



平成12年5月31日

インターネット上の音や映像を瞬時に探索する技術を開発

－圧縮・編集したデータも瞬時に探索可能な"学習アクティブ探索法"－

NTTは、インターネット上などに流通している圧縮等により品質が変化した音や映像データの中から、特定の音や映像を瞬時に探索する技術"学習アクティブ探索法"を開発しました。

"学習アクティブ探索法"は、98年10月にNTTが開発した高速探索技術である"時系列アクティブ探索法"を発展させ、従来の技術に加え、新たに学習機能を導入したものです。これにより編集や圧縮などによって品質が変化（低下）した音や映像に対しても正しく高速探索することが可能になりました。本技術を用いると、24時間分のデータ（インターネット上の30秒の音や映像を約3000件分に相当）の中から目的とする音や映像を約1秒で探索することが可能となります。

本技術により、インターネット上の音楽や映像の著作権管理システムや、音で探せる音楽情報検索サービスなどに応用することが可能となります。今後は本技術を核として、音、映像、文書などの膨大なメディア情報から興味ある情報をすばやく探索する技術の実現をめざすと共に、さらに幅広い応用を検討していきます。

○開発の背景

マルチメディア技術の普及によって、身のまわりにはさまざまな音や映像のコンテンツが氾濫するようになりました。そのため、視聴を希望する特定の音や映像が、膨大な情報のどこに存在しているかを探し出すのが困難になっています。また、自分が作った音や映像等のコンテンツの不正使用をチェックすることが難しいという著作権上の問題も深刻になってきています。

NTTではこれらの課題に対して、ヒストグラム照合の技術により、従来の"特徴ずらし照合法（注1）"に比べ約600倍の高速探索を可能とした"時系列アクティブ探索法（注2）"を開発し、テレビ放送信号の探索に適用してまいりました。

しかしヒストグラム照合の技術だけでは、テレビ放送などのように品質変動の少ない音や映像の探索には適用できるのですが、インターネット上に見られる圧縮（注3）や編集などによって品質が変化した場合には、探索精度

が低下するという問題点がありました。

○技術のポイント

"学習アクティブ探索法"は、蓄積した長時間におよぶ大量の音や映像（蓄積信号）の一件ごとの各時点に対して、順次、目的の音や映像（目的信号）と照合し、目的信号と類似した音や映像（音や映像同士の類似度が一定値を越える箇所）がどの蓄積信号のどの時点にあるかを高速に探す技術です（[図1](#)）。

既存の技術では、圧縮や編集などによって品質が変化した音や映像については、探索精度が低下するという問題がありました。本技術は、圧縮や編集などで生じる品質の変化などに対応して特徴がどのように変化するかを特徴変換テーブルの形で事前に学習しておきます。さまざまな品質変化の種類に対応した複数の特徴変換テーブルで複数の特徴を生成して探索するため、品質が変化した場合にも高い精度を維持できます。新たな要因による品質変化への対処が必要になった場合にも、追加学習を行うことで柔軟に対応できます。

また、精度を犠牲にしない高速探索を、ヒストグラム照合の技術により実現しており、一般的なパソコンで24時間分の音や映像の中から、目的とする部分を約1秒で探索することが可能です。

○探索手法

"学習アクティブ探索法"は、学習段階と探索段階という2つの段階で構成されています。

（1）学習段階（[図2\(a\)](#)）

探索の対象とする音や映像の品質の変化の仕方を事前に学習し特徴変換テーブルを作成する段階です。例えば、インターネット上の音や映像の編集過程で用いられているMPEG（[注4](#)）や毎秒のコマ数を減らす（コマ落ち）等の圧縮処理について、その編集や圧縮の前後で音や映像の特徴がどのように変化するかを学習します。このような学習は、学習用のデータ（例えば編集処理の前と後の音や映像データ）から自動的に行われます。

