
Diversifying Music Recommendations

Houssam Nassif
Kemal Oral Cansizlar
Mitchell Goodman
Amazon, Seattle, USA

HOUSSAMN@AMAZON.COM
KEMAL@AMAZON.COM
MIGOOD@AMAZON.COM

S. V. N. Vishwanathan
Amazon, Palo Alto, USA, and University of California, Santa Cruz, USA

VISHY@AMAZON.COM

Abstract

We compare submodular and Jaccard methods to diversify Amazon Music recommendations. Submodularity significantly improves recommendation quality and user engagement. Unlike the Jaccard method, our submodular approach incorporates item relevance score within its optimization function, and produces a relevant and uniformly diverse set.

1. Motivation

With the rise of digital music streaming and distribution, and with online music stores and streaming stations dominating the industry, automatic music recommendation is becoming an increasingly relevant problem. Various recommender systems have been proposed, including models based on collaborative filtering (Xing et al., 2014), content (van den Oord et al., 2013; Soleymani et al., 2015), context and emotions (Song et al., 2012). Most of those recommender systems focus on improving recommendation accuracy and user preference modeling, in order to produce individually more enjoyable items.

Unlike other digital and physical products, music content tends to have explicit clusters. An album contains multiple songs, all of which share the same album cover graphic, title and description. Furthermore, songs within the same album tend to belong to the same genre, and are usually played back to back. Due to their similar features, recommender systems tend to score same-album songs similarly.

It is common for music recommendations to be rendered in list form, which makes it easy for users to peruse on desktop, mobile or voice command devices. Naively ranking recommended songs by their personalized score results in lower user satisfaction because similar songs get recommended in a row. Duplication leads to stale user experience, and to lost opportunities for music content providers wanting to showcase their content selection breadth. This impact is amplified on devices with limited interaction capabilities. For example, smart phones have a limited screen real estate, and it is usually more onerous to navigate between screens or even scroll down the page (see Figure 1).

In fact, other factors besides accuracy contribute towards recommendation quality. Such factors include diversity, novelty, and serendipity, which complement and often contradict accuracy (Zhang et al., 2012). Since we also deal with user-constructed libraries, we focus on exploring methods to diversify music recommendations.

This work presents experiments to diversify Amazon Music mobile app recommendations. Amazon offers Prime members a free Prime Music benefit, with access to millions of songs and thousands of expert-programmed playlists. Customers can also upload their own music to their library, and mix it with Prime Music content to create personal playlists. Amazon Music developed a recommender system that assigns a personalized score to each music content.

2. Diversity Methods

Similar to cases in visual discovery (Teo et al., 2016), image search (van Leuken et al., 2009), blog posts (El-Arini et al., 2009), and news articles (Ahmed et al., 2012), we apply diversification to alleviate recommendation redundancy. We consider two different diversity methods, one based on Jaccard distance, and the other on submodularity.

2.1. Jaccard Swap Diversity

The Jaccard distance measures dissimilarity between two finite sample sets A and B :

$$d_J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}. \quad (1)$$

For each candidate music recommendation, we use an explanation-based diversification method to generate a set of weighted corresponding explanatory items (Yu et al., 2009). The explanatory items are latent features of the candidate, generated from content and behavioral features. We compute the Jaccard distance between two music recommendations by applying Equation 1 to their underlying explanatory items (Clarkson, 2006). We generate a list of $k = 40$ recommendations using the Algorithm Swap method (Yu et al., 2009), which iteratively maximizes top k pair-wise Jaccard distance, conditioned on score relevance.

2.2. Submodular Diversity

Alternatively, we formulate the selection and ranking of a diverse musical subset as a submodular optimization problem (Fujishige, 2005). Submodular functions are characterized by a diminishing returns property. For set S , subset $A \subseteq S$, elements $x, y \in S$, and submodular function $f : \{0, 1\}^S \rightarrow \mathbb{R}$, we have:

$$f(A \cup \{x\}) - f(A) \geq f(A \cup \{x, y\}) - f(A \cup \{y\}). \quad (2)$$

We divide all musical content into C categories, according to the same content attributes. Each scored candidate gets mapped to multiple categories based on its content and behavioral features. Each category c has its own submodular function f_c . To ensure that a candidate contributes no more than its personalized score, we use:

$$f_c(A) = \log \left(1 + \sum_i score(a_i) \right). \quad (3)$$

We diversify by maximizing the sum ρ of all category functions f_c , which is itself submodular:

$$\operatorname{argmax}_A \left(\rho(A) = \sum_c f_c(A) \right). \quad (4)$$

A near-optimal solution is achieved by an iterative greedy procedure (Nemhauser et al., 1978):

$$A_0 := \emptyset \text{ and } A_{i+1} := A_i \cup \left\{ \operatorname{argmax}_{a \in A \setminus A_i} \rho(A_i \cup \{a\}) \right\}. \quad (5)$$

