8-2008

Applications of Voting Theory to Information Mashups

Alfredo Alba

Varun Bhagwan

Julia Grace

Daniel Gruhl

Kevin Haas

See next page for additional authors

Alfredo Alba, Varun Bhagwan, Julia Grace, Daniel Gruhl, Kevin Haas, Meenakshi Nagarajan, Jan Pieper, Christine Robson, and Nachiketa Sahoo

Applications of Voting Theory to Information Mashups

Alfredo Alba* Varun Bhagwan* Julia Grace* Daniel Gruhl* Kevin Haas$^\bullet$
Meenakshi Nagarajan$^\circ$ Jan Pieper* Christine Robson* Nachiketa Sahoo$^+$

{ *alba**, *bhagwan**, *jhgrace**, *dgruhl**, *khaas*$^\bullet$, *nagarajan.5*$^\circ$ *jhpieper**, *crobson**, *nsahoo*$^+$ }
@ { **us.ibm.com*, $^\bullet$*yahoo-inc.com*, $^\circ$*wright.edu*, $^+$*cmu.edu*}

*IBM Almaden Research
650 Harry Road
San Jose, CA

$^\bullet$Yahoo! Inc.
2811 Mission College
Santa Clara, CA

$^\circ$Wright State University
Dayton, Ohio

$^+$Carnegie Mellon University
Pittsburgh, PA

Abstract

Blogs, discussion forums and social networking sites are an excellent source for people's opinions on a wide range of topics. We examine the application of voting theory to "Information Mashups" – the combining and summarizing of data from the multitude of often-conflicting sources. This paper presents an information mashup in the music domain: a Top 10 artist chart based on user comments and listening behavior from several Web communities.

We consider different voting systems as algorithms to combine opinions from multiple sources and evaluate their effectiveness using social welfare functions. Different voting schemes are found to work better in some applications than others. We observe a tradeoff between broad popularity of established artists versus emerging superstars that may only be popular in one community. Overall, we find that voting theory provides a solid foundation for information mashups in this domain.

1 Introduction & Motivation

Online forums are providing an ever increasing opportunity for self expression and exploration. For any given phenomena – interest in music, for example – there are many places on the web where opinions are expressed. In the music domain this includes sites where users comment on artists and songs, listen to music, view videos, enjoy remakes, covers and parodies, and so forth. These actions are all motivated by the same underlying objective: for music fans to become more engaged in, and active contributors to, the communities surrounding their favorite artists. It seems natural to combine data gathered across different sources to form a unified, focused view of community interests.

This is harder than it might at first appear, as the virtual footprint one leaves in each source differs. One manifestation of user contribution might be a posting to a discussion board, while in other communities the social presence of users is solely reflected in "view counts".

Some sites have large numbers of subscribers who are infrequently on-line, while others cater to very active audiences (often in the teenage demographic). Some sites are heavily prone to spamming and other manipulation.

Bringing all of this information together has the feel of a "Mash-Up," i.e. things that were not designed to work together combined for a new purpose. In doing so, we must deal with the differences between sites, the "noise" in various sites, wide discrepancies in site populations and demographics, business concerns about site data providers, etc. We will consider each in turn to motivate our approach.

Modality: One question we face when looking to combine data from multiple sources is how relevant each source is as an indicator of community interest. For example, does listening to music on a radio stream indicate the same degree of interest as posting an in-depth discussion of an artist on a discussion board? We have identified two major dimensions of modality - intentional verses unintentional and consuming (passive) verses producing (creative).

Intentional activities are those where we have a fair degree of confidence that the user has had to take steps to "make their mark". Examples would be navigating to a particular page, or typing a bands name into a search bar. This contrasts with unintentional activities where a song is played on the radio stream, or automatically when the user navigates to a site (for example, many MySpace pages have music playing by default). In general, we assume that intentional activities are a stronger indication of interest than unintentional.

Creative, producing activities are those where the user takes the time to author a post, compose a response, post their own cover of a song, etc. Passive, consuming activities involve the user listening to, watching or reading something created by someone else. Again, we assume that creative activities indicate more interest, if only because they require time and attention.

