

## 日本のテレビ放送における番組のシリーズを同定する手法の開発

藤岡 雄大

日本では現在、テレビ番組の内容や性格、放送形態、視聴形態それぞれについて多様化している。そして、近年においても、番組推薦手法などテレビ番組に関する様々な研究が行われている。このような番組推薦や、番組録画予約を実現するうえで、放送番組を同定することは欠かせない。しかし、放送番組の同定において基本となる番組表は、同定に適したきれいなデータではない。このテレビ番組の同定に関しては、これまでに複数の放送局間での放送に着目し、機械学習を用いてシリーズ番組の同定を行い、新番組の判定の難しさや機械学習の限界を示した例がある。

そこで、本研究は放送番組のシリーズに着目し、URL などより幅広い観点を比較検討し、同定に有効な手法を示すことを目指す。この研究を行うことによって、将来のテレビ番組推薦システムや録画予約システムなど、番組表やテレビコンテンツのメタデータを扱うシステムへの応用が期待できる。本研究では、SuperWork モデルの概念における Work に相当するものを「連続シリーズ」と定義し、この同定を行う。

使用する番組情報は「テレビ王国」の番組表を用いた。このうち、2021年9月から10月にかけての35日分について、東京と千葉の地上波局を対象に評価用データセットを構築し、それぞれの手法について検証を行った。提案手法として、まずタイトルなどのフィールドごとにいくつかの手法を提案し、そしてこれらの手法を組み合わせた手法 X と手法 Y を提案した。手法 X は、区切り文字を予め設定しタイトルから抽出したマスタータイトルの完全一致と番組詳細から URL を抽出し、双方を OR 条件で用いた。一方、手法 Y は、タイトルに対しジャロ・ウインクラー距離を求め、また番組詳細の URL を加味した類似度を用いた。

そして、同じ対象期間で正解データを作成し、その正解データと提案手法による同定結果を比較して、シリーズ数と放送回ペアの精度、再現率を指標に評価を行った。その結果、手法 X は、精度 86.7%、再現率 86.7%、手法 Y は精度 85.8%、再現率 66.4% となり、手法 X のほうが精度と再現率がともに優れていることが明らかとなった。また、この手法 X を改良するため、タイトルに対する手法として有望であると示されたジャロ・ウインクラー距離を、URL の同定に組み込むことにより、手法 X より再現率を約 0.03 ポイントの低下に抑えつつ、精度を約 0.2 ポイント向上できた。よって URL による同定の条件について、ジャロ・ウインクラー距離を加えることで精度向上に有効な可能性が期待できることが分かった。

今後は、他の番組群を対象としたデータセットによる検証や電子番組表以外のリソースと組み合わせた同定も検討の余地がある。また、テレビ番組とそれ以外の資源を結びつけることへの応用の可能性があり、利用用途により精度と再現率の優先度は異なるため、閾値によりどのように変化するか検討することが必要であることを課題としたい。

(指導教員 高久雅生)