Management*, *57*(2), 102058. https://doi.org/10.1016/j.ipm.2019.102058

Sattar, A., Ghazanfar, M. A., & Iqbal, M. (2017). Building accurate and practical recommender system algorithms using machine learning classifier and collaborative filtering. *Arabian Journal for Science and Engineering*, *42*(8), 3229-3247. https://doi.org/10.1007/s13369-016-2410-1

Techlabs, M. (2018, December 24). *What are product recommendation engines? And the various versions of them?* Medium. https://towardsdatascience.com/what-are-product-recommendation-engines-and-the-various-versions-of-them-9dcab4ee26d5

Thomas, M. R. (2020, December 1). *Online sales reach $10.8 billion on Cyber Monday, the biggest U.S. e-Commerce day ever, Adobe says*. CNBC. https://www.cnbc.com/2020/12/01/holiday-2020-shoppers-spend-10point8-billion-on-cyber-monday.html