

Pandora's Music Recommender

Michael Howe

1 Introduction to Pandora

One of the great promises of the internet and Web 2.0 is the opportunity to expose people to new types of content. Companies like Amazon and Netflix provide customers with ideas for new items to purchase based on current or previous selections. For instance, someone who rented "Star Wars" at Netflix might be presented with "The Matrix" as another movie to rent. The challenge in this strategy is to make suggestions in a reasonable amount of time that the user will mostly like based on the known list of what the user already enjoys. People will only pay attention to the recommendations of a service a finite number of times before they lose trust. Repeatedly suggest content that the user hates and the user will look for new content elsewhere. Also, a user will only pay attention to a service's recommendation if they arrive in a reasonable amount of time. Make the user wait longer than they are willing and they will again turn elsewhere for content suggestions. Accuracy and speed are critical to a service's success.

Pandora exposes people to music with an online radio station where a user builds up "stations" based on musical interests. The user indicates in each station one or more songs or artists that he or she likes. Based on these preferences Pandora plays similar songs that the user might also like. As Pandora plays the user can further refine the station by giving a "thumbs up" or a "thumbs down" to a particular song. A "thumbs up" means the user likes what he or she hears and wants to hear more music that is similar. A "thumbs down" means the user never wants to hear this particular song again and is not interested in similar types of music. With every bit of this information about the user's interest Pandora hopes to improve the user's trust in their ability to play music the user likes and to make these recommendations in a reasonable time frame. Beyond these goals Pandora also aims to recommend music that the user might have not otherwise heard because it is not well known by a large community of people.

2 Description of Pandora problem space

To make their music recommendations Pandora utilizes a classification system that is the heart of their service. Instead of making musical choices based on the song choices of other users with similar interests, Pandora recommends by matching up the user's artist and song likes with other songs that are similar. To figure out how songs are similar Pandora implemented what they call The Music Genome Project. They believe that songs are comprised of a series of characteristics. Similar to how the human genome describes a person, these characteristics describe a song. The Music Genome Project applies values for each song in each of approximately 400 musical attributes. According to Tim Westergren, the founder of Pandora, "The genome itself is a sort of a very large musical taxonomy. It's a collection of about 400

musical attributes that collectively essentially describe a song, that altogether comprise the basic elements of a song. So it's all the details of melody and harmony, rhythm and instrumentation, and form and vocal performance – a song broken down into all the most basic of parts. It's a bit like musical primary colors in a way."

The greatest challenge for Pandora is classifying songs in their database and building their musical taxonomy. To accomplish this, Pandora employs a team of trained musicians who perform a manual classification on each song before adding it to their database. The musicians spend their workdays listening to a collection of songs and tagging each according to approximately 400 musical attributes. Westergren says, "...we have a team of musicians that sit down and listen to songs one at a time and analyze each song along these attributes. So they literally score each attribute one at a time based on their musical breakdown of the song. So it's a bit like musical DNA. And then once you have this kind of fingerprint for a song, this musical fingerprint which is all done by human analysis..." This classification is slow, tedious and completely manual. A music buyer must keep a backlog of music to classify or else the employees sit idle. Even when working at full capacity each song requires approximately 20 minutes of an employee's time to categorize according to the attributes. Based on an ordinary eight hour workday, an employee can only feed approximately 24 songs into the music database on any given day. The only way to scale the service is to hire more employees, speed up the manual classification, automate the classification or determine a trusted way for more people to contribute attribute ratings. Any of these improvements would still require that the classification quality not diminish. Certainly, this process is the primary bottleneck for the service.

Once songs are properly classified in the database Pandora compares the description of musical tastes of a station selected by an individual user with the classification of the songs in the music database. This comparison returns a collection of songs that drive the playlist. According to Westergren, "...you then can calculate how close two songs are together by comparing the scores on these genes on which they're all described. So we have a mathematical algorithm that takes those numbers and turns it into a proximity measure and that's what we use to generate playlists." This proximity measure is the second key to the Pandora music service. The algorithm must run quickly enough so that the user does not wait too long to hear the next selection. This performance is especially important when the user changes the station or gives a "thumbs down" to a song. In both cases a new song must be chosen based on the user action and the music must be streamed to the user's computer in a reasonable time frame. The next section focuses on proximity measures and collaborative filtering.

