

E-021

映像コンテンツ理解支援のためのキーワード表示方式の検討
Investigation of keywords presentation method
easy to view and helpful to understand video content

岡本 昌直[†] 祖父江 美香[‡] 山本 けい子[‡] 田村 哲嗣[‡] 速水 悟[‡]
Masanao Okamoto Mika Sobue Keiko Yamamoto Satoshi Tamura Satoru Hayamizu

1. はじめに

近年、インターネット、情報処理技術の発展によりさまざまな映像コンテンツが普及している。そのため、映像コンテンツをリアルタイムで理解することを支援するシステムの開発が望まれている。その代表例として、字幕が挙げられる。これは映像コンテンツ内で発話された内容を、音声認識などにより書き起こし、発話にあわせて表示するものである。字幕は聴覚障がい者にとって重要な情報源であり、健常者にとっても、必要な情報がどうかすぐ判断できるといった観点からも重要である。字幕の問題点として、人間の読み取り速度の問題が挙げられる[1]。そこで、ユーザの理解を支援する研究として、字幕の要約結果のリアルタイム提示[2]、会議議事録のように話者が複数存在する際の字幕についての検討がなされている[3]。

一方で、書き起こし文からキーワードを抽出する手法も考えられる。キーワードは、映像コンテンツの内容を端的に表しているため、効果的にユーザに内容を伝えることが可能となる。そこで、キーワードの抽出方法、および、字幕・キーワードの出力表示方式が重要となる。

これまでに、TF-IDF など複数の特徴量を用いた、線形回帰によるキーワード自動抽出を行っている。字幕・キーワードの表示方式としては、図1に示すL字型がユーザの理解を支援するうえで有効であると報告した[4]。以上をふまえて、本研究ではジャンルにおける特徴量の重要度ベクトルの変化を調査し、最適なキーワードの文字数、キーワード数を検討する。加えて、キーワードのみ表示の条件下における被験者実験を行う。

2. キーワード自動抽出

2.1 重回帰分析による重要度推定

キーワードの重要度 \hat{y} を式(1)のように推定する。学習用テキストに対し、人手で各名詞 x の重要度を求める。これを正解とし、重回帰分析を用いて重要度ベクトル ω を推定した。

$$\hat{y}_i = \sum \omega_j x_{ij} + C \quad (1)$$

素性ベクトル x には、TF、IDF、TF-IDF、接続名詞数、係り受けされている節数、固有名詞の6種類を用いた[4]。

2.2 ジャンルにおける重要度ベクトルの変化

経済、社会、政治、健康の4ジャンルにて重要度ベクトルの変化について実験を行った。実験には、経済、社会、政治のジャンルはNHK時論公論、健康のジャンルには、NHKきょうの健康の書き起こし文を使用した。

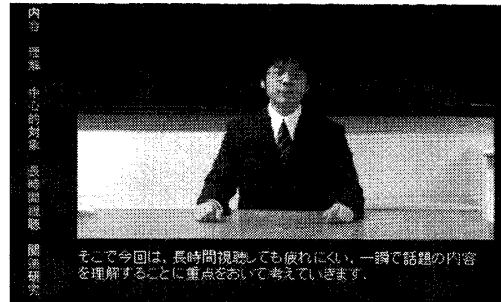


図1 L字型字幕表示方式

表1に人手で求めた重要度との相関を示す。素性ベクトルは各要素とも0-1で正規化されているため、重要度ベクトルの要素の大きさは、その素性の重要度を示しているといえる。

表1 重要度ベクトル推定結果

	経済	社会	政治	健康
TF	-74.1	166	-363	-118
IDF	-145	7.71	-21.7	-23.9
TF-IDF	283	62.8	598	295
接続	68.0	24.1	16.8	5.55
係り受け	145	11.1	3.59	12.4
固有名詞	637	-10.2	-3.92	28.6
相関係数	0.661	0.774	0.822	0.839

