

「多様なメディアからの復興情報抽出に関する研究」

伊藤慶明 (ソフトウェア情報学部 教授)、小嶋和徳 (ソフトウェア情報学部 講師)

<要旨>

本研究は、災害時にテレビやラジオから流れる様々な情報から必要な情報を検出・抽出することを目的とするものであり、平成 28 年度は以下の研究成果を得ることができた。音声情報からの情報抽出では、音声検索において深層学習をはじめとする様々な検索手法、および単語認識や音素認識などの情報を統合する手法を検討した、また一度検索した情報を再利用する手法も検討し、より効率よく必要な情報を獲得できる様にした。また画像からの情報抽出では、情景画像からの文字抽出において画像の特徴に合わせて前処理を変更することで、情報の取りこぼしを少なくする手法について検討した。

1 研究の概要

本研究は、災害時においてテレビやラジオなどから流れる様々な音声情報や映像情報から、必要な情報を抽出することを目的とするものであり、大きく音声情報からの抽出と映像 (画像) 情報からの抽出に分かれる。

音声情報からの抽出では、音声中の検索語検出の研究を中心に、より効率よく、より高速に検索する手法について検討している。また、映像 (画像) 情報からの抽出では、情景画像から文字部分を抽出する研究において抽出精度を向上させる手法を検討している。

2 研究の内容

音声情報からの抽出においては、以下の手法の検討を行った。(1) 音声クエリと音声ドキュメントを認識する際の単語認識および音素認識を統合する手法の検討、(2) 検索した結果の一位候補区間を新たにクエリとして再検索し、もとの検索結果と統合する手法の検討、(3) Monophone モデルの改良モデルの検討、(4) 深層学習 (DNN) の出力である Posteriorgram による照合における、Sparse Vector を用いた検索の高速化の検討、(5) DNN を用いた方法において、(5-1) サブワード間照合、状態間照合の音響距離を、DNN を用いることで高精度化する方法、(5-2) 音声クエリによる検索 (SQ-STD) においてフレームレベルでの連続動的計画法照合により、サブワード間照合、状態間照合よりも精緻な照合を行う方法、(5-3) SQ-STD において、DNN に加え音素や文字を出力とする Connectionist Temporal Classification (CTC) を照合に用いることで高精度な検索を行う方法の検討。

また、映像 (画像) 情報からの文字抽出では、画像から抽出できる高周波情報に注目し、ノイズが多い、またはぼやけた画像の特徴に注目し、画像によって処理を変更する手法について検討した。

3 これまで得られた研究の成果

音声情報からの抽出においては、以下の様な成果が得られた。(1) では単語認識結果と音素認識結果を統合する

際に照合結果を比較し、類似度が高い結果を重視し統合を行うことで検索精度が向上することが確認できた。(2) では最初の検索結果と一位候補をクエリとした場合の検索結果を統合することで、検索精度が向上することが確認できた。(3) では長母音を短母音二つで表現する方式、子音と拗音の連結を拗音とその後ろに続く母音を連結する方式を検討した結果、検索精度が向上することが確認できた。またここでは、DNN を利用する場合に情報を削減しない方が、検索精度が良いという見識が得られた。(4) では DNN で出力される Posteriorgram を使って照合を行う際、Sparse Vector を利用して計算の省力化を行うことで、検索精度を落とさずに計算時間が短縮することが確認できた。(5-1) ではサブワード間の照合および状態間の照合の音響距離を DNN で求めることでこれまでの手法に比べ検索精度が向上することが確認できた。また(5-2) では従来法である状態間レベルに比べて精緻なフレームレベルでの照合を行うことにより、検索精度が向上することが確認できた。さらに(5-3) では音声クエリによる検索において、DNN だけではなく音素や文字を出力する CTC を照合に用い、スコアを統合することで、検索精度が向上することが確認できた。

また映像 (画像) 情報からの抽出では、高周波情報が多い、つまりノイズが多い画像に対してはノイズ除去処理、高周波情報が少ない、つまりぼやけた画像に対してはエッジ強調処理を行うことで、従来手法に比べ再現率がかなり向上し、取りこぼしを防ぐのにかなり有効であることが確認できた。

4 今後の具体的な展開

音声情報からの抽出においては、DNN を用いた方法について、より検索精度が向上する手法を検討するとともに、検索速度を上げる手法についても検討を行う。さらには従来手法と提案手法を組み合わせることで、より有効な手法についても検討を行う。

映像 (画像) 情報からの抽出では、再現率を向上させたことで適合率が下がることが確認できており、今後は抽出された情報の中で明らかに文字ではないと分かるものを削除するなど、適合率も上げる方法について検討する。