

時系列データの俯瞰と縮約のための可視化の一手法

理学専攻 情報科学コース 林 亜紀

1 概要

近年、多次元の時系列データの可視化に関する様々な手法が提案されている。本論文では俯瞰と縮約によってこれらを可能にする可視化手法を提案し、クラシック楽曲を対象とした音楽構造の可視化事例と、システムログを対象とした時間的特徴変化の可視化事例を示す。

2 クラシック楽曲への適用事例

クラシック楽曲の中でもオーケストラ楽曲のような、たくさんのパート(楽器)が同時に異なるメロディを演奏する楽曲の楽譜は、段数が多いため、特に初学者にとって読み取るのは難しい。しかし、このような楽曲を扱う作曲家や編曲者、演奏者は曲の全体像を短時間で捉えたいという要求と、効率的に他の編成にアレンジしたいという要求を持つ。そこで本手法は、主に初学者を支援するために、楽曲の音楽構造の俯瞰表示と縮約表示を実現する。

2.1 提案手法

音楽構造の分析 本手法では楽譜データに MIDI を用いる。MIDI では、楽譜上の各パートの各音符について、発音のタイミング、音程、強さなどが数値で記述されている。まず、役割判定のためのパターンを付与する。パターンとは、1パートで構成される数小節単位の短い楽譜のことであり、本手法では経験者が付与するものとする。また、役割とは主旋律、伴奏(和音)、伴奏(低音)などである。続いて、与えたパターンと各ブロックをマッチングさせることで、各ブロックの役割を反復的に判定する。本手法のブロックという概念は、一般的にいうフレーズに該当し、同じパートの連続した音符で構成される。ブロックとパターンの類似度算出の際には、文献 [1] の RhythmicActivity(発音タイミングの特徴量) と MelodicActivity(音程遷移の特徴量) をもとに、各ブロックと各パターンのメロディとの距離をコサイン類似度によって求める。

縮約のための前処理 縦方向の縮約表示では、ユーザの指定した段数にスコアを縮約する。この際、元の楽曲で隣接している数段の中から、主旋律、伴奏、その他の順で、よりパターンとの類似度が高いブロックを選び、それを縮約結果の各段に割り当てる。縮約結果は、MIDI 形式で出力可能なため、効率的な編曲支援ができると期待される。横方向の縮約表示では、より簡潔に音楽構造を可視化するために、各パートやその役割が前の小節から変化しているかに着目し、縮約表示する。変化のない小節を縮約して表示することで、全体の長さを短く見せることができる。

ユーザインタフェースでの描画と操作 縦方向をパート、横方向を時間、役割を色として役割分析結果を描画する。ユーザは、インタラクティブに分析結果の俯瞰表示と操作、縦横両方向の縮約表示を適用することができる。また、役割分析の際に求めた、パターンとの類似性が高いものだけを表示する機能や、楽曲の再

生機能も実現した。

2.2 実行結果

チャイコフスキーの「花のワルツ」を用いた曲全体の俯瞰表示結果を図 1 に示す。楽譜の段数は 16 段で、ページ数は 33 である。この結果から、クラシック楽曲において典型的な、2つの主旋律が変奏を伴って繰り返され、音楽構造を構成する様子を把握することができる。また、後半に向けて登場するパート数が増えていく様子や、類似した主旋律が演奏されていてもそれを担うパートが変化の様子、徐々に対旋律が加わる様子も観察できる。図 2 は、同楽曲の一部の縮約結果(4段)を MIDI 形式で出力したもので、直接的な編曲支援に役立つと考えられる。

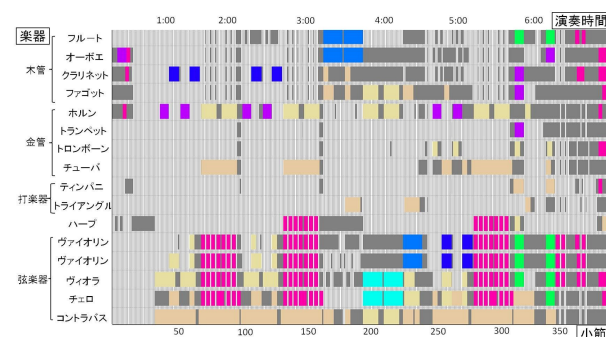


図 1: 提案手法の可視化(俯瞰表示)結果例



図 2: 縦方向縮約を MIDI 形式で出力した結果

3 システムログへの適用事例

クレジットカードの決済情報や Web ページのアクセスログなどシステムログは、情報量や属性数が多く、観察は困難である。しかしこれら进行分析することで、決済情報からは決済の傾向、不正決済特有の傾向を、アクセスログからはアクセスやエラー発生傾向、更新に適した時期などを知ることができる。本論文では、各属性に対する時間的特徴変化の俯瞰・縮約による全体像把握、効率的な属性選択の支援を行う Visual Analytics Tool を提案する。

3.1 提案手法

時間的特徴の俯瞰表示 X 軸を時系列、Y 軸を属性値(項目)、色を該当件数(暖色ほど集計値大)として、各座標に集計結果を描画する。X 軸には日付、曜日、月、

時間が、Y 軸には時系列を含む全属性が選択できる。属性値で描画するログを限定するフィルター機能や、集計値があまり大きくなかったものにも注目できるよう、色の割り当てを調整する機能も実現する。

Y 軸選択の推薦機能 有意性のある可視化結果を得られる属性を推薦する機能を提案する。選択された以下のいずれかの値を相対的に算出し、各属性の推薦度をインタフェース上のボタン色の濃度で提示する。