表1より、どのジャンルにおいても、一般的に単語重要度として用いられるTF-IDFが、重要度スコアの決定に大きく関与していることがいえる。他に、接続名詞数が高い重要度を示した。接続名詞数が小さい単語は、一般的な名詞、大きい単語は専門用語が多い傾向がみられた。つまり、専門用語はキーワードに適しているということを示している。より専門性の高いジャンルになると、接続名詞より固有名詞の重要度が高くなる。係り受けのスコアも全ジャンルにて、一定の重要度を示した。そのため、主語や目的語などの構文情報についても検討する必要がある。また、固有名詞では、人名、地名など、より詳細な情報を用いることにより、高精度なキーワード抽出を実現できる可能性がある。

3. キーワードの文字数・キーワード数

キーワードの文字数・キーワード数の最適数を調査するため、「字幕画面の見やすさ」、「キーワード提示による内容の分かりやすさ」の2つの項目を評価尺度として用いる。実験には2で使用した時論公論(ジャンル:経済)の映像を使用した。30秒を1フレームとし、全18フレームに被験者15名にて先述の項目について5段階評価を行った。図2は全18フレームでの

[†]岐阜大学大学院工学研究科

Graduate School of Engineering, Gifu University

[‡]岐阜大学工学部

Faculty of Engineering, Gifu University

「見やすさ」と「分かりやすさ」のスコアの2次多項式で近似した曲線であり、キーワード数による変化を示す。

「見やすさ」と「分かりやすさ」のスコアを足した値はキーワード数3のときが最も高く、文字数平均は13.4文字であった。フレームによっては、キーワード数5のときが最も評価が高いが、1語あたりの文字数が少なく、キーワードの総文字数がキーワード数3の文字数と大きな差は見られなかった。同様に、キーワード数1が最も評価が高い場合もみられたが、文字数が多いという結果がみられた。

図3にキーワードの文字数とそのスコアを示す。これによると、13文字から18文字程度が高い評価を得ていることが分かる。また、13文字とは、見出しにおいて最も適している文字数と考えられる数である[5]。

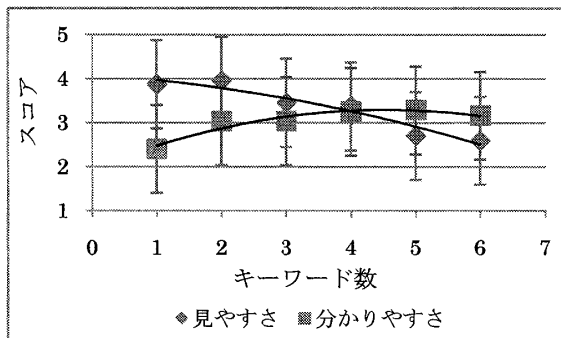


図2 キーワード数による評価

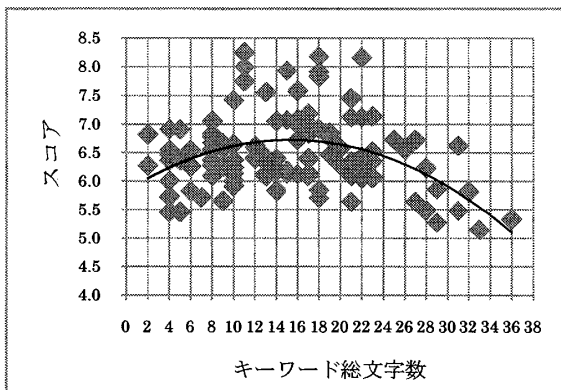


図3 キーワード総文字数による評価

4. キーワード表示方式の検討

先行研究では、字幕とキーワードの同時表示での検討がなされた。そこで本研究では、図4に示す、キーワードのみの表示位置についての調査を目的とし、被験者20名、5段階評価で評価実験を行った。図5に結果を示す。