（2）探索段階（[図2\(b\)](#)）

探索段階は、さらに3つのステップから構成されます。

◆第<1>ステップ（特徴抽出）

目的信号と蓄積信号の双方から、瞬時瞬時における音や映像の特徴を抽出します。

◆第<2>ステップ（ヒストグラム作成）

第<1>ステップで抽出した特徴からヒストグラム（[注5](#)）を作成します。ヒストグラムは目的信号と蓄積信号の双方について作成されますが、目的信号に対しては、学習段階で学習した特徴変換テーブルによってMPEG

やコマ落ちなどの品質の変化を考えた複数のヒストグラムを作成します。蓄積信号に対しては特徴からそのままヒストグラムを作成します。

◆第<3>ステップ（高速照合）

目的信号と蓄積信号のヒストグラムを照合することにより、蓄積信号のどの部分が目的信号と一致しているかを検出します。このとき、ヒストグラムの性質を利用することにより、探索もれが起きないことを保証したまま照合が不要なヒストグラムを瞬時に判別してその部分の照合計算を省くことにより、高速照合を達成しています。

品質が変化した音（音楽CD）や映像（テレビ放送）のそれぞれ24時間分から10秒の信号の断片を探索するという課題について、"学習アクティブ探索法"を評価しました。ヒストグラム照合の技術だけでは、低ビットレートのMPEGなどで極端に圧縮した映像や、家庭用ビデオでダビングを繰り返した映像、電話で送った音響信号などの品質の変化した音や映像を探索する場合には探索精度の低下が避けられませんでした。学習の技術を導入した"学習アクティブ探索法"では探索時間が1秒という高速性を保ったまま、高い精度が達成できることが実証されました（[図3](#)）。

○今後の展開

"学習アクティブ探索法"は、インターネット上の音楽・映像の著作権管理システム（[図4](#)）や、音楽で探せる音楽情報検索サービスなど、様々な応用が期待できます。

NTTでは"学習アクティブ探索法"を用いて各種応用システムを開発するとともに、この技術を核として、音、映像、文書などの膨大なメディア情報から興味ある情報をすばやく探索する技術の早期商用化をめざします。

○用語解説

（注1）特徴ずらし照合法

音や映像の探索で従来用いられていた方法で、音の波形や映像から計算される特徴ベクトルどうしをずらしながら照合する探索手法です。膨大な計算が必要になるという欠点があり、例えば24時間の音から10秒の区間を探索するのに、一般的なパソコンで約10分程度かかります。これまで、照合の仕方や照合箇所をずらし方などを簡略化することで高速化する手法もありましたが、その場合には探索もれなど探索精度の低下が避けられませんでした。

（注2）時系列アクティブ探索法

当社が98年10月に開発した、音や映像の高速探索技術です。ヒストグラム照合の技術を用いることにより、特徴ずらし照合法に比べ約600倍の高速探索を実現しました。テレビ放送などのように音や映像の品質が保たれ

ている場合には極めて有効性の高い技術ですが、極端に品質が変化した音や映像の探索は不可能でした。

(注3) 圧縮

音や映像を表現するデータ量を削減することをいいます。

(注4) M P E G

音や映像の圧縮方法を定めた標準規格の一つです。M P E G 1, M P E G 2などの種類があり、音楽データの圧縮によく用いられているM P 3もM P E Gの一つです。

(注5) 特徴のヒストグラム

信号の瞬時瞬時から得られた特徴をその性質に応じて複数種類に分類し、各分類ごとに特徴の発生回数を数えて度数分布グラフにしたものです。

- ・ [図1 音や映像の探索の概念図](#)
- ・ [図2\(a\) 学習アクティブ探索法のしくみ \(学習段階\)](#)
- ・ [図2\(b\) 学習アクティブ探索法のしくみ \(探索段階\)](#)
- ・ [図3 探索時間と探索精度](#)
- ・ [図4 インターネット上の音楽・映像の著作権管理システムの構成例](#)

<本件に関するお問い合わせ先>

N T T先端技術総合研究所

企画部 真鍋、活田、佐々木

Tel: (046) 240 5152, Fax (046) 270 2365

E-mail: st-josen@tamail.rdc.ntt.co.jp



[NTT NEWS RELEASE](#)