Discuss how data analytics can be used to influence people.

In the era of big data, fuelled by an ever-expanding flow of information where more than 2.5 quintillion bytes are created each day, data analytics has emerged as a key form of influence and plays a central role in monitoring and changing user behaviour at a global scale. Although powerful technologies built to interact with big data are abundant in every field, recommendation systems are seen to be the most influential. As the Internet grows exponentially in size, users are inundated with a massive volume of digital content. To prevent information overload, recommendation systems act as personalised filtering tools, shrinking the previously vast decision domain down to a small number of items by gauging and predicting user interest. By manipulating the content users view, they exert direct control over decision processes across almost every platform. In ecommerce, recommendations cross-sell complementary items to the consumer during check-out which boost average order value significantly, in the case of Amazon contributing towards more than 35% of sales alone [1]. On streaming sites, where attention spans are lost after only 60-90 seconds of searching, recommendation systems continuously push movies, videos and music suggestions to every user in real-time and are crucial to maintaining engagement – for example, they save Netflix $1 billion per year and within 20 years increased influence from 2% to 80% of stream time [1, 2]. Similarly, on social media and news platforms recommendations are responsible for delivering every piece of content on the page. As choices online are driven entirely by recommendations, such systems are omnipresent and dominate the information ecosystem.

The concept of automated recommending has existed since the beginning of computing. In the 1970s, the first personalised recommendation system was created, a computer-based librarian named Grundy which combined hard-coded stereotypes to form a simple user model. Although primitive, it utilised learning mechanisms to evolve its models – for example, it began by suggesting modern books with fast-paced plots to male users, but as many also preferred philosophical and classical novels this was reflected in future predictions. Even at such an early stage, Grundy was already effective in predicting user preferences, recommending books at a 72% success rate [3].

This human-curated method was adopted by many platforms in the early 2000s but being based off only bestseller lists was inevitably biased and failed to include more niche interests. However, the creation of Amazon's recommendation system in 2003 spearheaded a revolution in recommendation technologies which helped lead to their large-scale implementation and ubiquity today. Previous approaches typically used content-based filtering, which compares item attributes with user preference data to determine potential interest. Instead, Amazon's algorithm pioneered the collaborative filtering approach, integrating recommendations from other users with similar interests. Mimicking natural peer-to-peer recommendation (e.g. word of mouth), it discovers users with some items in common then suggests items one user has purchased but others haven't, resulting in higher accuracy without requiring any analysis of item attributes, although context from this information can be used to improve personalisation. For example, algorithms rely heavily on contextualised timing elements, considering whether the purchase of an item signifies short-term (a light bulb) or long-term interest (a badminton racquet), and predicting shifts in interest over time (e.g. buying baby swings shifting to future interest in children's books) [4].

Furthermore, Amazon's recommendation system marked a shift from traditional user-based to item-based collaborative filtering, which evaluates a specific item instead. For example, if a person bought a new computer mouse, the algorithm would create a list of all users who also purchased the mouse and suggest other products they have in common (e.g. a keyboard). Whereas user-based suffers from scarcity of data, as two users are unlikely to have purchased many items in common, item-based performs well with limited user data and only requires a single item to generate high-quality recommendations [5]. This partially alleviates the "cold start" problem, where recommending items to new users is difficult with unknown preferences and no historical data on purchases and ratings. As these algorithms matured, hybrid recommendation systems emerged which allowed the strengths of each approach to be composited. For example, Amazon Personalise combines machine-learning with a "bandit-based" recommendation system which can change the type of suggestions and the algorithm used based on how the user responds in real-time, allowing for more accurate recommendations that account for multiple metrics of satisfaction [6].

It is evident that recommendation systems benefit both users (providing personalised suggestions to aid with the decision making process) and platforms (stimulating demand and increasing revenue). However, there are concerns that by directly manipulating the discovery and consumption of information, these algorithms not only reflect consumer preferences but also shape them, a process that may be gradual but over time and with a large number of users distorts popularity significantly.

Sales diversity, which examines the variety of items bought at the individual level and sold at the market level, is one major area of influence. A study by the University of Pennsylvania found that although individual sales diversity saw a slight increase, as collaborative filtering takes many user preferences into account and exposes consumers to a wider variety of products, aggregate sales diversity fell by over 120% [7]. Items that are initially well-received will be suggested more frequently, resulting in a tendency for algorithms to favour a selection of popular items whilst under-representing the rest of the market. As these popularity differentials compound, concentration bias forms at an aggregate level and narrows consumer preferences.