If we look at the sources we examine for this paper they break down as follows: *Intentional Creative*: MySpace posts, Bebo posts; *Intentional Passive*: Youtube views; *Unintentional Passive*: LastFM listens. Obviously unintentional creative activities are hard to imagine.

Popularity and Population: One challenge of using Internet sites to gauge interest is with disproportionate populations and posting frequencies of different sites. The fraction of users posting on a particular site is often a small percentage of the total advertised users. For example, looking at 721 unique artists on MySpace, a discussion forum boasting over 250 million accounts, only 90,903 unique posters contributed in a 3-day window around February 14th. In a wider window of 3 months, we saw no more then 1.7 million unique posters. Each site we use has a different activity level of posting. The result is that raw counts, even within similar modalities, can vary by factors of 10 or more.

Organized Posting: Spikes in post frequency are often seen when a band asks their fan base to post about the new album, or other event. This has the effect of temporarily driving the band's apparent popularity up in the site's charts during that small time window, an effect we have observed many times during this study. This type of manipulation is not new to the online world. For example, it is suspected that the New York Times best seller list has on occasion been intentionally influenced by massive book buying campaigns (e.g. [16]).

Some voting systems can lessen the impact of point manipulation in particular sites. Drawing from different modalities where different user groups are represented and using fair voting methods diminishes the impact of point sources while preserving popular interest in a fair and unbiased manner.

Business Realities: User-generated content typically reflects the demographic of the community using the website. Some websites may be visited more heavily among the teen and tween groups, causing a bias toward younger audiences when reporting topical popularity. Websites which represent international audiences tend to portray different preferences than local websites. User behavior may vary based on the community and the usability of the website. As a result, top-N lists can vary wildly across the various sites, yet many of the sites pride themselves on the accuracy and quality of results displayed to their target audience.

Secondly, we have found that the "noise" on data from many sites can be substantial, often due to caching effects, update times, etc. Fortunately, one thing we have found to be more consistent is the total ordering of artists, even when the relative ranking of the artists is not well preserved. Systemic errors tend to hit all artists equally and misorderings tend to be more blatant and, thus, are caught and remedied sooner.

Finally, many sites have advertising business models that are based on absolute volume of views, searches, downloads, etc. As these metrics are connected to revenue, it is understandable that they are not often publicly available. To integrate information from such sites it may be necessary to pursue alternative methods such as percentages or ranked lists of popularity.

1.1 Contributions

In the remainder of the paper we will examine prior work in the area of combining results (Section 2), then outline our use of voting methods (Section 3) to create a very powerful framework for Information Mashups (Section 4) which can effectively deal with the merging of dissimilar modalities from very different sources. We perform experiments (Section 5) with eight popular voting schemes operating over four information sources covering three modalities and variants on two popular social welfare functions to evaluate the effectiveness of each scheme. We conclude (Section 6) with some observations on the success of this approach, and finally acknowledge (in Section 7) our partners without whom this research would not have happened.

2 Background

The process of gauging a community's collective interest in music can be described in two steps. First, the data sources need to be prepared for integration using metadata and information integration techniques[11]. Second, the sources need to be combined in a way that takes into account the aforementioned modality, interest and population differences. In order to combine the data sources we will apply voting theory.

Metadata Extraction for Integration: Preparing heterogeneous data sources for integration requires resolving heterogeneities between their elements [19]. Often, this is attempted at the feature low level, for instance combining 'address' elements between Google Maps and Craigslist for a housing MashUp. However, when using unstructured data sources, the first step of Metadata Extraction needs to be performed before attempting integration. Our work in extracting metadata from unstructured user posts draws from two domains of literature – *Information Extraction* (IE) and *Natural Language Mining (NLM)*.

