

音楽情報処理 最前線!

■第15回

未来の音楽の楽しみ方、作り方はどう変わるのか？
コンピュータは音楽を理解できるようになるのか？
コンピュータを使って音楽を研究する「音楽情報処理」
という研究分野が、世界的に注目を集めています。
本連載では、そうした最先端の研究事例を紹介していきます。

ニコニコ動画のコメントは自動生成できるか？ 音楽にコメントするシステム MusicCommentator

人間は音楽を聞いて「このギターソロは神」とか「哀愁を誘うメロディだね」などと言語を使って音楽を表現することができる。SF映画のように、いつの日かコンピュータもこのような能力を獲得できるのだろうか？本稿では、そのような究極の目標への第一歩として開発したシステム MusicCommentator を紹介する。

コメント能力を科学する

動画共有サービスであるニコニコ動画では、投稿された動画に対して不特定多数のユーザがコメントを行い、コミュニケーションを楽しんでいる。このコメントするという行為は人間の極めて高度な能力であるのだが、我々が普段意識することは少ない。そこで、ニコニコ動画の人気コンテンツの一つである「演奏してみた」カテゴリ(音楽演奏の様子を録画した作品群)を例に、ユーザがどのようにコメントを行っているか見てみよう。動画を見た直感から、すぐさま「鳥肌立った」などとコメントする。音楽演奏をよく聴いて「いいアレンジだ」などと評価する。ブラスバンド動画を見て、あるユーザが「トランペットで参加したい」と言えば、「じゃ、俺はクラリネットで」と呼応する。楽曲「思い出は億千万

のサビでは、お約束としてみんなで「おっくせんまん!」を同じ箇所でも多数コメントして「弹幕」を作る。楽曲が思い出のゲームに使われていたら、悲しい曲調でもないのに「涙出てきた」とコメントする。このように、我々が音楽にコメントする内容は、音楽自体の内容だけにとどまらず、他のユーザのコメント、自らの人生経験、心理的・生理的反応、文化的背景などの様々な要因が複雑に絡み合っている。にもかかわらず、それらを意識することなく適切なコメントを生成できる人間の能力には驚くべきものがある。

では果たして、将来、コンピュータも人間と同等のコメント能力を獲得できるのだろうか。本研究ではその長い道のりの第一歩として、音楽内容とコメントとの表層的な対応付けに基づいてコメントを行うシステム MusicCommentator を開発した。

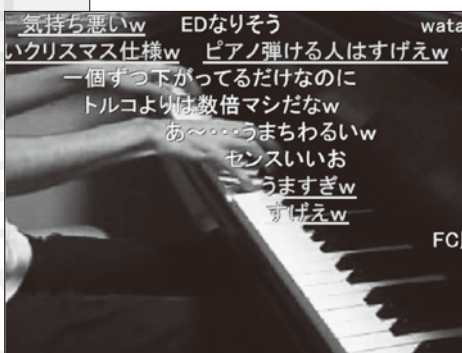
こんなコメントが生成された!

MusicCommentator は、「空気を読んで」音楽にコメントを行うシステムである。人間は新たにコメントしようと思ったとき、音楽内容だけでなく、他のユーザがどのようなコメントをしているかを参考にすることがある。すなわち、空気を読むわけだ。その場の流れにふさわしくないコメントをしてユーザ間のコミュニケーションを阻害してしまうことがないよう、MusicCommentator も音楽内容とユーザコメントの両方を考慮しながら新たなコメントを付与する。ユーザコメントがほとんどない場合でも、音楽内容だけからコメントすることもできる。

それでは、実際のコメント生成例を見ていこう。図1は、ニコニコ動画から取得した、タイトルに「弾いてみた」を含むある動画(sm1298190)に対してコメントを付与した結果である。事前に、他の「弾いてみた」動画75曲とそれに付与されたユーザコメントを多数集めて、システムに「空気の読み方」を覚えこませている(「学習」と呼ぶ)。覆面プレイヤーがギターで「となりのトトロ」を弾いているこの動画に対し、システムは「ギター教えてくれw」とか「なんという才能の無駄遣いw」などといったコメントを自動生成して付与した。このような表現は、他の類似した動画でユーザが用いていた表現から自動的に学習したものである。図2は、あるピアノ演奏動画(sm1594970)にコメントを付与した結果である。このプレイヤーは転調を繰り返しながらスーパーマリオのテーマ曲を弾いており、システムは「うますぎw」と「ピアノ弾ける人はすげえw」とを同時にコメントした。



