

Kolmogorov 記述量を用いた MIDI の分類

Clustering of MIDI with Kolmogorov Complexity

<http://www.tani.cs.chs.nihon-u.ac.jp/g-2006/fukumori/>

谷 研究室 福森 勤

Tsutomu Fukumori

概要

これまでに音楽を自動的に分類する様々な手法が提案されてきた。Ming Li らは Kolmogorov 記述量に基づいた similarity metric を提案し, Rudi Cilibrasi らは similarity metric を用いて音楽を自動分類する手法を提案した。本研究では Similarity Metric を用いて, MIDI の自動分類を目的とする。

1 はじめに

近年コンピュータ技術の発達に伴い, 様々な事柄が電子的に保存されるようになった, その中で音楽も「データ」として扱われるようになってきた。現在一部の専門家によって, 手動でデジタル化された音楽は分類されている一方で, これらの自動分類化が求められている。Ming Li, Xin Chen, Bin Ma, Paul M.B. Vitányi らは, Kolmogorov 記述量に基づいてデータ間の距離を表す similarity metric を提案した。これにより「データ」という観点から音楽を分類することが可能になり, 今まで無関係だとされていた事柄に対して新しい関係の発見が期待される。Rudi Cilibrasi, Paul Vitányi, Ronald de Wolf らの先行研究では, similarity metric を用いて, Classic や Rock など幅広いジャンルの音楽を取り扱い, それらの情報としての近さをグラフで表した。本研究室の三井による追実験では, 歴史的背景や地理的背景が MIDI に含まれているかを検証したが, 期待されるような結果は出なかった。そこで, 本研究では, 対象を MIDI に着目し, rubi 氏等の実験を再現することにより, 三井の結果の問題点を検証する。

第 2 節では MIDI の定義について, 第 3 節では Kolmogorov 記述量について, 第 4 節では similarity metric について述べ, 第 5 節では, 結果の表示方法について述べ, 第 6 節では, 実験結果について述べ, 第 7 節では, 今後の課題を述べる。

2 MIDI の定義

この節では, MIDI の基本的な構造について述べる。

2.1 前処理

今回の実験で取り扱ったデータは MIDI と呼ばれる音楽ファイルである。MIDI とは Musical Instrument Digital Interface の略で, 楽器間のデータ通信方法の規格を指す。音をデジタル化することによって, 音色というも

のは音程 (ピッチ)・波形・音量という 3 つのデータとして記憶される。この記憶されたデータを MIDI メッセージと呼び, 今後は 16 進数で表現する。メッセージの内容によっては複数のバイトにまたがるため, 最上位ビット (Most Significant Bit, 以後 MSB) がフラグのような機能を備えている。MSB が 1 の場合, そのバイトをステータス・バイト, 0 の場合データ・バイトと呼ぶ。ステータス・バイトにはメッセージの種類が示されており, データ・バイトにはデータの情報が書かれている。以下に MIDI メッセージの中から一例を紹介しておく。

三井による追実験では, ノート・オン/ノート・オフを抜き出し, MIDI チャンネルを固定する前処理が行われており, rubi 氏等による選考実験では, ペロシティを平均化し, MIDI チャンネルを固定する, 前処理が行われている。

ノート・オン/ノート・オフ ノート・オン/ノート・オフとは音を出す MIDI メッセージである。これらは必ずペアで存在し, 共に 3 バイト必要である。以下にメッセージ内容を示す。

ノート・オン

9n	kk	vv
メッセージ種類	ノート番号	ペロシティ

ノート・オフ

9n(8n)	kk	00
メッセージ種類	ノート番号	ペロシティ

n: MIDI チャンネル 楽器の設定を行う。ノート・オン/ノート・オフにおける, ステータス・バイトの下位 4 ビット (0~F) で表される。

kk: ノート番号 音程を表している。データ・バイトにあたるため, MSB は 0 で設定されている。よって, 00~7F

までで表される．ノート番号 60 とはドの音 (261.6Hz) に当たる．

vv:ベロシティ 音の強弱を伝えるパラメータのこと．ノート番号と同様にデータ・バイトあるため，00～7F で表される．00 とは，無音つまりその音を止めることである．ノート・オフメッセージには 9n～と 8n～の 2 種類があるが，一般的に 9n～が使われている．

実際の MIDI ファイルは，この演奏データに時間情報を加えた SMF(Standard MIDI File) と呼ばれる標準ファイルフォーマットとなっている．14 バイトで表現された「ヘッドチャンク」から始まり，「トラックチャンク」と呼ばれるトラックがいくつか連なって一つのファイルを構成している．それぞれのファイルで単位時間が定義されており，SMF はその時間を基準に MIDI メッセージを送信する順番を記憶しているだけである．つまり，プレイヤーでは「待つ」→「送信」が繰り返し行われているのである．