IE exploits the presence of structured content in the gathered data. Metadata is extracted by using manually coded or automatically induced wrappers [14, 13] for the specific data source. In our work, examples of such metadata include user's demographic information, post timestamps, etc. NLM analyzes the completely unstructured part of the

Rank	Bebo	LastFM	MySpace	YouTube
1^{st}	Rihanna	Red Hot Chili Peppers	Jeffree Star	Rihanna
2^{nd}	Paramore	The Beatles	My Chem. Romance	Alicia Keys
3^{rd}	50 Cent	Radiohead	Alicia Keys	Britney Spears
4^{th}	Cascada	Coldplay	Miley Cyrus	Avril Lavigne
5^{th}	Chris Brown	Smashing Pumpkins	Amy Winehouse	Backstreet Boys
6^{th}	My Chem. Romance	Bloc Party	Simple Plan	My Chem. Romance
7^{th}	Justin Timberlake	System of a Down	Britney Spears	Foo Fighters
8^{th}	Fall Out Boy	Bob Dylan	Avril Lavigne	Three Days Grace
9^{th}	Arctic Monkeys	Jack Johnson	Bow Wow	Kanye West
10^{th}	Elliot Minor	Green Day	Eminem	Linkin Park

Table 1. Top 10 Artists on 02/14/2008 by Source

content to extract information nuggets [12, 20]. In our work, we process user posts from MySpace and Bebo by employing a series of UIMA [11] annotators driven off of basic entity spotting. Examples of metadata extracted from every user post include artist and track.

Voting Systems We employ methods from Voting Theory to combine preferences observed in the disparate data sources. Many voting systems have been proposed since the field's inception in the 18th century [18], especially since the Nobel prize winning observation by Arrow [2] that a perfect voting scheme cannot exist. The modern method of objectively comparing the effectiveness of voting systems was established by Bergson in 1938 [5], and employs a metric called a Social Welfare Function. This method was recently re-evaluated and defined by Balinski and Laraki as a more general Social Decision Function [3]. The basic methodology is to define a mathematical criteria for the success of a voting system based on a list of characteristics which are desired (e.g. if the majority of voters prefer a single candidate to all others, that candidate is elected).

Our problem can be framed as a multiple-winner voting system where we ask - 'What artists should fill the top-N slots of our ranked list of popular artists?'. The challenge is apparent when we observe the unmistakable disparity between the top-10 lists of each of our four sources (See Table 1). We employ two major voting schemes – the majoritarian and positional schemes to combine rankings in a fair and effective manner

Two of the most popular majoritarian voting systems are the simple majority and plurality system [17]. While simple majority awards the choice with one more than 50% of the votes (or suitable thresholds), a plurality system selects a winner that attains the maximum number of votes without having to surpass the threshold. We use plurality voting extended to the multiple-winner scenario by counting and sorting votes to select top-N winners while also normalizing for source population sizes.

While majoritarian methods use information from binary comparisons between choices, positional methods take information about a source's preference orderings into account. These systems allow each voter to rank the candidates in order of preference. In our case, each source provides a ranking that we use to generate a combined preference ordering. We use variations of the popular Borda Count [7] that is known to generate a complete, transitive social ordering supported by a broad consensus, rather than the choice that is favored by a majority.

Rank Aggregation: Voting theory has been well investigated for creating a ranked list of alternatives from different online sources. This is the notion of rank aggregation, which was introduced to improve search applications on the Web and to combat "spam" [9]. Rank aggregation for information retrieval is typically measured using a weighted harmonic mean of precision and recall called an F-measure [15]. Thus the social welfare function for rank aggregation in information retrieval is precision and recall. Music popularity has a very different social welfare function, wherein each site's opinion should matter and be counted towards the total ranking. Rank aggregation has also been applied to database applications for combining similar results [10]. Again, the social welfare function in this approach is different, so these rank aggregation methods cannot be employed for music popularity. This leads us to explore different voting algorithms.

A recent study of rank aggregation measured the quality of different aggregation methods based on the effect that an individual source can have on the outcome [1]. This study inspires one of the social welfare functions we employ to study voting systems.

3 Approach

We have framed the problem of combining multiple ranks into that of a number of "voters" who wish to "elect" a ranked panel. The evaluation of success is performed by a social welfare function (SWF) which takes lists of all the voters' preferences, along with the outcome of the vote, and produces a "score" satisfactory the outcome is.

