REGULARITY AND FORECASTING OF MUSICAL TASTE[1]

Dan Gang and Daniel Lehmann[2]
MusicGenome.com Inc., Israel

Background

We describe the core technology of a system used to learn automatically the musical taste of a user. This technology, together with search services, is part of sophisticated listening stations located in retail shops and proposed to e-commerce and m-commerce vendors, in particular the cellular market.

Aim

The following paper introduces data, methods and results of a research that was conducted to learn some aspects of regularity and forecasting in musical taste.

Method

Information is gathered from chosen subjects: they rate a plurality of songs, describing indirectly their musical taste or preferences. The information is analyzed and reveals some patterns of regularity of the musical taste of the subjects. Based on this information we can predict with good accuracy the tastes of customers known only by a very small sample of their preferences. For example, if 30 subjects have rated a relatively large catalog of songs (a few thousands), given a small (a dozen) arbitrary set of songs rated by a new listener/ subject, it is possible to forecast successfully (with about 80% success rate) his rating for other songs of the catalog.

Results

1. A number of subjects (from 28 to 32) have been asked to reveal their personal taste on a catalog of songs (2200 to 5570 songs). To this aim the subjects provided ratings of songs on a scale of five values: from strong dislike to strong approval. The ratings were collected easily from the subjects and were found to contain a high degree of regularity and consistency in describing the subjects' preferences. Repeated ratings by the subjects for sub-set of the songs were found very consistent with differences of about 10% and with low standard deviation (almost all changes in the ratings were of one scale level).

2. A matrix d(i,j) whose entries are "pseudo-distances" for each pair of songs was built. The distance between songs i and j is small if typical subjects who like song i also like song j, while subjects who dislike song i also dislike song j. Using a small but fixed number of subjects, each of whom rated a large part of the catalog, allowed us to build a very detailed "pseudo-distances" matrix (as opposed, for example, to gathering information from a large number of subjects who rate only a small portion of the catalog). We assume that, if the pseudo-distances express the ratings of a large enough group of diverse enough subjects, a pseudo-distance between song i and song j close to zero reliably means that most persons who like song i also like song j.

3. Given a matrix of pseudo-distances d(i,j) and a vector of ratings (i.e., the learning set), of some arbitrary songs (S1..Sn) by a new

[1] Protected by pending patent

[2] Prof. Daniel Lehmann is on leave from the Hebrew University, Computer Science Institute.

ISBN: 3-931852-66-0
ISSN: 1617-6847

subject, it is possible to forecast successfully his rating for a new song i. The predicted rating for a song i is: the weighted average of the ratings for songs S1..Sn, where the rating of song Sm (m=1..n) is weighted by a quantity that is inversely proportional to some fixed power of d(m,i). We present data supporting the claim that, even with a number of subjects as low as 15 the method described above provides a reasonably successful prediction. The quality of the forecasting for a new user is a function of the size of the random sample he has rated, the number of subjects used to build the pseudo-distances, and their variety.

Conclusions

The method developed here combines two different types of information to predict the rating the user would give to each musical item considered. The first type of information features ratings of songs reflecting the taste of the user. The second type of information involves a number of listeners who rate a large number of songs of the catalog. The fact that an algorithm to forecast personal musical taste can be developed is evidence of the deep structure of regularities in regard to musical taste. This work can provide an initial quantification and qualification of the complex relations among listeners, the musical pieces and the musical taste.

ISBN: 3-931852-66-0
ISSN: 1617-6847