3 Kolmogorov 記述量の定義

この節では，Kolmogorov 記述量の基本的な定義について述べておく．

3.1 Kolmogorov 記述量

ある事柄に含まれている情報について考える．与えられた有限の対象 (たとえば文字列) に含まれる情報量を，その対象を生成するプログラムのサイズ (すなわちビット数) と定義する．ただし，この場合プログラムとは，空の状態からスタートし，すべてを出力するものとする．どんなプログラミング言語を選んでも，それが妥当であれば，情報量は定数分の差しかないことが証明されている．

対象の集合からなる定義域 $D \ni x$ について， s をプログラム言語， $|p|$ をプログラムのサイズ， $S(p) = n = S(p)$ となる関数， $n(x)$ を列挙する方法を仮定したときの x の順位とすると，記述量は以下のように定義される．

$$K_s(x) = \min\{|p| : S(p) = n(x)\}$$

ただし， x を生成する p が存在しないときは，

$$K_s(x) = \infty$$

となる．

また，ある対象 y が与えられているときの対象 x の記述量を定義する．文字列 x を，計算可能関数 f ，それに文字列 p と y により $f(p, y) = x$ と記述することを考え

る．そこで，任意の計算可能関数 f をインタプリタ関数あるいは複合関数と，任意の $p \in \{0, 1\}^*$ をプログラムと呼ぶことにする．インタプリタ関数 f のもとで，補助情報 y に対する x の記述量 K_f を以下のように定義する．

$$K_f(x|y) = \min\{|p| : p \in \{0, 1\}^* \text{ \& } f(p, y) = x\}$$

ただし，条件を満たす p がいないときは，

$$K_f(x) = \infty$$

となる．

また， x と y を区別できる形で出力させる最短のプログラム長を， $K(xy)$ と表す． $O(\log K(xy))$ の誤差範囲では，

$$K(xy) = K(x) + K(y|x)$$

となることが証明されている．

3.2 K のアルゴリズム的性質

$K(x)$ を正の整数から正の整数への関数と考える．

定理

- (a) 関数 $K(x)$ は帰納的ではない．しかも，どのような帰納的部分関数 $\phi(x)$ を考えても，もしその値が無限個の点で定義されているならば，その定義上のどこかの点で $\phi(x) \neq K(x)$ となる．
- (b) 引数 t に対して単調減少で (全域的な) 帰納的関数 $H(t, x)$ が存在し， $\lim_{t \rightarrow \infty} H(t, x) = K(x)$ ．すなわち， $K(x)$ のよい近似は計算可能．ただしこれは一様近似ではない．

3.3 情報量

相対記述量 $K(x-y)$ が絶対記述量 $K(x)$ よりかなり小さい場合，「 y が x についての情報を多分に含んでいる」と考えることができる．したがって，定数分の差異を無視すれば，「 y に含まれる x の情報量」は

$$I(x : y) = K(y) - K(y|x)$$

とみなすことができる．また， $K(x|x)$ となる x を選べば，

$$I(x : x) = K(x)$$

となる．ここで考えるべきことは，情報の対称性である． $O(\log K(xy))$ の範囲では，

$$K(xy) = K(x) + K(y|x)$$

が成り立つので、これを用いると、

$$I(x : y) = I(y : x)$$

が成り立つ。

4 similarity metric

数学において距離空間とは、任意の 2 点間で距離が定められた空間のことをいう。

定義

ある集合 X 上の距離とは、実数値関数 $d : X \times X \rightarrow R$ で任意の $x, y, z \in X$ に対して次のような性質を満たす。

$$d(x, y) \geq 0$$

$$d(x, y) = 0 \Leftrightarrow x = y$$

$$d(x, y) = d(y, x)$$

$$d(x, y) \leq d(x, z) + d(z, y) : \text{三角不等式}$$

これをもとに、情報距離について考える。

Ming Li らの研究では、情報に関する距離を標準化している。任意の文字列 x, y について、以下のように決める。

$$d(x, y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}}$$

また、 $K(y) \geq K(x)$ としたとき、

$$d(x, y) = \frac{K(y|x)}{K(y)}$$

これに対して先に述べた情報量の公式を用いると、

$$d(x, y) = \frac{K(y) - I(x : y)}{K(y)}$$

となる。また 2 節で示した通り、 $O(\log K(xy))$ の範囲では $K(xy) = K(x) + K(y|x)$ が成り立つので、

$$d(x, y) = \frac{K(xy) - K(x)}{K(y)}$$

と表すことができる。

実際の実験では、この式とともに以下の 3 つの理想理論のもとに行われた。

- (1) 要求された情報距離 $d(x, y)$ は漠然と長い文字列 x, y によって得られる。
- (2) Kolmogorov 記述量は帰納的でないため、計算不可能である。
- (3) 実用的な方法で情報距離を近似する際、圧縮方法の一つである“ bzip2 ”を用いる。