Terminology: We (s)elect the most popular artists $a \in A$, given multiple, disagreeing sources $s \in S$. For each artist, we calculate the combined evidence e_a. Such evidence might be number of views, number of listens, number of positive posts, etc. For each source s we use e_a to create a partial ordering (i.e. an ordering where ties are permitted) top n list, T_s. The number of votes for a given artist a from a source s is denoted as v_{as}. This may differ from e_a in cases where some evidence is considered more or less important. The total number of votes V_s within a source s is the sum of the votes for all artists in s: $V_s = \sum_a^A v_{as}$. We also consider the rank r_{as}, which is the position of an artist a within source s. The total number of ranks in a source s is R_s, which is defined as $max_a(r_{as}), a \in A$, the lowest assigned rank. Note that there may be more artists than ranks, because multiple artists may share a rank position.

Majoritarian Schemes

These schemes look at combinations of v_{as} to determine the total ordered list of artists.

Total Votes: The simplest approach to creating a combined ranked list is the summation of comments, views and listens from all sources. We determine the combined evidence e_a for artist a by adding the votes of all sources for this artist: $e_a = \sum_s^S v_{as}$

Though natural, this approach has several shortcomings. It is sensitive to large "miscounts" and other errors from sources, and it tends to result in the very largest sources dominating the chart.

Weighted Votes: Different modalities indicate different levels of effort by the users. It may therefore be reasonable to multiply the counts by a weight $w(s)$. The function $w(s)$ may be determined by the modality of source s or other factors such as population: $e_a = \sum_s^S v_{as}w(s)$.

It is not easy to determine the "correct" weights for each modality or source. Moreover, these weights may need to change over time, e.g., when user behavior changes or when a source increases its user population.

Semi-Proportional Methods: The two previous approaches are sensitive to differences in user populations. In general, larger communities create more votes for a given artist. If the goal is to determine a ranking that will appeal to all communities, some way of boosting smaller communities' voices is needed. One approach is to normalize the votes v_{as} for artist a from source s by the total number of counted votes from that source V_s. $e_a = \sum_s^S \frac{v_{as}}{V_s}$

This gives equal weight to each source, as well as equal weight to the different modalities. However, this may give too much power to smaller communities, and especially heavily weighted sources that only mention a few artists.

Delegates: Delegates combine semi-proportional voting with a (manual) weight for each source, and have many of the same problems as weighted votes. Sources are allotted delegates, often based on user population size. Delegate numbers can be controlled, e.g., by setting minimum and maximum thresholds, therefore limiting the influence of large communities and boosting the influence of smaller ones directly. Here, the proportional amount of votes is multiplied by the number of delegates for the source d_s: $e_a = \sum_s^S \frac{v_{as}}{V_s} d_s$.

Positional Schemes

The above voting schemes are based on the total number of votes cast for each artist. A different class of voting methods is based on rank rather than proportional vote count. Within each source, the rank (1 = best) can be determined by simply sorting artists by number of votes. These ranked lists are then combined from different sources.

Simple Rank: Known as a Borda count, a "score" is assigned to each candidate based on its rank, with the lowest possible rank assigned to missing entries (usually 0). Thus the lowest rank, $\bar{r}_{as}$, is calculated as $\bar{r}_{as} = R - r_{as}$ for each source s, where $R = max_s^S(R_s)$. The evidence then becomes: $e_a = \sum_s^S R - r_{as}$.

Inverted Rank: The use of simple Borda count has the effect that all rank positions are equidistant. We often desire to reward higher ranks more and lessen the impact of one bad rank. This can be accomplished by using the Nauru method of inverted rank: $e_a = \sum_s^S \frac{1}{r_{as}}$.

The appeal of this formula is its simplicity, while avoiding many of the problems described for the other ranking methods. Here, missing artists within a source are just left out (basically contributing 0). The distances between ranks are stable against lowest rank R_s.

Other Methods

There are a number of other methods that have been popular as voting systems, and which we consider for combining rankings.

Run Off: From the top, select artists one at a time from each source in a fixed order. Once the same artist has been selected by at least 50% of sources it is added to the elected list and further mentions of it are ignored. This repeats on unselected artists to fill the remainder of the list.

