

情報システム工学科 平成 19 年度後期「自主課題研究」
研究テーマ:Web データベースを用いたリコメンデーションシステム

担当教員 笠原禎也 名列番号 042 堂前幸司

1 目的

今日、私たちのまわりには多くの情報が錯綜している。その中で個人が所有する情報を集団全体で共有することにより、その人にとって有益な情報を探し出すという考えがある。その手法について調査し、実装・検証する。

2 内容

2.1 Web データベースについて

Web データベースを用いることにより、ネットを通じて複数のユーザがデータベースを利用することができる。

2.2 協調フィルタリングについて

協調フィルタリングを用いて音楽を推薦するシステムを構築する。各ユーザの音楽の再生数を評価値とし、推薦する音楽は次の手順で求めることができる。

[Step 1]ユーザ(Y)と他のユーザ(O)を選択

[Step 2]相関係数の計算

相関係数は次式から求まり、ユーザ間の相関関係を調べることができる。ただし X_i と \bar{X} はそれぞれの評価値と平均値である。

$$r = \frac{\sum_i (Y_i - \bar{Y})(O_i - \bar{O})}{\sqrt{\sum_i (Y_i - \bar{Y})^2} \sqrt{\sum_i (O_i - \bar{O})^2}}$$

[Step 3]予測値の計算

ユーザがまだ聞いたことのない音楽をどう評価するかは、次式のように他のユーザによる評価値に相関係数の重みをつけてその合計を計算することで求まる。そして最も大きな値となる音楽を推薦する。

$$p = \bar{Y}_B + \frac{\sum_{J \in \text{other}} (J_x - \bar{J}) r_{YJ}}{\sum_J |r_{YJ}|}$$

2.3 リコメンデーション機能の製作

実際に Web データベースを用いたリコメンデーションシステムを構築し、シミュレーションを繰り返し行った。

3 結果と考察

シミュレーションの結果、ユーザがまだ聞いたことのない音楽に対してそれぞれ予測値が求められ、さらにその中で最も値の大きい音楽が推薦されていることが確認された。しかし、相関係数がどれも小さな値となり、あまり結果に反映されていないことがわかった。

そこで推薦の精度をよくするためにユーザが数回しか聞いたことのない音楽を評価対象から外すフィルタを用いて再度シミュレーションを行った結果、フィルタをかける前とは異なる結果が得られる場合もあった。また、ユーザが以前に一度だけ聞いたことのある音楽が計算した結果、推薦される(再発見)という面白い結果も得られた。

4 感想

実際に自分でシステムを構築することによって原理やその手法などがよく理解できた。また、リコメンデーションを行うためには多くの情報が必要であり、複数のユーザがそれぞれの情報を共有しあうという一種のコミュニケーション体系に大変関心が持てた。情報量が増加し続ける今日、このような技術が必要となっていくと考えられる。

参考ページの URL

協調フィルタリングに関する研究動向

<http://www.race.u-tokyo.ac.jp/~fukuhara/Research/paper/98/cofil.pdf>