

1E-3

視聴者入力情報からの時系列キーワード抽出システムの試作

大黒 泰平[†] 加藤 友規[‡] 土居 清之[‡] 亀山 渉[‡]

早稲田大学大学院国際情報通信研究科[†] 日本テレビ放送網株式会社[‡]

1. はじめに

我々は、番組視聴者が番組コンテンツのどのシーンに対してどのような感想を持ったのかについて、チャットの発言内容からキーワードを抽出し、メタデータとして再構築することを試みてきた¹⁾²⁾³⁾。

本稿では、抽出したキーワードを具体的にメタデータの形に再構築することを目標とし、発言情報の解析、メタデータ構造の定義、及びシステムの実装を行った為、報告する。

2. 視聴コンテンツと実験

番組視聴者のチャットでの発言傾向に最も大きな影響を及ぼす要因の1つに、番組コンテンツのジャンルが挙げられる。本研究では、番組コンテンツのジャンルが、番組視聴者のチャットでの発言にどのような影響を及ぼすのかを検証する為、最も典型的な番組ジャンルであると考えられる、1.バラエティ番組、2.スポーツ番組、3.ドラマ、4.ニュース番組の4種類の番組コンテンツを使用した。これら4種類の番組コンテンツについて、我々が行ってきた手法¹⁾²⁾³⁾に従い、実験を実施した。実験の内容は、「番組コンテンツを視聴しながら番組コンテンツに関する感想情報をチャットで書き込む」という形式のものである。約15名の被験者を募り、PC端末上に表示された番組コンテンツを視聴しながら、同じ画面上に表示されたチャットページ上で自由に会話を行わせた。

実験の後、番組コンテンツの内容の区切りをセグメントとして分割し、番組コンテンツの面白さを被験者が主観的に5段階で評定するアンケートと共に、各セグメントの発言傾向について分析を加えた³⁾。

3. 発言内容の分析

3-1. ジャンルによる発言傾向の差異

図1、図2に、ニュース番組とスポーツ番組における発言頻度と被験者アンケートで得られたポイントの分析結果を一例として示した。図1より、グラフ中にAで示したセグメントにおいて、チャットでの発言頻度を示すポイントと番組視聴者から得たアンケートのポイントが共に高い。よって、ニュース番組は番組視聴者が主観的に興味を持った場面では、チャットでの発言頻度が上昇する傾向があるものと推測することができる。一方、図2では、アンケートのポイントとチャットでの発言頻度を示すポイントとの間に関連性が見受けられない。よって、スポーツ番組は番組視聴者が主観的に興味を持った場面と、チャットでの発言頻度が上昇する場面には関連性が見受けられない傾向があると推測することができる。

ジャンル毎の分析結果については、バラエティ番組とニュース番組に関しては、セグメント毎に番組視聴者の感想の推移と連動して有意にキーワードを抽出できることが確認できた。その理由として、バラエティ番組やニュース番組は、番組のトピックがこまめに変わる、番組

視聴者同士での議論を促すような番組内容である、という点が挙げられる。一方、スポーツ番組とドラマについては、セグメント毎に有意にキーワードを抽出することができなかった。その理由として、スポーツ番組については話の展開が早すぎてチャットが展開に追いついていかない、ドラマについては、ジャンルの性質上、個々のセグメントという短い単位で感想を述べられるものではない、という点が挙げられる。