Round Robin: Select some order between sources, e.g. by population, and then select artists in a round-robin fashion. If an artist has been selected previously, then move to the next one from that source.

4 Engineering Our MashUp

The engineering process of building a MashUp can be broken down into the following four steps. We visit each of these in the context of the work presented in this paper.

Data Capture: Obtaining reliable volumes of data is not a trivial task. For assessing mass popularity and interest in music, one of the main problems is that polling large samples of people directly is problematic and expensive. This problem gets compounded when the polling needs to be done in a repeatable and robust fashion, which is necessary to create a consistent data model that can be tracked over time. Challenges in polling are well known [4].

We thus look to the wealth of information latent in online music communities. Using Internet crawls, public API calls, RSS feeds, etc., we gather user comments on music artist pages from MySpace and Bebo, viewcounts of music videos from YouTube, and audio listens on music tracks from LastFM. Table 2 shows how these sources range across modalities, the total number of user interactions we gathered per site, and the total number of music artists on each site. We gather comments from MySpace and Bebo with a crawler, and content from YouTube and LastFM using the public APIs of those sites.

Data Cleansing: This is a multi-step process in our system. We clean user posts using filter-phrases to purge profane content. Crawled artist profiles are cleaned by discarding spam band profiles.

Metadata Generation: Applications process the gathered data to obtain relevant metadata. We use wrappers to extract structured information from user posts. Some of these include user's demographic information, post timestamps, number of listens, etc. We also analyze the free form text, specifically user comments to derive information nuggets, e.g., spam, sentiments and artist/track related mentions. Using simple arbitrary window-based entity spotting techniques backed by domain dictionaries which have been successfully employed in similar applications [12, 20], we implement a chain of UIMA annotators to extract the metadata.

Integration and Normalized Representation: We use MusicBrainz and FreeDB sources to map different versions of artists and track names from the various sources to a standard form. We use a data hypercube (also called an OLAP cube [6]) stored in a DB2 database to explore the relative importance of various contributing dimensions. The dimensions of the cube are generated in two ways: from the structured data in the posting and listens (e.g., age, gender of the user commenting, number of track plays), and from the measurements generated by the annotator methods (e.g., number of positive, negative, spam comments). We use simple projection of this cube to generate the top-N list for each source which is then fed into the voting algorithm.

5 Experiment

Our experimental evaluation of the different ranking methods tests each of the aforementioned voting systems. We employ two social welfare functions (SWF) to evaluate the effectiveness of the method. As discussed previously, a SWF is defined as a mathematical criteria for the success of a voting system based on some desired characteristics.

Source	Modality	Total Count	Artists
Bebo	Comments	292136	377
LastFM	Passive Listens	1644898	398
MySpace	Comments	41282	771
YouTube	Active Views	17098279	331

Table 2. Sources and Counts for 02/14/2008.

We calculate the SWF as a score where points are awarded for increased social welfare of a ranking system. This allows us to quantitatively measure the "happiness" of each contributing site with the overall ranking.

Precision Optimal Aggregation: Our first SWF is based on the Precision Optimal Aggregation method introduced by [1], which measures how many artists from each source's top-n list made it into the overall top-n list. In our implementation, for each artist in the top-10 list on any single site's rating, we award 2 points if that artist appears in the overall ranking as well, up to a total of 10 points. This equates to a desire that any given site have at least half of its top-10 list represented in the overall top-10 list. Thus, our Precision Optimal Aggregation SWF is defined as $P_{swf} = \sum_S \min(2 * |T_s \cap T|, 10)$, for top-10 lists T_s for each source and top-10 list T overall.

Spearman Footrule: For our second SWF, we employ the Spearman Footrule distance[8]. This SWF emphasizes the preservation of position in the rankings. In our implementation, for each artist in a the top-10 list on any single site's rating, we award points for how close to the same position that artist appears in the overall ranking. This results in the definition of our Spearman Footrule Distance SWF as $S_{swf} = \sum_S \sum_{a=1}^{10} \max(10 - |r_a - r_{as}|, 0)$. Spearman has been proven a good approximation of a related SWF, Kendall tau distance. We employ the Spearman Footrule as a SWF because it is appreciably less computationally intensive (minutes vs. days), unsurprising as computing Kendall tau is NP-hard [9].