The potential scale of cumulative advantage was simulated by David Sumpter in the book "Outnumbered", which modelled customers purchasing books through a preferential attachment mechanism where already popular items become more popular. The first customer buys two books at random (representing items with early success on the platform), then in subsequent rounds books by pairs of authors who have been previously bought together are more likely to be chosen. After 500 rounds, success was in most cases reinforced as authors with many connections benefited from the preferential attachment process and authors with few at the start often made lower sales. As this model is entirely probabilistic, multiple runs resulted in different starting books and thus completely different rankings at the end – top authors in one simulation were placed bottom in others. This suggests that recommendation systems create success entirely based off limited initial popularity, in this case with just a one item lead [8].

The main limitation of this model is that it only uses synthetic data generated with simple probability calculations and has no way of reflecting the quality of items, an inherent factor in determining popularity as worse items will struggle to gain early sales in the first place. However, interactions between real consumers and items allow for an evaluation of popularity bias which takes user preferences and quality into consideration. Such significant changes in choice behaviour raise implications that recommendation systems not only manipulate popularity but also perception of quality, as one often acts as a signal of the other. In a study by Salganik, M.J. and Watts, D.J., participants were given the chance to download songs and were assigned to one of two groups – independent with no popularity information at all, and social influence where songs were listed by rank and had download counts displayed (a rudimentary form of recommendation where successful songs are prioritised). Additionally, an "inverted" social influence group was created by reversing the ranking halfway through the experiment (a song in first place would be moved to last and a song in last place to first) which assessed the extent to which false popularity information affected perception of quality. Expectedly, top songs gained significantly more downloads than ones lower in rank, at a much faster rate with popularity information which made successful songs salient and more attractive to users – subjects in the social influence group were about six times more likely to listen to the top song than middle songs (interestingly the bottom song also had three times more listens, as songs at the end of the ranking also stood out to users who clicked out of curiosity). However, after inversion the bottom song immediately began growing faster than the top song which lost most of its download trajectory. Overall, it was found that rankings of almost all songs were permanently affected, with the projected ranks of previously popular songs falling significantly and unpopular songs boosted.

Although in many cases songs began moving back in the direction of their previous ranks, suggesting that the effects of inversion may only be transitory, it is clear that perception of quality (which download count is an indicator of) is easily swayed and over time massively distorted by popularity information. Even in the unchanged group, song rankings did not entirely reflect actual appeal (based on the downloads in the independent group) – the two groups were originally very strongly correlated but began to divert over time as popular songs gained cumulative advantage which inflated downloads above those in independent conditions [9].

Popularity bias is a major unintended consequence of recommendation systems, responsible for manipulating consumer preferences across every domain. However, growing academic and industrial attention has led to the development of many approaches to mitigate its effects, such as increasing the representation of less popular items, as well as raising awareness of impacts on consumer behaviour. As algorithms evolve and reflect preferences with increased accuracy, they will help users more efficiently navigate the continuously growing digital landscape without losing personal choice. Having emerged as incredibly valuable pieces of technology and a key driving force behind algorithmic innovation, recommendation systems will remain the largest source of influence over almost all human activity on the Internet.

References

[1] S. Arora, “Spiceworks,” 16 December 2021. [Online]. Available: https://www.spiceworks.com/marketing/customer-experience/articles/recommendation-engines-how-amazon-and-netflix-are-winning-the-personalization-battle/.

[2] C. A. N. H. Gomez-Uribe, “The netflix recommender system: Algorithms, business value, and innovation.,” *ACM Transactions on Management Information Systems (TMIS),* vol. 6, no. 4, pp. 1-19, 2015.

[3] E. Rich, “User modeling via stereotypes,” in *Cognitive Science*, The University of Texas at Austin, 1979, p. 329–354.

[4] E. Roman, “Data 607 - DISCUSSION/ASSIGNMENT 11: RECOMMENDER SYSTEMS,” 16 November 2022. [Online]. Available: https://rpubs.com/enidroman/966999.

[5] B. S. J. Y. Greg Linden, “Item-to-Item Collaborative Filtering,” EEE Computer Society, 2003.

[6] L. M. Brent Rabowsk, “AWS,” 17 November 2020. [Online]. Available: https://aws.amazon.com/blogs/media/whats-new-in-recommender-systems/.

[7] K. H. Dokyun Lee, “"People who liked this study also liked”: an empirical investigation of the impact of recommender systems on sales volume and diversity.",” 2015.

[8] D. Sumpter, “Outnumbered: Exploring the Algorithms That Control Our Lives,” UK, Bloomsbury, 2018, pp. 87-90.

[9] D. J. W. Matthew J. Salganik, “Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market,” *Social psychology quarterly,* 2008.