

後藤 真孝

(独)産業技術総合研究所 情報技術研究部門・上席研究員

コンテンツ共生社会のための類似度を可視化する情報環境の実現

§1. 研究実施体制

(1)「後藤」グループ

- ① 研究代表者:後藤 真孝 ((独)産業技術総合研究所 情報技術研究部門、上席研究員)
- ② 研究項目
 - ・音楽を中心とした類似度可視化情報環境の実現と全体統括

(2)「森島」グループ

- ① 主たる共同研究者:森島 繁生 (早稲田大学 理工学術院、教授)
- ② 研究項目
 - ・音楽連動動画を中心とした類似度可視化情報環境の実現

(3)「中村」グループ

- ① 主たる共同研究者:中村 聡史 (京都大学大学院 情報学研究科、特定准教授)
- ② 研究項目
 - ・Web インタラクションを中心とした類似度可視化情報環境の実現

§ 2. 研究実施内容

■ 研究のねらい

本研究は、音楽や動画のようなメディアコンテンツを豊かで健全に創作・利用する「コンテンツ共生社会」の実現に向けて、音楽を中心とした研究開発によって、膨大なメディアコンテンツ間の類似度を人々が知ることができる(可知化する)情報環境のための技術基盤を構築する(図1)。そして、創作支援技術と鑑賞支援技術を研究開発することで、コンテンツの創作や鑑賞を誰もが能動的に楽しめる社会や、過去のコンテンツに敬意を払う文化、感動体験重視型のコンテンツ文化の実現を目指す。具体的には、メディアコンテンツとして代表的で重要性が高い音楽あるいは音楽連動動画(ミュージックビデオやダンス動画等)等を中心に、研究代表者らがこれまでに研究開発してきた音楽自動理解技術を応用し発展させつつ、人間と情報環境の調和的な相互作用を引き出せるよう、ユーザを巻き込んだ研究開発に取り組む。さらに、人間とコンテンツとの関係、過去のコンテンツと未来のコンテンツとの関係が豊かで持続発展可能な共生関係となる「コンテンツ共生社会」を実現するために、類似度を可知化する情報環境を活用した「コンテンツ生態系」という新たな概念を提唱し、メディアコンテンツの創作と鑑賞を総合的に支援する「コンテンツ生態系技術」を開発する。この「コンテンツ生態系技術」では、過去のコンテンツが新たなコンテンツを育む「コンテンツ循環」をユーザと共に持続発展させることを狙う。そのために Web 上で公開されている音楽関連コンテンツを主な対象とした「コンテンツ生態系 Web サービス」を構築・公開して実証実験する。

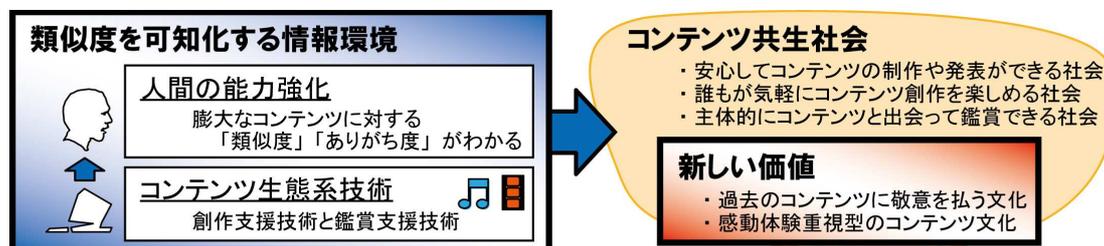


図1. 「コンテンツ共生社会」のための「類似度を可知化する(知ることができる)情報環境」

■ 研究進捗状況と今後の見通し

平成 23 年度は本研究プロジェクトの初年度として、研究成果を創出するだけでなく、後藤グループ、森島グループ、中村グループのメンバを含めた関係者全員に、プロジェクトの狙いや、具体的な成果目標を周知して共有することが予定通り完了し、グループ間連携のための共通計算機環境等の整備を進めた。また、本プロジェクトの対外的な名称を、音楽に関連して音と画像を共に扱うことから「OngaCREST Project (音画 CREST (おんがくれすと) プロジェクト)」と命名し、<http://ongacrest.jp> を取得して広報準備を進めた。平成 24 年度以降は、その環境の上で具体的な連携を開始し、基礎研究の成果を蓄積しつつ、応用研究としてコンテンツ生態系 Web サービスのプロトタイプシステムを公開して対外的に見える形で研究成果を創出しながら、研究を進めていく。

■研究成果

以下、平成 23 年度の代表的な研究成果を述べる。

・ 能動的音楽鑑賞サービス「Songle」のベータ版プロトタイプシステムの実現と試験公開

「コンテンツ生態系 Web サービス」の鑑賞支援技術の実現に最優先で取り組み、音楽音響信号理解技術によって音楽の聴き方をより豊かにするための能動的音楽鑑賞サービス「Songle (ソングル)」(<http://songle.jp>) のプロトタイプシステムを実現した。従来、研究開発段階の音楽インタフェースや技術を、日常生活で人々に使ってもらうのは容易でなかった。Songle では、Web 上で人々に能動的音楽鑑賞インタフェースを体験してもらうことで、音楽鑑賞がより能動的で豊かになる質的な変化を日常生活で起こすことを目指している。我々がこれまでに培ってきた音楽自動理解技術を応用し、Web 上で公開されている任意の楽曲(MP3形式の音響信号ファイル)中の音楽情景記述(音楽要素)を推定する。Songle のユーザは、その結果が可視化された様子を見ながら、楽曲の再生を楽しむことができる。現在の実装では、(i) 楽曲構造(サビ区間と繰り返し区間)、(ii) 階層的なビート構造(拍と小節の先頭)、(iii) メロディライン(メロディの歌声の基本周波数(F0))、(iv) コード(根音とコードタイプ(構成音))の四つの代表的な音楽情景記述を自動推定し、可視化して音楽内容に基づくブラウジングを可能にした。さらに自動推定では誤りが不可避であるため、効率的な誤り訂正インタフェースを Web 上で提供した。これを Web 上のサービスとして実装して、ベータ版の試験公開を開始している。以上の成果を、情報処理学会 査読付きシンポジウム「インタラクション 2012」に投稿したところ採択され、ベストペーパー賞を受賞した¹⁾。また、査読付き国際ワークショップ「CrowdSearch 2012」(WWW 2012 併設)にも投稿し、採択された²⁾。

