ブログにおける楽曲評価の学習による分類

060427361　村林 翔（佐川 雄二）
名城大学　理工学部

1. はじめに

現在の Web では情報量が非常に多いことがかえって情報の迅速な収集を阻害している．音楽に関しても，Web 上から楽曲をダウンロードできるようになり，電子化が進んでいる．そのため，ブログや CD レビューなどにおける感想や評価も多くなっている．しかし，ユーザが聴きたい楽曲の感想や評価を知りたい場合，多くの情報を見なければならないため，知りたい感想や評価を得るのに時間がかかってしまう．

そこで個人の意見や感想が強く反映されるブログにおいて，書いてある楽曲の感想や評価を収集し，それらを良い評価と悪い評価の 2 値評価で分類し，ユーザに提供するシステムを提案する．本研究では，機械学習による 2 値評価分類のシステムの作成を行う．

2. ジャンル

それぞれのユーザには音楽に対する好みが存在する．ある人が Pop と感じる音楽でも，他の人が聴いた場合 Rock と感じることがある．また，楽曲を票差する際にもジャンルにより表現が異なることが考えられる．

そこで本研究では，楽曲のジャンルを指定してもらうことで，好みによる評価の違いを解消する．ジャンルには，iTunes や CD の通販サイト等を参考に以下の 7 つを使用する．

- Pop
- Rock
- Hip-Hop
- R&B
- Reggea
- Jazz
- Classic

3. 学習に使用するブログ

3.1 ブログの収集

機械学習に使用するブログを Ameba，Yahoo!ブログ，JUGEM 等のブログサイトから上記のジャンルに当てはまる楽曲の感想をそれぞれのジャンルにつき，40 件ずつ，計 280 件のブログを収集した．

3.2 アンケートの実施

3.1 で収集したブログに対し，アンケートを実施する．それぞれ 1 つのブログにつき 3 名に以下の 2 項目について回答をしてもらう．

- ジャンルの種類
- 良い評価か，悪い評価か

3.3 アンケート結果

3.2 から得られたアンケートをもとにそれぞれのブログを分類する．分類は，3 名のうち 2 名以上の 2 項目が一致するブログについてのみ行う．一致しないブログについては除外する．分類するカテゴリは，7 つのジャンルに対し，良い評価と悪い評価の 14 カテゴリとする．

4. 機械学習

4.1 使用する学習法

本研究では，カテゴリに分類をする点やテキストが対象となる点からナイーブベイズによる学習を使用する[1]．ナイーブベイズの式については以下の式で表される．

$$P(c_i \mid d_j) = P(c_i)\prod_{k=1}^{n} P(t_k \mid c_i)$$

ここでの c_i はカテゴリ，d_j はテキスト，n は d_j に含まれる単語数，t_k は単語を表す．$P(c_i)$ と $P(t_k|c_i)$ については以下の式で求める．

$$P(c_i) = \frac{N_i}{N} \qquad P(t_k \mid c_i) = \frac{1 + N_{kj}}{M + \sum_{l=1}^{M} N_{li}}$$

ここでの N は学習用データのテキスト数，N_i は学習用データのうち c_i に分類されたテキスト数，N_{kj} はカテゴリ c_i 中に単語 t_k が出現した回数，M は単語の種類数を表す．

4.2 学習結果を用いた分類

分類対象となるブログを text ファイルに保存し，ジャンルを選択することで，どちらに分類されるかの確率を結果として出力する．

5. まとめ

楽曲評価が書かれているブログを収集し，学習を行うことにより，対象となるブログを分類するシステムについて述べた．今後はブログの品詞を限定することやストップワードの削除により正確性を高めることを考えている．またその後は，実際に Web 上にあるブログを自動で検索，収集するシステムを組み込むことを考えている．

参考文献

[1]斉藤大：機械学習を用いたテキスト分類手法の調査：東京大学(2006).