A-P07

深層学習による音響信号の特徴抽出に関する実験的考察

北海道科学大学 ○丹羽 孔明, 会津大学 成瀬 継太郎, 北海道科学大学 大江 亮介,
北海道科学大学 木下 正博, 北海道科学大学 三田村 保, 北海道科学大学 川上 敬

要旨

深層学習は多段階的な特徴抽出器を自己形成し, 混合した情報の中から特徴となるパーツを学習できることが知られ, また特徴ベクトルから元の情報を復元できる. 本稿では深層学習により音楽の音響信号をネットワークに学習させ, また音響信号を学習したネットワークが音楽を記憶することを示すことで, 深層学習によるネットワークが音響信号に復元可能な特徴ベクトルを学習することを実験的に考察する.

1. はじめに

コンピュータサイエンスの方法論による音楽の自動生成は L. Hiller と L. Isaacson によって初めて実現された. この後, 現在までに多くの自動作曲システムが提案されている. D. Cope の Experiments in Musical Intelligence (EMI) [1] は数あるシステムの中でも質の良い音楽を自動的に作曲する代表的なシステムである. EMI はある音楽の断片に続く断片をあらかじめ分析して得られた音楽のルールに従って決定する. 近年では深層学習のアプローチによる自動作曲も提案され, Deep Belief Networks (DBNs)[2] によるコード進行からジャズの即興演奏のメロディ生成[3] や Restricted Boltzmann Machine (RBM)[4, 5]を拡張した RNN-RBM を用いた音楽生成[6]が行われている.

これまでの音楽の自動生成の研究では音楽のデータが記号により表現されてきた. しかし記号による表現では楽器の音色や同じ楽譜でも演奏する楽器や演奏者によって印象が異なるというような人間の感じている音の微細な違いの表現は困難である. 元の音響信号を復元できるような特徴ベクトル列をある時刻の音響信号から得ることができれば, その特徴ベクトルの時系列を学習させることにより環境音,音声のような明確な楽音以外の音を含む音楽や記号で表現することが困難であるような特徴を含んだ音楽を自動生成することも期待できる.

本稿では深層学習が与えられたデータから多段な特徴抽出器を自己獲得し, かつ抽出して得た特徴ベクトルから元のデータを復元できるという特徴に着目する. これを構成する Restricted Boltzmann Machine(RBM)と Conditional RBM[7]によって音楽の音響信号をネットワークに学習させ, また音響信号を学習したネットワークが音楽を記憶することを示すことで, 深層学習によるネットワークが音響信号に復元可能な特徴ベクトルを学習することを示す.

2. 深層学習(Deep Learning)

Deep learning または深層学習と呼ばれる機械学習手法が大きく注目されている. これは多層のニューラルネットを良く学習させるための方法論である. 深層学習は深いネットワークを 1 層ごとに浅い層のネットワークとみなして教師なしの学習を行い, 学習後の層を積み重ねることで深いネットワークの全体の学習を実現する. このとき個別に学習した各層がそれぞれ特徴抽出器となることが知られている. 各層の学習は入力信号を良く復元する隠れ層のベクトルとネットワークのパラメータを探索するように行われるため, 学習後に得られる特徴ベクトルから入力データを復元することができる.

3. 深層学習による音響信号からの音楽の学習

3.1 ネットワークの構成

入力変数が実数値をとる Gaussian-Bernoulli RBM(図 1 中の左の図)[9]と RBM が時系列を学習できるように拡張された Conditional RBM(図 1 中の右の図)を積み重ねたネットワークに音楽の音響信号を学習させる. ネットワークは図 2 に示す 3 層の構造とし, 第一層を 16,000 ユニット,第 2 層を 200 ユニット,第 3 層を 200 ユニットとする. 第 2 層は RBM と Conditional RBM で共有され, 図中の past 層が示す層はある時刻における過去の第 2 層のベクトルを保持するヒストリーベクトルと呼ばれるベクトルであり, 第 2 層の出力値を 7 ステップ分保存する. 各ユニット数の設定は事前の実験により経験的に決定している.