Our analysis focuses on the period around the 50^{th} Grammy Awards ceremony – a time period with a fair amount of commentary on music, which was held on February 10th 2008. Table 1 shows the top ten artists for each of the sources as of 2008/02/14.

Method Evaluation

We will now look at each of the eight voting methods to see how they merge these four top-10 lists from Table 1 into one combined top-10. For each method, performance is defined as the efficacy of the technique in maximizing our two Social Welfare Functions.

Rank by Total Votes	P_{swf}	S_{swf}	C_s
Rihanna			
Alicia Keys			
Britney Spears			
Avril Lavigne			
My Chemical Romance			
Backstreet Boys			
Foo Fighters			
Three Days Grace			
Kanye West			
Linkin Park			
Total	22	149	

Table 3. Top 10 using Total Votes

Rank by Weighted Votes	P_{swf}	S_{swf}	C_s
Rihanna			
Alicia Keys			
Britney Spears			
Paramore			
Avril Lavigne			
My Chemical Romance			
Kanye West			
Foo Fighters			
50 Cent			
Cascada			
Total	28	153	

Table 4. Top 10 using Weighted Votes

Rank by Proportions	P_{swf}	S_{swf}	C_s
Rihanna			
Alicia Keys			
Jeffree Star			
Britney Spears			
My Chemical Romance			
Avril Lavigne			
Paramore			
Foo Fighters			
Red Hot Chili Peppers			
Fall Out Boy			
Total	30	146	

Table 5. Top 10 using Semi-Proportional

Rank by Delegates	P_{swf}	S_{swf}	C_s
Rihanna			
Jeffree Star			
Alicia Keys			
Britney Spears			
My Chemical Romance			
Avril Lavigne			
Amy Winehouse			
Linkin Park			
Paramore			
Simple Plan			
Total	26	151	

Table 6. Top 10 using Delegates

Total Votes: Table 3 shows the top-10 produced by merging the sources through simple summation of the votes for each artist. The rest of the columns are color coded, in order: Green=Bebo, Orange=LastFM, Blue=MySpace, Red=YouTube.

The P_{swf} column shows the contribution of each artist to the overall Precision Optimal Aggregation SWF for that source. S_{swf} does the same for the Spearman Footrule SWF. The bars represent how "happy" each source is with the artist being ranked at this position. The graphs in C_S express the contribution to the combined ranking for the artist from each source. In this case (the number of votes), is clearly dominated by YouTube, because YouTube has significantly more data points (views) than the other sources.

The bottom of the table shows the total SWF scores for P_{swf} and S_{swf}, expressed as the raw score. In P_{swf}, each source can contribute up to 10 points, for a maximum score of 40 (best). For S_{swf}, each source can contribute up to 100 points, for a theoretical maximum of 400. We also include the total influence each source had on the top-10 list. As can be seen in C_s, YouTube clearly dominates.

Weighted Votes: Next, Table 4 shows the impact of assigning weights to the different sources based on modalities. In this experiment, we used a modality-based multiplier of 500 for comments (Bebo and MySpace), 10 for active click-views (YouTube) and 1 for passive listens (LastFM). Overall, the list is still being dominated by the absolute number of YouTube views, while Bebo is able to promote some of its artists (Paramore, 50 Cent, Cascada). This dual influence can be seen in the totals. Also, slight improvement in the two SWFs can be seen as well. One could start tweaking the weights until a more even mix between sources is reached. Our experiments suggest that this is a painstaking process and the weights do not remain fixed over time.

Semi-Proportional Methods: The next method calculates the proportional support for each artists within each source and then combines these proportions by summing across sources. We first note that Table 5 contains candidates from all sources. However, this method prefers sources that have few strong favorites (e.g., Bebo). While P_{swf} continues to improve, S_{swf} is actually worse than the Total Votes scheme.

