

5 C - 5

LVQとHMMを用いた 音楽コード認識システムの実装

小杉 信太郎 山口 文彦 中西 正和

慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

本研究では音楽音響信号から音声認識でしばしば用いられるケプストラムを学習ベクトル量子化 (LVQ) によってベクトル量子化しその量子化されたベクトルを隠れマルコフモデル (HMM) によって学習させ音楽コードを認識するという手法を用いることでコード認識システムを実装した。

2. 本システムの構成

本システムは音響信号特徴量抽出部、ベクトル量子化部、記号復合部から構成される。

音響信号特徴量抽出部

コードの特徴をとらえるのに、ふつう、フーリエ解析や帯域フィルタバンクを用いて、構成音の分解を行い構成音の種類によってコードを決定するのが主流の方法であるが、音響信号の特徴を捉えるのに本システムでは音声認識でしばしば利用されるケプストラムを利用する。ケプストラムをコードの特徴量とする場合、構成音の分解はできない。このため本システムでは1つのコードのケプストラムを1つの特徴量とみなす。また、ケプストラムは低ケフレンシー部に基本周波数情報が数多く含まれていると考えられるので、低ケフレンシー部の40点を特徴量としこれを1つのベクトルとする。

ベクトル量子化部

ケプストラムによる特徴量ベクトルを抽出したのち、それらをLVQ[4]により最適に学習されたコードブックベクトルによってベクトル量子化のクラス分けを行なう。LVQはベクトルの学習によってベクトル量子化を最適におこなう為の手法である。このクラス分けを行う部分をベクトル量子化部と呼ぶ。この処理部の主要な目的はHMMを用いた記号復号部へのデータ入力である。入力されたケプストラムデータはこのベクトル量子化部によって、データがどのクラスに属するか判定される(図1)。

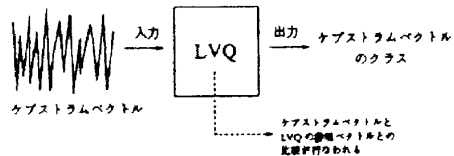


図1: ケプストラムのクラス分け

記号復号部

記号復号部はベクトル量子化部によって逐次出力されるコードクラスの記号時系列をあるセグメント長をひとつのコードとして対応させるように、復号するための部分である。この部分の簡略図を図2に示す。この部分の必要性はコードはある瞬間において決定

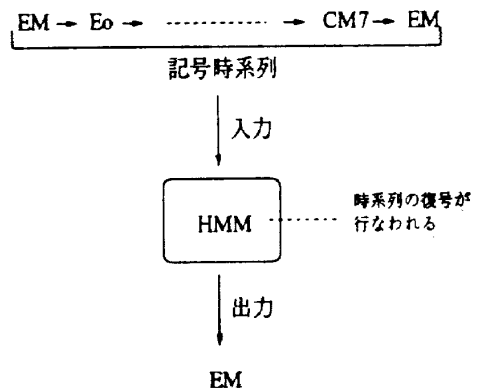


図2: 記号復号部

されるものではなく、ある程度の時間長の中で決定されるという特徴があるからである。すなわち、ベクトル量子化部において出力されるクラス名によるコードの分類のみでは約0.1秒分のコードクオリティーのみでしかなく、それによってコードの分類をおこなうことは非現実的であると考えられるからである。本システムでは記号時系列の復元には音声認識でしばしば利用される left-to-right の HMM を用いる。HMM を用いれば多くの記号列が1つのコードに対応するか、そして最も可能性のあるコードが何であるかを最適に決定できると考えられるからである。コード認識のための HMM はそれぞれのコードごとに作成する。HMM 学習用のデータには生コードのケプストラムデータをベクトル量子化したものを使用する。

Speaker Verification

Shintaro KOSUGI Fumihiko YAMAGUCHI Masakazu NAKANISHI
Department of Computer Science, Faculty of Science and Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223, Japan