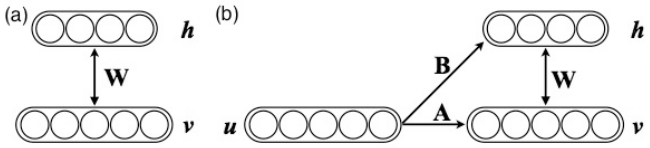

Fig. 1 (a) Restricted Boltzmann Machine (RBM) and (b) Conditional RBM

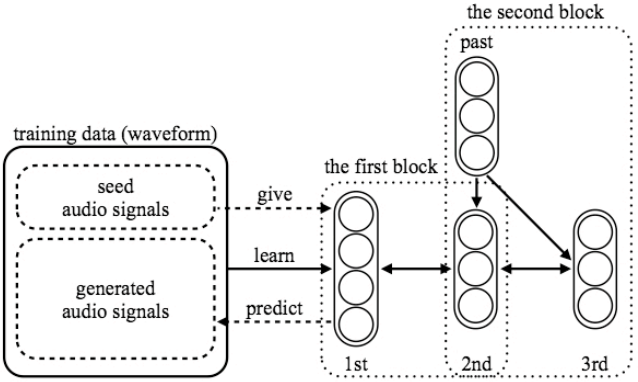

Fig. 2 Structure of network model for learning of music audio by Restricted Boltzmann Machine (RBM) and Conditional RBM

3.2 音響信号の前処理

学習用の音楽の音響信号は,まず 16kHz/16 bit のモノラル信号となるようにサンプリングする. このサンプリングされた信号は最小値-5 から最大値 5 の範囲となるように正規化する. 次に窓幅 16,000 サンプルで 16,000 サンプルごとにフレームに分割し,各フレームが 1 秒間の波形信号を表すデータを作成する. このとき時間方向は重複させない.

3.3 深層学習による音響信号からの音楽の学習

ネットワークに前処理を加えた音楽の音響信号を学習させ

る. RBM と Conditional RBM の学習パラメータは共通した値を設定し, 学習率 0.0001 として Contrastive Divergence (CD_k) 学習アルゴリズム[8]により各層 6000 エポックの学習を行う.

学習用のデータには市販されている音楽の音響信号から The Beatles の"Let It Be"を用意し, 0 秒から 30 秒までの間をネットワークに学習させ, 0 秒から 10 秒までを学習後のネットワークに与えて, その後続の音響信号をネットワークより得た. 元の音響信号とネットワークから得られた音響信号は図 3 に示すスペクトログラムとなり, 10 秒以降の音響信号をネットワークが想起できていることが確認された. ここで図 3 中の上の図が元の"Let it be"の音響信号,下の図がネットワークから得られた音響信号のスペクトログラムである.

ネットワークから得られた音響信号はスペクトログラム上では全ての周波数帯域でそのパワーが確認でき, 元の音響信号のスペクトログラムの判別が難しい. これは学習誤差などからくるノイズの影響と考えられ, 人間のテスターが対象の音響信号を聴取したときには, 元の音楽に白色雑音が含まれたように聞こえる. この雑音をよく除去することができれば, より元の音響信号に近い信号が得られる. 本稿ではノイズの除去法には触れないが, この音響信号に含まれる雑音の除去は, 深層学習による音響信号からの音楽の学習と自動生成において解決しなければならない問題の 1 つとなる.

深層学習により音楽の音響信号を学習したネットワークは学習した音楽の断片の音響信号を与えると, その後続の音響信号を想起することが確認された. これは音楽をネットワークが記憶していることを示し, また雑音が混ざっているが人間のテスターにとって元の音楽と認識できる音響を得られたことは, ネットワークが音響信号を復元できるような特徴ベクトルを抽出していることを示唆する.