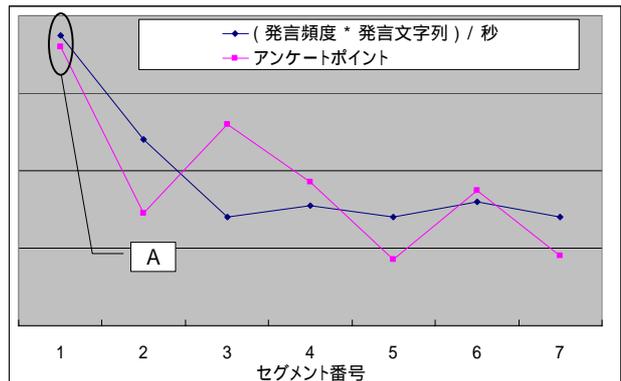


図1: ニュース番組の分析結果

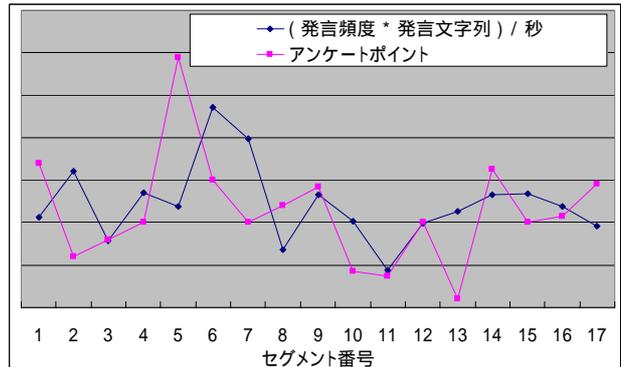


図2: スポーツ番組の分析結果

3-2. キーワードの抽出

チャットの入力情報を元に、番組視聴者の発言内容に対して番組コンテンツのセグメント毎に形態要素解析⁴⁾をかけた。解析は、図1のAで示した、ニュース番組における1セグメントに対して行った。形態要素解析の内、一例として固有名詞を抽出した結果を図3に、形容詞を抽出した結果を図4に、記号「！」(Exclamation mark)を抽出した例を図5に示す。各図は、形態要素解析の結果の出力であり、当該セグメント内において、番組視聴者が発言した内容を品詞毎に分類した結果を示している。各図より、当該セグメント内で番組視聴者がどのような固有名詞・形容詞・記号を発言したのかが伺える。

1. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
2. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
3. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
4. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
5. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
6. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名
7. イチロー	イチロー	イチロー	名詞-固有名詞-人名-名

図3: 固有名詞の抽出

A Prototype to Extract Time-based Keywords from Viewers' Chat Inputs to TV Programs

[†] Taihei DAIKOKU, E-mail: taihei_d@fuji.waseda.jp

Wataru KAMEYAMA

GITS, Waseda University

[‡] Tomonori KATO

Kiyoyuki DOI

Nippon Television Network Corporation

1.	暑い	アツイ	暑い	形容詞・自立	形容詞・アウオ段	基本形
2.	いい	イイ	いい	形容詞・自立	不変化型	基本形
3.	ない	ナイ	ない	形容詞・自立	形容詞・アウオ段	基本形
4.	嬉しい	ウレシイ	嬉しい	形容詞・自立	形容詞・イ段	基本形
5.	いい	イイ	いい	形容詞・非自立	不変化型	基本形
6.	いい	イイ	いい	形容詞・自立	不変化型	基本形
7.	すごい	スゴイ	すごい	形容詞・自立	形容詞・アウオ段	基本形
8.	すごい	スゴイ	すごい	形容詞・自立	形容詞・アウオ段	基本形
9.	カッコいい	カッコイイ	カッコいい	形容詞・自立	不変化型	基本形
10.	すごい	スゴイ	すごい	形容詞・自立	形容詞・アウオ段	基本形
11.	カッコいい	カッコイイ	カッコいい	形容詞・自立	不変化型	基本形
12.	すごい	スゴイ	すごい	形容詞・自立	形容詞・アウオ段	基本形
13.	新しい	アタラシイ	新しい	形容詞・自立	形容詞・イ段	基本形
14.	すごい	スゴイ	すごい	形容詞・自立	形容詞・アウオ段	基本形

図4：形容詞の抽出結果

1.	!	!	!	!	記号-一般
2.	!	!	!	!	記号-一般
3.	!	!	!	!	記号-一般
4.	!	!	!	!	記号-一般
5.	!	!	!	!	記号-一般
6.	!	!	!	!	記号-一般
7.	!	!	!	!	記号-一般
8.	!	!	!	!	記号-一般
9.	!	!	!	!	記号-一般
10.	!	!	!	!	記号-一般

図5：記号の抽出結果

図3より、野球選手の名前が7回と、頻繁に発言されている。よって番組視聴者は、野球選手に対して興味を持っているということが推測できる。また、図4より、「すごい」、「うれしい」、「カッコいい」といった形容詞が頻繁に発言されていることから、番組視聴者が野球選手に対して抱いている興味の種類の、「賞賛」に当たるものであるということが推測できる。更に、図5より、「！」が9回と頻繁に出現していることから、番組視聴者は、エキサイティングな感情と共に番組コンテンツを視聴しているということが推測できる。このように、当該セグメントにおいて、セグメント内で番組視聴者が抱いた感想情報が、キーワードの形で抽出されていることを確認できる。

