Analysis of music tagging and listening patterns: Do tags really function as retrieval aids?

Jared Lorince[1], Kenneth Joseph[2], and Peter M. Todd[1]

[1] Cognitive Science Program
Indiana University, Bloomington, Indiana, USA
`{jlorince,pmtodd}@indiana.edu`
[2] Computation, Organization and Society program
Carnegie Mellon University, Pittsburgh, PA, USA
`kjoseph@cs.cmu.edu`

Abstract. In collaborative tagging systems, it is generally assumed that users assign tags to facilitate retrieval of content at a later time. There is, however, little behavioral evidence that tags actually serve this purpose. Using a large-scale dataset from the social music website Last.fm, we explore here how patterns of music tagging and subsequent listening interact in an effort to determine if there exist measurable signals of tags functioning as retrieval aids. Specifically, we describe several methods for testing if the assignment of a tag tends to lead to an increase in listening behavior. Results indicate . . .

1 Introduction

In social tagging systems, users assign freeform textual labels to digital content (music, photos, web bookmarks, etc.). These individual tagging decisions are aggregated into a folksonomy [4], a "bottom-up" classificatory structure developed with little or no top-down guidance or constraints. There are a variety of reasons for which users tag content, but it is overwhelming assumed that tagging for future retrieval – assigning a tag to an item to facilitate re-finding it at a later time – is users' principal motivator. But is this a valid assumption?

Collaborative tagging systems are often designed, at least in part, as resource management platforms that expressly facilitate the use of tags as retrieval aids, but the freeform, and often social, nature of tagging opens up many other possible reasons for which a user might tag a resource. There is much non-controversial evidence for such alternative tagging motivations, (sharing resources with other users, evaluation, etc.) but the problem with the retrieval aid assumption runs deeper than there simply existing possible alternatives. There is, in fact, almost no behavioral evidence that tags are ever actually used as retrieval aids. While there is much data available user tagging habits (i.e. which terms are applied to which resources, and when), to our knowledge there is no published research providing behavioral evidence of whether or not tags, once applied to items, actually facilitate subsequent retrieval. This is an issue largely driven by a lack of data: While a web service can in principle track users' interaction with tags (for instance, if users use tags as search terms to find tagged content), there are no

available datasets containing such information, nor can it be crawled externally by researchers.

The problem is not intractable, however. While measuring how existing tags are utilized remains beyond our reach, an alternative approach is to examine how patterns of user interaction with tagged versus untagged content vary. In other words, if tags do serve as retrieval aids, we should expect users to be more likely to interact with resources (e.g. visit bookmarked pages, listen to songs, view photos, etc.) upon the application of a tag.

In the current paper we test this hypothesis using a large-scale dataset from the social music website Last.fm using, consisting of complete listening and tagging histories from more than 100,000 users. From this we extract user-artist listening time series (each representing the frequency of listening over time to a particular artist by a particular user), and compare those time series in which the user has tagged the artist, and those that are untagged. Specifically, we address the following two questions:

– RQ1: Does comparison of tagged versus untagged time series provide evidence that tagging an artist increases probability of listening to that artist in the future?
– RQ2: Do certain tags prove to be particularly associated with increases in future listening, and if so, can we identify attributes of such "retrieval-targeted" tags as opposed to others?

We describe the various analytic methods we bring to bear on these questions in Section 4, but first present related work (Section 2) and details of our dataset (Section 3). We close in Section 5 with synthesis and interpretation of our results, as well as a plan for future work.

2   Related work

3   Dataset

Last.fm incorporates two specific features of interest to us here. First, it implements a collaborative tagging system (a "broad" folksonomy, following Vander Wal's [3] terminology, meaning that multiple users tag the same, publicly available content) in which users can label artist, albums, and songs. Second, the service tracks users' listening habits both on the website itself and on media players (e.g. iTunes) via a software plugin. This tracking process is known as "scrobbling", and each timestamped instance of a user listening to a particular song is termed a "scrobble".

Here we utilize an expanded version of a dataset described in earlier work [1, 2] that includes the full tagging histories of approximately 1.9 million Last.fm users, and full listening histories from a subset of those users (approximately 100,000) for a 90-month time window (July 2005 - December 2012, inclusive). Data were collected via a combination of the Last.fm API and direct scraping of publicly available user profile pages. For further details of the crawling process, see [1, 2].

For our current purposes, we consider only those users for which we have both tagging and listening histories. For each user, we extract one time series for each unique artist listened to by that user. Each user-artist listening time series

consists of a given users's monthly listening frequency to a particular artist for each month in our data collection period, represented as a 90-element vector.

The over 2 billion individual scrobbles in our dataset define a total of XXX such time series. In XXX of these cases, the user has assigned at least one tag to the artist within the collection period (we refer to these as tagged time series), while in the remaining cases (XXX) the user has never tagged the artist. We summarize these high level dataset statistics in Table 1.

Total users	104,829
Total scrobbles	2,089,473,214
Unique artists listened	X,XXX,XXX
Unique artists tagged	X,XXX,XXX
Total user-artist listening time series	XX,XXX,XXX
Total tagged time series	X,XXX,XXX
Total untagged time series	88,944,512

Table 1: Dataset summary

4 Analyses & Results

5 Conclusion

References

1. Lorince, J., Todd, P.M.: Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In: Proceedings of the 5th Annual ACM Web Science Conference. pp. 215–224. ACM (2013), `http://dl.acm.org/citation.cfm?id=2464516`
2. Lorince, J., Zorowitz, S., Murdock, J., Todd, P.M.: "Supertagger" behavior in building folksonomies. In: Proceedings of the 6th Annual ACM Web Science Conference. pp. 129–138. ACM (2014), `http://dl.acm.org/citation.cfm?doid=2615569.2615686`
3. Vander Wal, T.: Explaining and Showing Broad and Narrow Folksonomies (2005), `http://www.vanderwal.net/random/entrysel.php?blog=1635`
4. Vander Wal, T.: Folksonomy Coinage and Definition (2007), `http://vanderwal.net/folksonomy.html`