Songle (<http://songle.jp>)

□ 産総研独自の音楽理解技術を使用

- ・ 4種類の音楽的要素を推定して表示

表示された「音楽地図」を見ながら
繰り返しや音楽的要素に気づくことができる

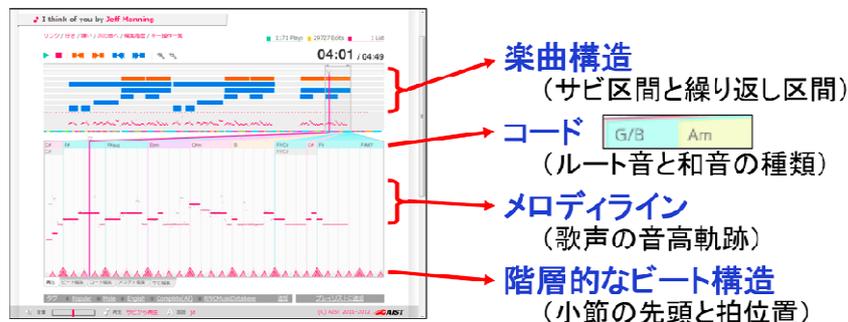


図2. 能動的音楽鑑賞サービス「Songle (ソングル)」のプロトタイプシステム

・音楽連動動画の類似度計算・登場人物探索の検討とプロトタイプシステム構築

動画共有サイト上に存在する音楽連動動画を主な処理対象として、音楽連動動画の類似度計算に向けての基礎検討を行なった。画像の色彩、構図、登場人物に着目した類似画像検索システムのプロトタイプシステムを構築し、性能評価を行なった。画像の色彩とは、画面全体から受ける色あいの印象であり、第一の試みとして色ヒストグラムを用いることで色彩の印象の類似した画像がおおよそ検索可能であることが分かった。構図とは、画面に写っている登場人物やオブジェクトなどの形状の印象を意味し、最初の試みとして高速でノイズに頑強な SURF 特徴量に対して k-means クラスタリングを施すことにより低次元のパラメータを抽出し、評価実験により、おおよ構図の近い画像を選択できる可能性があることを確認した。同一の登場人物に関しては、HOG 特徴量を利用し、アノテーションが予め付与された人物が描画された静止画像をクエリとして類似性検索することにより、動画の 1 フレームの画面内において、注目キャラクタの有無およびその存在位置を認識できることが分かった。次に、音楽連動動画中の登場人物探索の試みとして、注目する人物の顔画像に着目し、その人物の登場するフレームを時間連続体として処理対象とすることで、顔認識アルゴリズムを動画のフレーム単位で適用する場合と比較して、顔向きや遮蔽に対する頑健性を大幅に向上できることを確認した。

・ソーシャルアノテーション収集・管理システムの研究開発

ソーシャルアノテーションを効率的に収集するクローラおよび収集したアノテーション情報を蓄積する管理システムを開発し、蓄積コンテンツからの簡易的な検索を可能とする検索システムのプロトタイプ版を開発した。動画共有サイトに投稿されている動画に関連したソーシャルアノテーション収集のためのクローラおよびデータベースの開発および整備を行うとともに、実際のクローリングを開始した。また、収集したソーシャルアノテーションを様々な観点で取得できる内部向け Web API を整備した。さらに、我々がこれまでに培ってきた動画に対するコメントからの主観的な評価抽出手法を本プロジェクト用に改良し、今回開発したクローラで収集したデータに適用した。動画の各再生時刻におけるそれぞれの主観的評価量を返す Web API の開発も開始した。

§3. 成果発表等

(3-1) 原著論文発表

●論文詳細情報

1. 後藤 真孝, 吉井 和佳, 藤原 弘将, Matthias Mauch, 中野 倫靖: "Songle: ユーザが誤り訂正により貢献可能な能動的音楽鑑賞サービス", 情報処理学会 インタラクション 2012 論文集, pp.1-8, March 2012. (査読付きシンポジウム Proceedings、ベストペーパー賞受賞)
2. Masataka Goto, Jun Ogata, Kazuyoshi Yoshii, Hiromasa Fujihara, Matthias Mauch, and Tomoyasu Nakano: PodCastle and Songle: Crowdsourcing-Based Web Services for Retrieval and Browsing of Speech and Music Content, Proceedings of the First International Workshop on Crowdsourcing Web Search (CrowdSearch 2012), April 2012. (in press)
3. Kazutoshi Umemoto, Takehiro Yamamoto, Satoshi Nakamura, Katsumi Tanaka, Search Intent Estimation from User's Eye Movements for Supporting Information Seeking, Proceedings of the International Conference on Advanced Visual Interfaces (ACM AVI 2012), May 2012. (in press)