- (1) 全体の最大値 特異な集計が存在
- (2) 各項目の最大値の平均 特異な集計が頻発
- (3) 項目の合計値の最大値 特異な項目が存在
- (4) エントロピーの低さ (ちらばりの大きさ)
- (5) エントロピーの高さ (ちらばりの小ささ)

可視化結果の縮約 俯瞰表示では、例えば URL が数千～数万種類におよび、属性によっては画面上の水平な帯が増え、拡大操作でちらつきが見られる場合がある。そこで、2 種類の縮約機能を提案する。

ソートによる縮約では、以下のいずれかの値を項目ごとに算出してソートし、スライダーで表示項目数を調節する。

- (1) 各項目の最大値
- (2) 増加量の最大値
- (3) エントロピーの低さ (ちらばりの大きさ)
- (4) エントロピーの高さ (ちらばりの小ささ)

ソートによる縮約の際、一定の周期性がある項目など、有意性があるにも関わらず描画されない項目も出てくる。そこで、項目をクラスタリングする縮約を提案する。k-means 法で 30 個にクラスタリングした後、各クラスタの代表を描画することで縮約する。加えて特定のクラスタに属する項目全体の描画も可能である。

3.2 実行結果

不正の傾向分析 (2007/7-12 月の不正決済) 図 3(左上) は X:曜日 (左が日曜), Y:商品コード (推薦 (4) より) とした結果をソート (1) で縮約した結果の一部である。元の結果では商品コードがない決済が 1 行目に配置され、商品コードのついていない項目には青色ばかりが割り当てられてしまったが、特異な項目を取り除いて色の再割り当てを行うことで、可読性が向上した。海外商品 (1 行目) は継続的に多いが、電化製品 (2 行目) や鉄道 (3 行目) は土日に多く、通販 (4 行目) は平日に多い。図 3(左下) は電化製品に限定して X:日付, Y:加盟店コード (推薦 (3) より) とした結果である。2 行目の加盟店で継続的に不正が見られる。Y:カード ID とすると、同一 ID ではなかった。さらに図 3(右) で 2 行目の加盟店を [2] の散布図で観察した。X:金額, Y:不正種別, 色:曜日としたところ、土日に 10 万円前後の偽造による不正が集中しており、この加盟店が偽造カードの標的となっている可能性が浮かび上がった。

アクセス・エラーの傾向, 更新時期の検討 図 4(左上) は, X:月, Y:URL として, (1) でソート・縮約を行った例である。元の項目数は 6000 で, 縮約により可読性が大幅に向上した。アクセス数が多いことが自明な top や index に加えて, 教員のある授業全 11 回のうち,

7,8 回目の資料だけアクセス数が多いことが分かった。図 4(左中央) で, 7 回目の資料について, X:時間 (左が 0 時), Y:曜日 (下から日, 月...) としてアクセス時期を観察した。月曜午前の授業に向けて, 日曜夕方から月曜朝にアクセスが多く, 特に日曜の 23 時にアクセスが集中しており, 更新はそれ以前に行う必要があることが分かった。続いて, 図 4(左下) で, X:月, Y:ステータスコード (推薦 (4) より) としたところ, 最も一般的な成功 (200) よりも, 部分的成功 (206) が多いことが分かった。この結果からアクセス増の原因がこの 206 ではないかと考えられる。続いて図 4(右) で, 全授業資料のアクセス数を X:URL, Y:ステータスコード, 色:転送量, 高さ:件数として [2] の散布図で観察した。予想通り 7,8 回目の授業のステータスコード 206 のアクセス数が突出している。観察を続けると 206 が発生した IP アドレスは多様であり, 多くの学生が 7,8 回目の資料の入手に苦労している可能性が発見できた。

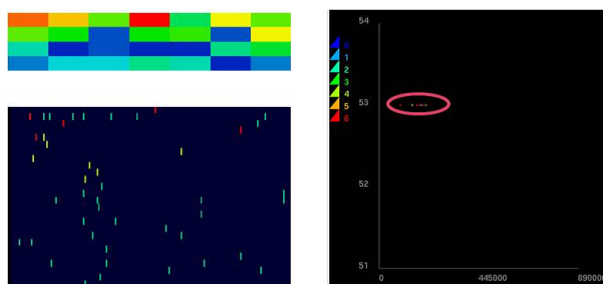


図 3: 2007/7-12 月の不正決済の傾向分析例

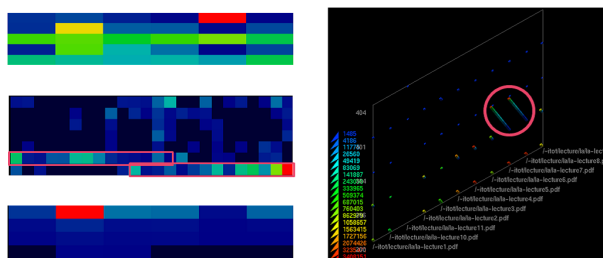


図 4: 2011/4-9 月のアクセスログ分析例

4 まとめ

本論文では, 多次元時系列データを効果的に俯瞰・縮約表示する可視化手法を提案し, 2 つの適用事例を紹介した。

謝辞

楽曲への適用に際し貴重なご助言をいただきました慶應義塾大学松原正樹氏, 決済情報をご提供いただきました株式会社インテリジェントウェイブ様に感謝いたします。

参考文献

- [1] 松原正樹 et al., ScoreIlluminator: スコア色付けによるオーケストラスコアリーディング支援システム, 情報処理学会論文誌, Vol. 50, No. 12, pp. 1-12, 2009.
- [2] C. Sakoda et al., Visualization for Assisting Rule Definition Tasks of Credit Card Fraud Detection Systems, IEEEJ Image Electronics and Visual Computing Workshop, 2010.