Rank by Borda	P_{swf}	S_{swf}	C_s
Fall Out Boy			
Coldplay			
Metallica			
Marilyn Manson			
My Chemical Romance			
Alicia Keys			
Green Day			
Red Hot Chili Peppers			
Amy Winehouse			
Lily Allen			
Total	20	74	

Table 7. Top 10 using Borda Count

Rank by Inverted Rank	P_{swf}	S_{swf}	C_s
Rihanna			
Red Hot Chili Peppers			
Jeffree Star			
Alicia Keys			
My Chemical Romance			
Paramore			
The Beatles			
Britney Spears			
Avril Lavigne			
Radiohead			
Total	32	127	

Table 8. Top 10 using Nauru Rank

Rank by Runoff	P_{swf}	S_{swf}	C_s
Rihanna			
Alicia Keys			
My Chemical Romance			
Britney Spears			
Avril Lavigne			
Amy Winehouse			
Linkin Park			
Kanye West			
Justin Timberlake			
The Killers			
Total	26	198	

Table 9. Top 10 using Runoff

Rank by Round Robin	P_{swf}	S_{swf}	C_s
Rihanna			
Jeffree Star			
Red Hot Chili Peppers			
Paramore			
Alicia Keys			
My Chemical Romance			
The Beatles			
50 Cent			
Britney Spears			
Miley Cyrus			
Total	30	123	

Table 10. Top 10 using Round Robin

Delegates: We can modify the weight of each source by assigning different numbers of delegates, for example based on user population. In this experiment, we use the following delegate numbers based roughly on population: Bebo: 300, LastFM: 500, MySpace: 1000, and YouTube: 500. When comparing Table 6 with the previous one, the greater influence of MySpace in this scheme becomes apparent. This produces SWFs that are nearly as good as Weighted Ranking, not surprising since they are similar techniques.

Borda Count: Table 7 shows the result of using the standard Borda count method. It gives each source roughly the same influence, which allows LastFM to promote more of its favorites. However this equal weight comes at a price – it has very poor SWFs. It seems that trying to make everyone happy all the time is not a winning strategy here.

Inverted Rank: Table 8 shows the top-10 when using the Nauru method, or inverted rank. This approach strongly favors the top artists from the different sources, leading to something similar to a round robin method. All sources get to contribute strongly to some of the entries. Moreover, artists are not penalized as much for not occurring in all sources. This method actually is the best for P_{swf} since the result has something for everyone. However S_{swf} does not do as well as order is not being preserved.

Run Off: Table 9 runoff method does surprisingly well, especially at preserving the order of results. This is reflected in both a good overall distribution of influence, as well as the best S_{swf} score of any examined.

Round Robin: As Table 10 shows this approach does fairly well at representing all sources (with a P_{swf} of 30) but poorly at preserving order (with a S_{swf} of 123).

6 Conclusion

At a fundamental level, each music site engenders different behavior from its users. To add to the complication, each site does not cleanly partition music fans; in other words it is highly likely that fans use many sites in parallel. For this reason, no site can be labeled as the "holy grail" when it comes to measuring buzz and popularity, and just because a site has a very large user base doesn't necessarily mean it is more influential. Thus, providing a unified view of music buzz and popularity that captures different types of endorsements from all sites is key. It is also important to note that some sites have popular artists that are very site-specific and exist almost exclusively in that medium (such as Jeffree Star on MySpace). These border cases can be difficult to handle; what is the fair way to handle artists that have extensive endorsements from one site but are almost unmentioned in others?

We have shown that voting theory provides some excellent approaches for compiling a top ten list that takes into account all sources. In compiling and analyzing rankings based on different voting schemes, we found that the lists often provide a trade-off between determining a winner based strictly on shear volume by simple summing (thus the YouTube top-10 very heavily influences the ranking) versus depth across all modalities (ranking who is most popular in each site, then combining these ranks and thus discounting the volume difference between different sites). Lists that take into account popular artists on each site are far more useful for several reasons. Having a broad base of support on many sites is a measure of popularity in and of itself. Secondly, they take into account that it takes significantly more effort to write a comment than view a video, though video views are several orders of magnitude more numerous than comment volumes.