3 Investigation of algorithms potentially used by Pandora

The paper "Item-based Collaborative Filtering Recommendation Algorithms" by Sarwar, Karypis, Konstan and Riedl discusses the approaches taken by recommender systems. In general this problem is called collaborative filtering. The goal is to produce a list of items in which a user would have interest. Producing this list can be based on a number of different factors, including buying habits of a user, buying habits or similar users or in the case of

Pandora the musical interests of the current user. Once the data is collected, either implicitly from the actions of a user or explicitly from user ratings (as in the case of Pandora), predictions are made as to what content to deliver. There are a couple techniques typically used for making this prediction.

First, memory-based algorithms take advantage of the full user database to generate a prediction. As indicated in the above mentioned paper, “These systems employ statistical techniques to find a set of users, known as *neighbors*, that have a history of agreeing with the target user (i.e., they either rate different items similarly or they tend to buy similar sets of items). Once a neighborhood of users is formed, these systems use different algorithms to combine the preferences of neighbors to produce a prediction or *top-N* recommendation for the active user. The techniques, also known as *nearest-neighbor* or user-based collaborative filtering are more popular and widely used in practice.” This is the type of algorithm used by a site that takes user data and provides a set of content based on similarity to a user group. They take all the data from all users and seek to intuit likeable content by breaking down this information into groups of like-minded individuals.

Second, model-based collaborative filtering algorithms make predictions after making a model of ratings. They compute a recommendation based on the user’s ratings on other items. According to Sarwar, et al, “The clustering model treats collaborative filtering as a classification problem and works by clustering similar users in same class and estimating the probability that a particular user is in a particular class *C*, and from there computes the conditional probability of ratings.” In this case the ratings given by other users are used to guess what the ratings would be on other similar items. This information is then used to make a prediction.

Pandora differs from other services in that it is not interested in relating the interest of one user to another. Most services heavily rely on user’s opinions to make recommendations for other users. They build up a database of user preferences and seek to relate those preferences to a current user. Pandora, on the other hand, does not utilize information about the number or types of people who like or dislike particular songs to make music selections. Only the musical classification of a song determines whether a station will play that song. Although the general methods differ, the underlying framework is similar between these different service types. Beneath it all the service need to make a decision about what other content is similar to the user’s preferences based on the current classifications. A site like Amazon looks for other users with very similar taste to the current user and then makes recommendation based on algorithmic techniques filtering these people liked. Pandora looks for music whose genome is very similar to the tastes of the user and chooses songs closest to this definition with their own filtering technique. Either way, the service builds a collection of similar content based on a group of near values.

The paper “Application of Dimensionality Reduction in Recommender System –A Case Study” also by Sarwar, Karypis, Konstan and Riedl furthers this discussion of targeting this collection of content. A common way the area of like content is calculated is with nearest-neighbor techniques via the memory-based algorithms. Basically, a neighborhood of content is built based on a proximity measure that defines what is similar. According to Sarwar, et al, “Neighborhoods need not be symmetric. Each user has the best neighborhood for him. Once a neighborhood of users is found, particular products can be evaluated by forming a weighted composite of the neighbors’ opinions of that document.” In other words the key is to build up a

neighborhood for a particular user's preferences. It does not matter if this neighborhood works exactly for other users, only for this particular instance. In contrast to a site like Amazon whose neighborhood is made of people and their interests, Pandora's neighborhood would consist of a group of songs.

The key to Pandora's method of recommending music from its database is to utilize an efficient and effective proximity measure algorithm to determine the neighborhood of music to play on a station. This is not a particularly difficult problem as the set of songs in the database is fairly static, only growing daily on the order of hundreds of songs (based on the number of musicians classifying). All that is required is math that combines a station's preferences into a single value and then compares this value with the values of the songs in the database (that are pre-defined). The algorithm that accomplishes this task occurs in polynomial time. The time required upon initially adding a song to the database is $O(n)$, where n is the number of attributes characterizing the genome. Adding a song to the database only occurs once per song and is not time-critical, so this time is fairly unimportant. The more important factor is the time it takes to choose the next song to play from the current station. This can be done in $O(mn)$ time, where m are the number of songs/artists in the current station and n is the number of attributes characterizing the genome, to create the value for the station plus a fixed time to query the database for the closest matches to this value. Pandora could increase the order by doing a more complicated database query. Regardless, the order could stay polynomial. A quick turnaround time is important at this stage since content streaming will certainly be time-consuming and a large song retrieval time could be user prohibitive.