◀図1 ギター演奏動画 sm1298190 に対するコメント生成例: 下線を引いたコメントがシステムによる自動生成。



◀図2 ピアノ演奏動画 sm1594970 に対するコメント生成例: 下線を引いたコメントがシステムによる自動生成。

音楽と単語の対応付けを学習する

では、どうやってコメントを自動生成しているのだろうか？図3をみながら、人間にはなぜコメントできるのかを考えてみよう。音楽も言語も知らない赤ん坊には無理である。しかし、成長するにつれ言語を覚え、他人と「この曲の○○は××だったね」な

吉井 和佳
(よしい かずよし)

2008年 京都大学大学院情報学研究所 博士後期課程修了。博士(情報学)。現在、産業技術総合研究所 情報技術研究部門 メディアインタラクション研究グループ研究員。機械学習手法を用いて音楽推薦や可視化などの研究を進めている。

後藤 真孝
(ごとう まさたか)

1998年 早稲田大学大学院理工学研究所 博士後期課程修了。博士(工学)。現在、産業技術総合研究所 情報技術研究部門 メディアインタラクション研究グループ長。2007年～2008年度 音楽情報科学研究会主査。計算機によって実世界の音楽・音声コンテンツを自在に扱える技術の確立を目指し、音楽・音声の音響信号の自動理解と、それに基づくユーザインタフェースの研究を中心に、様々な研究課題に取り組んでいる。

「音楽情報科学研究会」へ参加してみませんか？

情報処理学会 音楽情報科学研究会(SIGMUS)は、コンピュータと音楽とが関わり合うあらゆる場面を活動対象とする学際的な研究会で、年5回の研究発表会を開催しています。研究会に会員登録すると、研究発表会の参加費が無料になるだけでなく、過去の全研究発表会の論文のダウンロードなどの特典があります。研究会の登録方法や研究発表会の開催に関する最新情報などは <http://www.sigmus.jp/> をご覧ください。

どとコミュニケーションを行う。これを繰り返すと、知らず知らずのうちに他人が利用した表現を取り込んで自らの表現が豊かになってくる。その結果、適切にコメントできるようになるのである。ということは、システムにも「学習」と「運用」の機構を実装すればいいはずだ！

機械学習(マシンラーニング)とは、そのような機構を実現するための汎用的で強力な枠組みである。例えば、スパムメール判別や機械翻訳などに広く利用されている。MusicCommentatorの概略を図4に示す。本システムは、多数の楽曲の音楽音響信号とそれらに対する時刻付きユーザコメントを利用して、どのような音楽内容に対し、どのような単語が、どのくらい使われやすいかのモデル^{注1}を構築する。モデルとは、人間がコメントを行うときの感覚に相当する。例えば、ポピュラー音楽中のギターソロの区間では、当然「ギター」「ソロ」「弾く」といった単語が使われやすくなる。一方、「うまい」や「すごい」といった単語は、動画内の時刻にかかわらず普遍的に使われるだろう。ある音楽内容に対する各単語の使われやすさは確率値で表現できて、与えられた学習用データに最もよく合致するように自動的に調整を行う。すなわち、システムは学習を通じて音楽内容と単語とを対応付ける感覚を磨くのである。

ここで問題なのは、ある音楽内容を表現するのに適切な単語が分かって、それらをうまくつなげて文章にしなければならないことだ。人間が助詞や助動詞などを適切に使用することができるのは、単語の並びに関する感覚を身につけているからである。ということは、単語の並びモデル^{注2}を構築する必要ができて、これもやはり大量のユーザコメントから機械学習することになる。

システムは結局、音楽内容と単語の対応付けモデルと単語の並びモデルとを併用してコメントを生成しているのである。人間だと、文法的に正しい文を生成しようとすると、本当は使いたい単語を使いにくい場合もある。システムも、文法と内容の整合性がとれるように調整を行い、最終的なコメントを生成する。