3. Results and Discussion

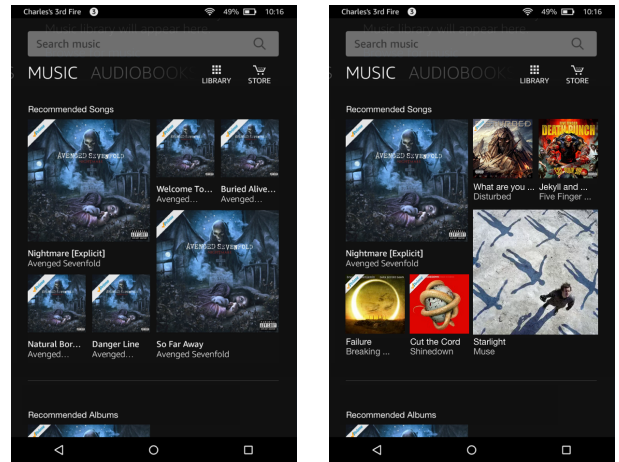
We test the effectiveness of our diversity methods using an online experiment to improve customer engagement with

the Amazon Prime Music app track, album and playlist recommendations (Figure 1). We run a 3-way A/B test on top of the Amazon Music recommender system: baseline, Jaccard Swap diversity, and submodular diversity. Recommender uses item-to-item collaborative filtering and provides item score and explanatory set (Linden et al., 2003). We use the artist and album features as attributes/categories for the Jaccard/submodular methods. Baseline ranks by recommender score and lacks diversity. The experiment lasted 4 weeks, with equal allocation of at least 700,000 customers per treatment. We evaluate the treatment impact on engagement by tracking the number of minutes streamed. We compare the method’s lift (Nassif et al., 2013) via Welch’s t-test.

Table 1. Increase in number of minutes streamed.

Treatment	Jaccard Swap	Submodular
Baseline	0.40% ($p = 0.18$)	0.64% ($p = 0.03$)
Jaccard Swap		0.24% ($p = 0.41$)

Table 1 shows experimental results. Based on minutes streamed, both diversity measures fare better than baseline. This result reinforces the notion that diversity affects recommendation quality (Zhang et al., 2012). Only submodularity’s 0.64% lift improvement is significant (Figure 1).



(a) Baseline recommendation

(b) Submodular diversity

Figure 1. Effect of diversification on Amazon Prime Music mobile app personalized album recommendations.

Our submodular solution outperforms the Jaccard approach. One reason may be that maximizing the submodular function ρ (Equation 4) by iteratively picking the category with the highest gain (Equation 5) produces a uniformly diverse stream, as the diminishing returns property holds for any contiguous part of the recommended list. Jaccard Swap doesn’t guarantee such a smooth list.

Another possible reason is due to Algorithm Swap, which

does not necessarily retain the most relevant content at the head of the list. The swap can sacrifice a highly relevant song in order to increase overall diversity (Yu et al., 2009). The submodular approach ensures that the most relevant song appears first, followed by a mix of the most relevant songs within each category.

References

- Ahmed, A., Teo, C. H., Vishwanathan, S. V. N., and Smola, A. Fair and balanced: Learning to present news stories. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 333–342, 2012.
- Clarkson, K. L. Nearest-neighbor searching and metric space dimensions. In Shakhnarovich, Gregory, Darrell, Trevor, and Indyk, Piotr (eds.), *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*, pp. 15–59. MIT Press, 2006.
- El-Arini, K., Veda, G., Shahaf, D., and Guestrin, C. Turning down the noise in the blogosphere. In *Proceedings of SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 289–298, 2009.
- Fujishige, S. *Submodular functions and optimization*. Elsevier, 2005.
- Linden, G., Smith, B., and York, J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.
- Nassif, H., Kuusisto, F., Burnside, E. S., and Shavlik, J. Uplift modeling with ROC: An SRL case study. In *International Conference on Inductive Logic Programming (ILP)*, pp. 40–45, 2013.
- Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. An analysis of approximations for maximizing submodular set functions i. *Mathematical Programming*, 14(1):265–294, 1978.
- Soleymani, M., Aljanaki, A., Wiering, F., and Veltkamp, R. C. Content-based music recommendation using underlying music preference structure. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2015.
- Song, Y., Dixon, S., and Pearce, M. A survey of music recommendation systems and future perspectives. In *Proceedings of International Symposium on Computer Music Modelling and Retrieval (CMMR)*, pp. 395–410, 2012.
- Teo, C. H., Nassif, H., Hill, D., Srinivasan, S., Goodman, M., Mohan, V., and Vishwanathan, S. V. N. Adaptive, personalized diversity for visual discovery. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*, 2016.
- van den Oord, A., Dieleman, S., and Schrauwen, B. Deep content-based music recommendation. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 2643–2651, 2013.
- van Leuken, R. H., Garcia, L., Olivares, X., and van Zwol, R. Visual diversification of image search results. In *Proceedings of International Conference on World Wide Web (WWW)*, pp. 341–350, 2009.
- Xing, Z., Wang, X., and Wang, Y. Enhancing collaborative filtering music recommendation by balancing exploration and exploitation. In *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, pp. 445–450, 2014.
- Yu, C., Lakshmanan, L., and Amer-Yahia, S. It takes variety to make a world: Diversification in recommender systems. In *Proceedings of the International Conference on Extending Database Technology (EDBT)*, pp. 368–378, 2009.
- Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., and Jambor, T. Auralist: Introducing serendipity into music recommendation. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 13–22, 2012.