

word2vec による MIDI 形式楽曲データの解析

理工学研究科 数理情報学専攻

T14M001 川野 康平

指導教員 馬 青

概要

現在、類似楽曲検索の需要が高まっている。音楽のオンライン販売やストリーミング配信などのサービスが始まり、膨大な楽曲群の中から自分好みの楽曲を探し出す機能が求められている。しかし、音楽作品は、漫画や小説などと比べて要約が難しく、一つ一つの楽曲を試聴して自分が求める楽曲かどうかを判定するには膨大な時間的なコストや労力が必要となる。そこで、自動で似た楽曲を検索するための様々な手法が提案されている。

類似楽曲を検索する手法には、解析する楽曲データに対し周波数解析を行って特徴検出をするものや、楽曲を未知の言語で書かれた文章であるとして、自然言語処理の手法を応用するものなどがある。これまでの自然言語処理的な手法では、TF-IDF などを用いて楽曲の特徴ベクトルを生成し、それらのベクトル間の距離を定義することで楽曲間の類似度を測っている。そのため、楽曲全体の統計的な特徴は考慮されているものの、楽曲の特徴が登場する順番や文脈が考慮されないという問題がある。楽曲を聴いて受ける印象は、自然言語の文章のように登場するフレーズの順序や文脈に依存する。類似楽曲と判断するためには、局所的なフレーズの一致などを見るのではなく、フレーズの順序や文脈を考慮する必要がある。

そこで、本研究では曲フレーズが持つ文脈情報を含んだ楽曲の特徴ベクトルを作り出すために、word2vec を用いた楽曲間の類似度を測る手法を提案する。word2vec とは自然言語処理の分野で提案された単語特徴ベクトルを獲得する手法の一実装である。単語特徴ベクトルはニューラルネットワークによる学習で獲得される。word2vec では、単語とその周辺にある単語を用いて単語の共起確率を計算する。これにより、ある単語は周辺にどの単語があるときに登場しやすいかが学習され、word2vec により獲得される単語特徴ベクトルは、それぞれの単語がどのような文脈で使用されるかを表現したものになる。

word2vec は自然言語処理の分野の手法であるため、単語を持たない楽曲にそのまま適用することができない。そのため、楽曲に word2vec を適用するために、MIDI 形式の楽曲データを一度 MML 形式に変換し、更に、MML 形式の楽曲データから単語を作る処理を行った。この単語を本論文では音声単語と呼び、MML 文から固定の長さで切り出すことで生成した。

2 曲、あるいは 3 曲の楽曲に対して音声単語を生成し、word2vec を用いて特徴ベクトルを獲得した。また、ある楽曲の一部を別の楽曲に挿入した楽曲に対しても同様に特徴ベクトルを獲得した。得られた特徴ベクトルは主成分分析を用いて 3 次元に図示することで、その特徴を捉えられることを確認した。

次に、曲数を増やし、独自に定義した類似尺度を用いて楽曲間の類似度を測り、結果を検討した。実験には JPOP と演歌調の楽曲、クラシック楽曲を使用した。クラシック楽曲や演歌調の楽曲はそれぞれ JPOP とは違う音楽構造をしているため、これらの楽曲間の類似度は小さくなることが期待できる。実験の結果、いくつかのパラメータでジャンルによる楽曲間の尺度分離を行えることが確認できた。今後の課題として、音声単語の生成方法や、word2vec による学習の際に考慮する周辺語の数の調節などが挙げられる。