以上より、 $bzip(x)$ を文字列 x を bzip2 で圧縮したときのファイルサイズとすると情報距離 $d(x, y)$ は以下のように近似される。

$$d(x, y) / \approx \frac{bzip(xy) - bzip(x)}{bzip(y)}$$

この距離関数をもとに、データの分類を進めていく。

5 quartet method

次に、結果の表示方法について述べる。この実験では、“ quartet method ”という系統樹の一つを用いている。“ quartet ”とは、2 つの葉をもつ 2 つの subtree が連結したグラフを指す (Figure1)。ある n 個のデータ集合を S としたとき、 $S \ni u, v, w, x$ に対して quartet は $uv|wx$ と表現する。“ $|$ ”は 2 つの subtree に分解されることを表している。

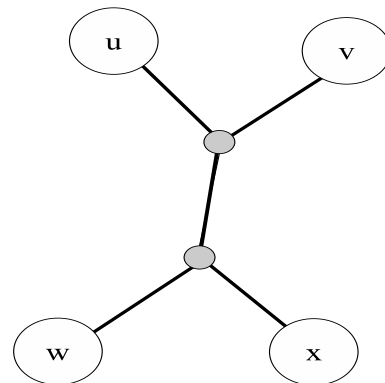
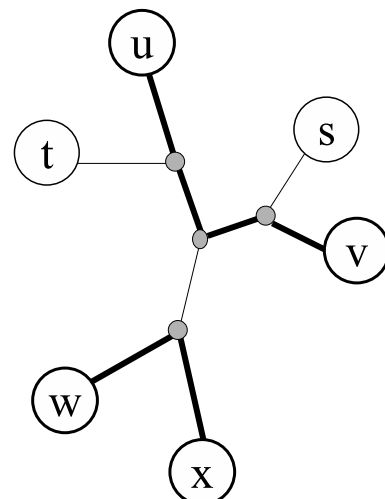


Figure1:quartet

quartet に対する cost を次のように定義する。

$$C_{uv|wx} = d(u, v) + d(w, x)$$

また、ある木 T が与えられたとき、 u から v までの辺と w から x までの辺が交わらないような $uv|wx$ のことを“ consistent ”と呼ぶ (Figure2)。



6.2 実験 2

スコットランド, 日本, アメリカ, フランス, スペインからそれぞれ 2~3 曲の民族音楽を集め, 実験を行った. tranceform は 70,000 回で固定した.

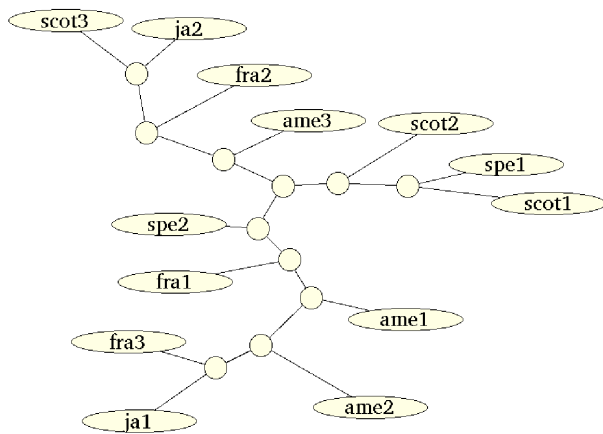


Figure6 :tranceform:70,000 $S(T) = 0.81024$

$S(T)$ の値は高くなっているのにも, かかわらず実験 1 と同様に, 地域ごとに際立った特徴が出なかった.

7 今後の課題

本実験では, MIDI ファイルから地域ごとの類似性を見つけることは出来なかったが, 違う前処理を行うこと

により, 地域ごとの類似性を見つけることが出来る可能性がある. また, MIDI ファイル以外の音楽ファイルを対象として見るならば, 地域ごとの特徴が出る可能性がある.

参考文献

- [1] Ming Li, Xin Chen, Bin Ma and Paul M.B. Vitányi. The Similarity Metric. In Proceedings of the 14th annual ACM-SIAM symposium on Discrete algorithms (SODA), pp.863-872, Baltimore, Maryland, 2003.
- [2] Rudi Cilibrasi, Paul Vitányi and Ronald de Wolf. Algorithmic Clustering of Music (2003)
- [3] Ming Li and Paul M.B. Vitányi. 渡辺 治 翻訳 Kolmogorov Complexity and its Applications. コンピュータ基礎理論ハンドブック (1994)
- [4] リットーミュージック出版編集部【編】 MIDI バイブル I MIDI 1.0 規格 基礎編 (1997)
- [5] <http://www.momo-mid.com/>
- [6] <http://www.tani.cs.chs.nihon-u.ac.jp/g-2005/taako/>