

画像解析と音声認識を利用した授業動画の自動チャプタ分割

龍谷大学理工学部数理情報学科

T180022 北村優詩

指導教員 佐野 彰

概要

新型コロナウイルスは私たち学生の生活に様々な悪影響を与えた。今後も新たな変異株の影響から、更なる負担を及ぼすことが懸念される。一方で、コロナ禍ではオンライン授業という画期的な授業形式を生み出された。人が密集せずコロナ禍でも教育が継続できることや移動コストの削減等、対面授業にはないメリットが多く存在する。

本研究では授業動画を自動でチャプタ分割するシステムを作成した。本研究には2つの目的がある。1つは、コロナ禍の学校教育で大量にストックされていく授業動画を用いた復習や閲覧をスムーズに行えるようにすること。もう1つは、コロナ収束後もこれらの授業動画を要所で活用することでオンライン授業の質を高め、推進することである。

本研究では、自動チャプタ分割に画像解析のみを利用したもの、音声認識のみを利用したもの、画像解析と音声認識の両方を利用したものの3種類のシステムをPythonを用いて作成した。画像解析を用いた方法では、OpenCVを活用した平均画素法という画像縮小処理を施し、映像の細かな動きに妨げられない工夫や、動画のシーンに合わせて閾値が変動するような処理を行い、授業動画からチャプタ分割すべきカット位置を検出している。音声認識を用いた方法では、inaSpeechSegmenterという音声区間検出ライブラリで無声区間を抽出し、それぞれの区間の長さが動画の分割点として適切かを判定している。さらに、作成されるチャプタへの情報付与の手段としてSpeechRecognitionとMeCabを利用した音声認識からテキスト情報の抽出を試みた。

結果として、画像解析のみでは過剰な数の検出をしてしまい、誤検出も多かった。音声解析のみでは、検出すべき場所が比較的長い無声区間に含まれることがわかったが、長さというデータだけでは判定が難しく、検出漏れが高頻度で確認された。画像解析でカット検出した位置が動画の分割点として適切かどうかを音声区間検出によって判定する手法では、誤検出の発生頻度に改善がみられた。誤検出や未検出が僅かに発生するため手動で作成するチャプタには劣るが、簡易的なチャプタとしての活用は見込める精度である。一方で、音声認識によるテキスト情報の抽出では、単語の出現数とその単語の重要性にある相関関係が弱く、チャプタへの適切な情報付与が行えなかった。

今後の課題として、自動チャプタ分割では、さらに精度を高めるために重要な検出のみに絞る工夫が必要である。チャプタへの情報付与では、音声認識ではなく、映像からテキスト情報を読み込むなどの異なる手法を取り入れる必要がある。