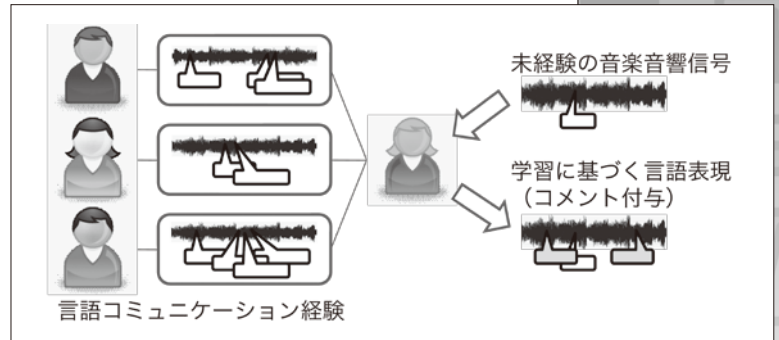
言語を操るコンピュータ

音楽に限らず、自らが体験したことの中身を言葉で表現することは、人間を人間たらしめている重要な能力である。言葉があったからこそ、人間は互いに自らの体験を伝達しあい、進化を果たすことができた。我々は音楽をテーマに取り上げ、音楽がどのような仕組みでどのように言語表現に「変換する」ことができるのかに関して計算モデルを構築した。現状では表層的な対応付け関係を学習しているにすぎないが、これが言語表現として音楽を「理解する」ことが可能な計算モデルを確立するための重要な一歩になれば、と考えている。

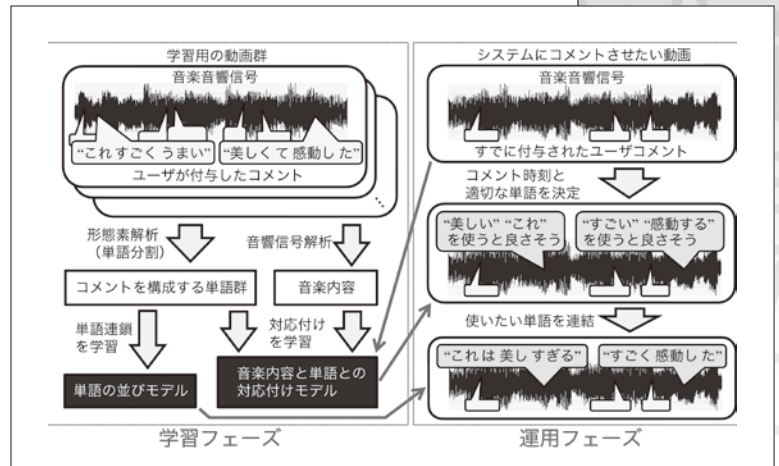
今後は、コンピュータによるメロディ・リズム・ハーモニーなどに関する多角的で高度な音楽理解を可能にするため、音楽解析技術を改善していきたい。同時に、機械学習の手法を洗練することで、人間の学習機構に近い処理が可能になることを期待している。将来的には、我々が一生かけても視聴できないほどの動画をコンピュータが代わりに視聴してくれて、言語で語り合いながら検索や推薦を行ってくれるような時代がやってきて欲しい。

注1：具体的には隠れマルコフモデルを用いる
注2：具体的にはn-gram言語モデルを用いる

▼参考文献：
吉井 和佳, 後藤 真孝: "MusicCommentator: 音楽に同期したコメントを自動生成するシステム", 情報処理学会 音楽情報科学研究会 研究報告, Vol. 2009-MUS-81, No. 20, pp. 1-6, July 2009.



▲図3 言語表現の獲得：他人と言語を使ってコミュニケーションを行う中で、音楽の表現方法を学習し、自分なりにコメントできるようになる。



▲図4 MusicCommentatorの概要：学習フェーズでは、音楽内容と単語の対応付けモデルと単語の並びモデルを構築する。運用フェーズでは、得られたモデルを用いて適切な単語をつなぎ合わせて適切な時刻にコメントを付与する。