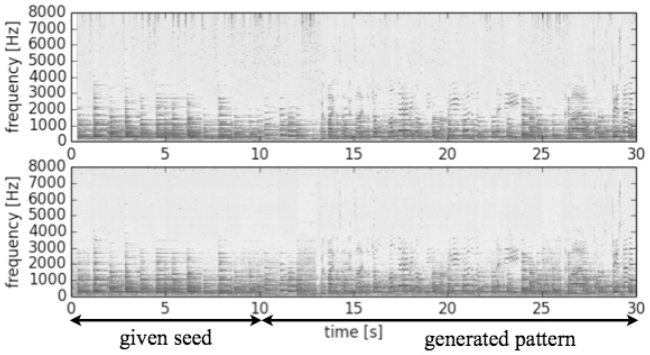

Fig. 3 Power spectrogram of the original music ("Let it be" by The Beatles), and the reconstructed music from predicted audio signal by network model. Upper image is spectrogram of the original audio signal (training data), and lower image is spectrogram of the reconstructed music from predicted audio signal by network model. Where parameters for Fast Fourier Transform (FFT): window size is 512 points, overlap size is 0 points, and window function is hamming window.

4. おわりに

これまで音楽の自動生成の研究では音楽のデータが記号によって表現されてきた. しかし, 記号による表現では音色や同じ楽譜でも演奏する楽器や人によって印象が異なるというような人間の感じている音の微細な違いの表現は困難である. ある時刻の音響信号から可逆変換のできる特徴ベクトル列を得ることができれば, その特徴ベクトルの時系列を学習させることにより環境音,音声のような明確な楽音以外の音を含む音楽や記号で表現することが困難であるような特徴を含んだ音楽を自動生成することも期待できる.

深層学習はネットワークの 1 層ごとを非常に浅いネットワークととらえてそれぞれを教師なしの学習を行うことで, 深いネットワークを良く学習させるための方法論である. この教師なし学習は入力層に与えられた入力信号を良く復元することのできるネットワークのパラメータを推定することで進められる. これにより得られた各層は学習用の入力データから自動的に特徴抽出器を獲得することが知られ,ネットワークの各層から得られる特徴ベクトルは元のデータに復元できる.

深層学習により音楽の音響信号を学習したネットワークは学習した音楽の断片の音響信号を与えると, その後続の音響信号を想起することが確認された. これは音楽をネットワークが記憶していることを示し, また雑音が混ざっているが人間のテスターにとって元の音楽と認識できる音響を得られたことは, ネットワークが音響信号を復元できるような特徴ベクトルを抽出していることを示唆する.

深層学習によるネットワークが記憶した音楽を加工したり, あるいは未学習の音楽の断片から人間にとって音楽に聞こえるような音のパターンを時系列に生成したりできれば音響信号を用いた音楽の自動生成システムの実現が期待できる. しかし, 本稿で述べた構造のネットワークにおいて未学習の音楽の断片からは人間のテスターが楽音として認識できる音は得られていない.

今後は深層学習の方法論により音楽を学習し記憶したネットワークにおいて, 未学習の音楽の断片から人間のテスターが楽音と認識できるような音の時系列の生成を試み,また学習後のネットワークから得られる音響信号に含まれる雑音を除去する方法論についても検討を行う.

文献

[1] Cope, D., "Computers and Music Style", Computer Music and Digital Audio Series, 1991.

[2] Bengio, Y., "Learning deep architectures for AI", Foundations and trends in Machine Learning, 2.1, 1-127, 2009.

[3] Bickerman, G., Bosley, S., Swire, P., and Keller, R. M., "Learning to Create Jazz Melodies Using Deep Belief Nets", First International Conference on Computational Creativity, 2010.

[4] Smolensky, P., "Information processing in dynamical systems: Foundations of harmony theory", Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol 1, pp.194-281, MIT Press, 1986.

[5] Hinton, G. E., "A Practical Guide to Training Restricted Boltzmann Machines", Tech. Rep. Department of Computer Science, University of Toronto, 2010.

[6] Lewandowski, N. B., Bengio, Y., and Vincent, P., "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription", 29th International Conference on Ma- chine Learning (ICML 2012), 2012.

[7] Taylor, G. W., and Hinton, G. E., "Factored conditional restricted Boltzmann machines for modeling motion style", Proceedings of the 26th annual international conference on machine learning, pp.1025-1032, 2009.

[8] Hinton, G. E., "Training products of experts by minimizing contrastive divergence", Neural computation, vol.14-8, pp.1771-1800, 2002.

[9] Cho, K. H., Ilin, A., Raiko, T., "Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines", Artificial Neural Networks and Machine Learning-ICANN 2011, pp.10-17, 2011.