システムでのコード決定法

本システムでは逐次コードの出力を1ヶブストラムの入力につき1つのコードを出力として返すが、基本的にコードは小節単位で決定するので、コードの決定は次のようにする。まず、本システムでは入力にはコードの定まる1小節であるとする。この仮定のもとでコードの決定は1小節中のシステムの出力結果が最も多いものであるとする。すなわち、1ヶブストラムの出力を1つの票として、多数決によるコードの決定をおこなう(図3)。

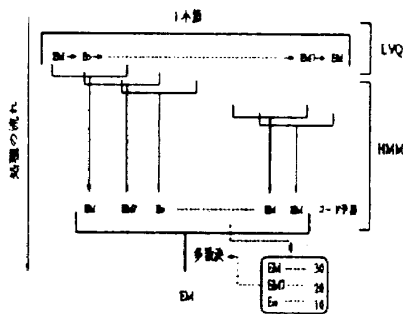


図3: コードの認識手順

3. 実験および考察

本システムの仮定

本システムでは現時点では以下の仮定のもとで実装される。

- 音種は1つに限定する。
- 音の幅は1オクターブに限定する。
- 雑音は存在しない。

実験1

実験1として、生コードのみを入力としたときのシステムでの認識率を調べる。ここで生コードとは、システムのコード学習用に用いたものと同じものを指す。

実験1の結果および考察

HMMの再推定回数が0回のときは全く認識能力はない。HMMの再推定回数を10回にしたとき認識率は軒並向上し100%の認識率を得られたコードもあった。このことから、HMMの学習が確かめられた。再推定回数が20回のときは10回のときよりも低い平均認識率であったが、これは過学習によるHMMの認識率の低下が原因であると考えられる。また、認識率が0%のものもあり、これはこのコードが他のコードと誤認識されているためで、コードの分類精度については完全とは言いがたい。再推定回数が10回のとときと20回のとときを比較すると10回のほうが認識率

が高い場合と低い場合があることがわかる。また、認識率の値に極端な差が生じていることがわかる。あるコードについては100%の認識を得られているが、一方では0%の認識のような差がでている。これは、HMMによるコードの認識は可能であるが、パラメータの再推定回数によって認識率が著しく変動すると考えられる。今回、コードの学習においてコードの総数が多いことと、HMMのパラメータの推定に多くの時間を要するため、すべてのコード認識HMMのパラメータ再推定回数を同じにして、一つ一つが最適化されているかどうか確かめることなく推定した。このため、それぞれのコード認識HMMの精度にばらつきが生じたと考えられる。このことから、ばらつきを減らすためにひとつひとつのHMMについて精度の確認をする必要がある。

実験2

実験2では、実際に曲をデータとして入力したときのシステムのコードの出力結果を正しいコードと比較する。

実験2の結果および考察

曲中のコード認識では、曲のコード感が多く含まれる小節については当然だが認識率が高い。また、逆にコード感があまり感じられない小節では誤認識する率が高くなる。このための対策として、HMMの学習段階において、生コードを学習材料とするのではなく、曲そのものを学習対象とする方法が考えられる。このようにすれば、生コードを用いたときよりHMMの観測系列の変動をより学習することができ、曲でのコード認識率が向上するはずである。

今後の方針

本システムのコード認識はコード認識HMMの学習には生コードを使用したのが、これより曲をデータとして与えた方が学習には最適であると考えられる。また、本システムでは音域や音色にかなりの制限がある。これらの制限をできるだけ除くための方法を考える必要性がある。

参考文献

- [1] 古井 貞熙「音響・音声工学」近代科学社, 1992.
- [2] 古井 貞熙「デジタル音声処理」東海大学出版, 1985.
- [3] 中川 聖一「確率モデルによる音声認識」電子情報通信学会, 1988.
- [4] T.Kohonen「Self-Organizing Maps」, Springer-Verlag, 1995.