5 Open issues and challenges regarding the problem

Overall, the method of recommendations employed by Pandora appears to be successful. Based on personal experience, anecdotal evidence from message boards and adoption by larger scale services like msn.com, the recommendations seem to resonate with users. Unfortunately, there is no precise and measureable manner of determining whether a recommendation actually resonates with a particular user. Only the overall success of the company in attracting new users based on word of mouth sharing will prove success. However, the overall signs seem to indicate that their method is attracting loyal customers.

Really, the major issue with the method employed by Pandora is scalability. Once they get music into their database, their algorithm is efficient and effective. However, the big issue is getting the songs into the database. Although they currently have over half a million songs in their collection, this is a small subset of the total music in existence. This is even a small subset of the music they classify as "good enough" (the company explicitly does not want to classify all music; just the music they feel is worthy). In addition to not being able to classify all songs in existence, they are unable to keep pace with the creation of new music. Being that part of their mission is to provide an outlet for the discovery of lesser known music, solving the issue of scalability seems rather important. As mentioned above, the primary ways to provide increased scalability are to hire more employees, speed up the manual classification, automate the classification or determine a trusted way for more people to contribute attribute ratings. Hiring more employees or speeding up the manual classification does not fully solve the

scalability problem. There is still a practical bound on the order of hundreds or thousands of songs that could be classified per day. Automating the classification seems like a good idea, but is really dependent on the growth of artificial intelligence. Teaching a computer to process all the nuances of musical classification is a problem that will most likely not be adequately solved soon.

The most promising scalability strategy seems to be allowing users to contribute classifications in line with a more collaborative filtering strategy. There are many people in the internet world that would be qualified to classify music. Not all these individuals would be willing to participate, but based on the amount of time people spend classifying other content online, some of them probably would be. Providing these users the ability to classify a song is a simple technology task. The issue is more related to trust. How can Pandora maintain a high-level of confidence in their classifications when they don't know the qualifications of the individual who provided a classification? This is not an easy problem to solve. One thought is for Pandora to publish a list of songs for general users to attempt classification. After an individual user has classified a certain number of songs, compare their classifications with those of the employee. If the results fall within a certain threshold of similarity, allow the user to become a classifier. Give the user an incentive such as free membership for classifying a certain number of songs in a given time period. Important to this plan is building in some type of automated verification system. For instance, part of the list given to the user to classify could include random songs already classified by an employee. Require that a user only maintains classifier status if the comparison stays within the required threshold over a representative sample. To further this validation Pandora could track the number of thumbs up and thumbs down given to songs classified by the user. If this ratio falls outside a certain range, review the classification of a sample of the user's songs. Obviously, these are general ideas that would require an initial and ongoing investment by Pandora. However, the net volume of songs that could be classified would potentially increase dramatically and aid in one of the company's biggest problems: scalability.

Despite any issues, the techniques utilized by Pandora to recommend music have been very successful. Their classification system provides a powerful way to show similarities between many types of music and allows individuals to easily discover new songs with little effort. Their primary challenge moving forward will be to maintain their quality while scaling their music library.

References

1. Sarwar, B. M., Karypis, G., Konstan, J., and Riedl, J. "Application of Dimensionality Reduction in Recommender System –A Case Study." In *ACM WebKDD 2000 Web Mining for E-Commerce Workshop*.
2. Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. "Item-based Collaborative Filtering Recommender Algorithms. Accepted for publication at *the WWW10 Conference*. May, 2001.

3. Pekalska, E., Duin, R. P. W., "Learning with general proximity measures", In Journal of Machine Learning Research, (2004).
4. Creative Generalist Q&A: Tim Westergren. Available: <http://creativegeneralist.blogspot.com/2006/10/creative-generalist-ga-tim-westergren.html> (Accessed 5 March 2007).
5. Algorithm and Blues: How Pandora's matching service cuts the chaos of digital music. Available: <http://www.fastcompany.com/magazine/101/pandora.html> (Accessed 5 March 2007).
6. Inside the Net 6: Tim Westergren of Pandora Media (radio interview). Available: <http://www.twit.tv/itn6> (Accessed 5 March 2007).