Compared to the simple, and largely used metric of adding "counts" to create a list of popularity, we find that for the SWFs we examine, there are voting schemes that produce results that are up to 45% better. Looking across the range, voting scheme performance can vary by as much as 168%. We know that among voting systems there can be no clear winner – there are many different ways to think about ranking in diverse communities. Each approach has its tradeoffs, but selecting a SWF that well describes what is important in the final ranking allows evaluation and selection of the proper approach.

Independent of ranking, almost all of the top 10 artists of any voting system appear in the top 40 of all other voting systems. Thus, it is not hard to discern which artists are generally popular, but gauging the level of that popularity can be problematic; viewing a video on YouTube or making a comment on MySpace are both forms of endorsement that cannot necessarily be compared equally. So, top artists are relatively easy to agree on, but the ordering of those top artists is much more controversial as the support base of an artist varies by site.

7 Acknowledgments

We would like to thank Geoff Goodwin, Head of BBC Switch, for his vision, support and encouragement, and Beth Garrod, BBC Producer, for providing domain expertise and guidance. We would also like to thank Daniel Danciu, Christoph Schwab-Ganser, and Matt Wiseman on the Google YouTube team for their timely assistance.

References

[1] S. Adali, B. Hill, and M. Magdon-Ismail. The impact of ranker quality on rank aggregation algorithms: Information vs. robustness. In *ICDEW '06: Proceedings of the 22nd Int. Conf. on Data Engineering Workshops*, page 37, 2006.

[2] K. J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1951, 2nd ed., 1970.

[3] M. Balinski and R. Laraki. A theory of measuring, electing, and ranking. *Proceedings of the National Academy of Sciences*, 104(21):8720, 2007.

[4] M. Balinski and R. Laraki. A theory of measuring, electing, and ranking. *PNAS*, 104(21):8720–8725, May 2007.

[5] A. Bergson. A Reformulation of Certain Aspects of Welfare Economics. *The Quarterly Journal of Economics*, 52(2):310–334, 1938.

[6] E. Codd, S. Codd, and C. Salley. *Providing OLAP (on-line Analytical Processing) to User-analysts: An IT Mandate*. Codd & Date, Inc., 1993.

[7] J.-C. de Borda. *Memoire sur les elections au Scrutin*. Histoire de l'Acad. R. des Sci., 1981.

[8] P. Diaconis and R. Graham. Spearman's Footrule as a Measure of Disarray. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):262–268, 1977.

[9] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. Rank aggregation methods for the Web. *Proceedings of the 10th Int. Conf. on World Wide Web*, pages 613–622, 2001.

[10] R. Fagin, R. Kumar, and D. Sivakumar. Efficient similarity search and classification via rank aggregation. *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 301–312, 2003.

[11] D. Ferrucci and A. Lally. Uima: an architectural approach to unstructured information processing in the corporate research environment. *Nat. Lang. Eng.*, 10(3-4):327–48, 2004.

[12] D. Freitag. Information extraction from html: application of a general machine learning approach. In *AAAI '98/IAAI '98: Proceedings of the 15th Conf. on Artificial Intelligence/Innovative Applications of AI*, pages 517–523, 1998.

[13] D. Freitag and N. Kushmerick. Boosted wrapper induction. In *Proceedings of the 17th National Conf. on Artificial Intelligence/Innovative Applications of AI*, pages 577–83, 2000.

[14] N. Kushmerick. *Wrapper Induction for Information Extraction*. PhD thesis, University of Washington, 1997.

[15] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel. Performance measures for information extraction, 1999.

[16] M. McIntyre. Hubbard hot-author status called illusion. *San Diego Union*, April 15 1990.

[17] W. H. Riker. *Liberalism Against Populism*. Waveland Press, Inc,, Prospect Heights, 1982.

[18] D. G. Saari. *Geometry of Voting*, volume 3 of Studies in Economic Theory. Springer-Verlag, 1994.

[19] A. P. Sheth. Changing focus on interoperability in information systems: From system, syntax, structure to semantics. *Interoperating Geographic Information Systems*, 1999.

[20] S. Soderland. Learning to extract text-based information from the world wide web. *Proceedings of 3rd Int. Conf. on Knowledge Discovery and Data Mining*, 1997.