4. メタデータ設計

ニュース番組のAのセグメントにおけるキーワード情報を元に、メタデータを構築した。本稿ではこのようなメタデータを、「キーワード抽出メタデータ」と呼ぶこととする。図6に、作成したキーワード抽出メタデータの構造を示す。なお、抽出したキーワードの内、固有名詞と一般名詞に関しては、その語自体が有意な意味を持っていると考えられるため、TV-Anytime Forum⁵⁾が規定しているメタデータの内、SegmentInformationTableの中にキーワード情報として定義することとした。また、その他の品詞については、番組視聴者が番組コンテンツに対して抱く感想情報を一定の「雰囲気」として定義し、雰囲気情報としてのメタデータを定義した。キーワード抽出メタデータ内の雰囲気情報の定義には、TV-Anytime Forumが規定しているメタデータの内、AtmosphereCSを利用し、ProgramInformationTable内に記述した。なお、これらの雰囲気情報は別個のメタデータとしてAtmosphereCS内のライブラリに定義されている。そのライブラリ内の雰囲気情報を参照するという形式でキーワード抽出メタデータ内では雰囲気情報を定義している。本セグメントで抽出された雰囲気情報として、以下の4種類が挙げられる。

1. Astonishing, 2. Exciting, 3. Fun, 4. Happy.
- 1は、すごい・すばらしい、などのキーワードに対応する。2は、感動・感心・興奮・感銘・!、などのキーワードに対応する。3は、カッコいい・たのしい・楽しい、などのキーワードに対応する。4は、うれしい・嬉しい、などのキーワードに対応する。

```

1. <?xml version="1.0" encoding="UTF-8"?>
2. <TVAMain xmlns="http://www.tv-anytime.org/2001/08/mef" >
3.   <ProgramDescription>
4.     <ProgramInformationTable>
5.       <ProgramInformation programId="crj://www.km.giti.waseda.ac.jp/news01.rm">
6.         <BasicDescription>
7.           <Genre preferenceValue="40" href="at:8.3">
8.             <Name>Astonishing</Name>
9.           </Genre>
10.          <Genre preferenceValue="40" href="at:8.16">
11.            <Name>Exciting</Name>
12.          </Genre>
13.          <Genre preferenceValue="40" href="at:8.19">
14.            <Name>Fun</Name>
15.          </Genre>
16.          <Genre preferenceValue="40" href="at:8.23">
17.            <Name>Happy</Name>
18.          </Genre>
19.        </BasicDescription>
20.      </ProgramInformationTable>
21.    </ProgramDescription>
22.    <SegmentInformationTable>
23.      <SegmentList>
24.        <SegmentInformation segmentId="1">
25.          <SegmentLocator>20041004115449-20041004115834</SegmentLocator>
26.          <Description>
27.            <Keyword>イテロ</Keyword>
28.          </Description>
29.        </SegmentInformation>
30.      </SegmentList>
31.    </SegmentInformationTable>
32.  </TVAMain>

```

図6：キーワード抽出メタデータの構造

5. 評価と考察

チャットにおいて発言された内容には、番組コンテンツとは直接関係のない情報や番組コンテンツの評価情報としては不適切な情報が多分に含まれている。これらの情報に対して形態要素解析をかけることにより、各々のセグメントにおけるキーワードを品詞ごとの単位で抽出することができた。また、抽出したキーワードをメタデータの形に再構築することにより、セグメント単位で番組視聴者の感想情報を記述したメタデータをリアルタイムで自動的に生成していけるものと考えられる。

6. まとめと今後の課題

本稿では、インターネットにおけるチャットでの発言情報を元に、番組コンテンツに対する発言からキーワードを抽出し、番組コンテンツに対する批評情報を定義したメタデータを自動的に生成する為の手法を提案した。結果、バラエティ番組とニュース番組に関しては、セグメント毎に番組視聴者の感想の推移と連動して有意にキーワードを抽出していけることが確認できた。また、番組視聴者の興味情報と連動して有意にメタデータを生成できることを確認することができた。

今後は、スポーツ番組やドラマといった番組ジャンルにおける同等の手法の提案が必要である。また、発言内容を的確に形態要素解析するための品詞情報ボキャブラリーの拡充、及び雰囲気情報を定義する為のメタデータ内の語彙の拡張などを行う必要がある。また、本手法を実際のフィールドにおいて適用可能であるのかどうかを検証する為の、大規模な実証実験を行う必要がある。

参考文献

- 1) 大黒他, “番組コンテンツにおけるユーザ入力情報からの時系列キーワード抽出に関する一考察”, IPSJ 第66回全国大会 4U-6(Mar 2004).
- 2) 大黒他, “番組に対する視聴者入力情報からの時系列キーワード抽出の効率化に関する検討”, ITE 年次大会 22-2(Aug 2004).
- 3) 大黒他, “番組に対する視聴者入力情報からの時系列キーワード抽出の改善に関する検討”, FIT2004, D-035(Sep 2004).
- 4) Cha-Sen: <http://chasen.aist-ara.ac.jp/hiki/ChaSen/>
- 5) TV-Anytime Forum: <http://www.tv-anytime.org/>