(1), (2)については、映像コンテンツの背景色と文字色が重なり、文字が見えにくくなる場面がみられた。その場面が評価を下げた要因であると考えられるため、今後、文字色について検討する必要がある。

一方、(3)~(6)については、映像コンテンツ外にキーワードを配置することにより、色が重ならず、見やすいという評価を得た。また、日常的に映画やニュースなどで横書きの字幕を見慣れているため、縦書きよりも横書きが評価を高めたと考えられる。キーワード・字幕の同時表示では、アスペクト比を変更無く字幕を付与できるため、多用されるL字型字幕表示方式の評

価が最も高い。以上のことから、字幕表示方式において、「見慣れている」という要素が重要であるといえる。

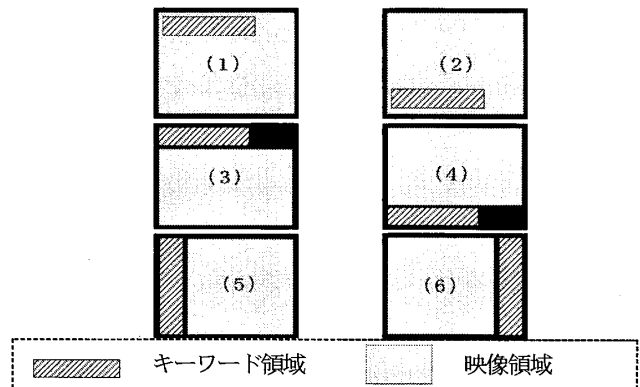


図4 キーワード表示位置

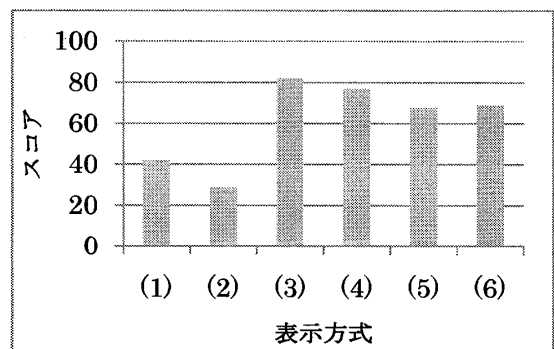


図5 キーワード表示位置の評価

5. まとめ

本研究では、ユーザの映像コンテンツの理解を支援することを目的とし、キーワード自動抽出、および表示方式の検討を行った。キーワード自動抽出では、ジャンルによる重要度ベクトルの変化を調査し、各ジャンルにおける有効な素性を確認した。キーワード表示方式における被験者実験では、横書きが最も評価が高く、映像コンテンツの理解度を高めることができると考えられる。

今後の課題として、より適切なキーワードを抽出するため、音声特徴量または、文字認識など画像処理を用いる手法、重回帰分析以外のモデルでの検討が挙げられる。表示方式については、文字色・フォントの検討が課題に挙げられる。

参考文献

- [1] 難波 誠一, “画像情報と提示時間”, NHK 技術研究所, Vol.35, No.1, pp.1-19 (1983).
- [2] 堀 智織, 古井 貞熙, “単語抽出による音声要約生成法とその評価”, 電子通信情報通信学会誌, D-II NO.2, pp.200-209 (2002).
- [3] 藤井 絢子, 南條 浩輝, 吉見 毅彦, “会議の情報保障を目的とした吹き出し型字幕提示方式の検討”, 情報処理学会研究報告, 2009-SLP-75-14, pp.75-82 (2009).
- [4] 岡本 昌直, 上地 春奈, 山本 けい子, 田村 哲嗣, 速水 悟, “キーワード抽出による映像コンテンツの理解支援とその心理的評価”, 言語処理学会 第15回年次大会, P2-27, pp.578-581 (2009).
- [5] 佐藤 理史, “13文字で何が伝えられるか: ウェブニュースボックス見出しの分析”, 言語処理学会 第14回年次大会 C3-7, pp.508-511 (2008).