

# 聴講者の興味をひく研究発表タイトルの作成支援

## a Support System to Make Titles for Attracting Audience Interest

西原陽子<sup>1\*</sup> 砂山渡<sup>2</sup> 谷内田正彦<sup>1</sup>

Yoko NISHIHARA<sup>1</sup>, Wataru SUNAYAMA<sup>2</sup> and Masahiko YACHIDA<sup>1</sup>

<sup>1</sup> 大阪大学大学院基礎工学研究科, <sup>2</sup> 広島市立大学情報科学部

<sup>1</sup> Graduate School of Engineering Science, Osaka Univ., <sup>2</sup> Faculty of Information Sciences, Hiroshima City Univ.

### 1 まえがき

国内国外を問わず、多くの研究会や会議がある。研究者は成果を発表し、聴講者と議論を行い、新たな知見を得ることができる。多くの人に発表を聴講してもらうためには、興味を持ってもらえるタイトルやアブストラクトを作成することが望ましいが、その作成方法は明らかではない。本発表では興味をひく研究発表のタイトルを作成する支援システムを提案する。提案システムはタイトル集合が入力されたときに、分かりやすさと面白さの2つから、タイトルに評価値をつけ、評価値の高い順にタイトルを出力する。これによって新たな聴講者を獲得することを目指す。

### 2 関連研究

本章ではキャッチコピーの作成支援と、研究発表の支援に関する従来研究をまとめ、本研究の位置づけを明確にする。

松平らは遺伝的プログラミングと電子化辞書を用いて、キャッチコピーの作成を支援した [3]。このシステムはユーザが単語を入力すると、電子化辞書から入力された単語に関連する単語を出力し、遺伝的操作によってキャッチコピーの候補を作成する。多くのキャッチコピーを得ることができる利点があるが、研究内容を正確に表現し、意味の整合がとれたキャッチコピーが得られるわけではない。そのため本研究では人間がタイトルを作成し、システムは作成したタイトルを評価することで、タイトル作成を支援する。

世の中に広まる研究発表のタイトルを抽出するシステムがある [4]。このシステムではタイトルに含まれる単語の新規性、認知度、単語の組合せの斬新さからタイトルを評価する。このシステムではタイトル中の2つの単語だけ評価するため、抽出されるタイトルは必ずしも有効でないことが多い。提案システムでは、タイトルに含まれ、研究内容を表す単語を全て用いて、タイトルを評価する。

また、研究発表を申し込む際には、タイトルと抄録を添えるが、分かりやすい文章を書くために、文章の推敲を支援するシステム [2] や文章の作成を支援するシステム [6] がある。これらのシステムでは論文など、複数の文からなる文章を対象としているが、本研究では1つの文節からなるタイトルの作成を支援する。

研究発表を多くの聴講者に興味を持ってもらうためには、プレゼンテーションを工夫することも重要になる。発表スライドの構造化を支援するシステム [1] や、発表スライドを自動生成するシステム [7] もある。これらの手法では、聞きに来た聴講者を更に引きつけることが可能である。本研究では、タイトルを工夫することで研究内容に詳しくない人に興味を持たせ、新たな聴講者を獲得することを目指す。

### 3 提案システム構成

図1に提案システムの構成を示す。提案システムは入力にタイトルの集合をとる。入力されたそれぞれのタイトルを茶筌を用いて形態素解析し、単語を取り出す。続いて、タイトルの分かりやすさを評価し、次にタイトルの面白さを評価する。最後に分かりやすさと面白さの評価値からタイトルに順位を付け、順位が高い順にタイトルを出力する。以下では各処理を説明する。

#### 3.1 タイトルからの名詞の抽出

入力されたそれぞれのタイトルから単語（名詞）を抽出する。このとき、研究内容との関係が少ないと考え、以下の単語は抽出しない。

支援, システム, 分析, 提案, 手法, 適用, 解析, 同定, 発見, 技術, 開発, 実験, 抽出, 生成, 改善, 構築, 検証, 評価, 検討, 表示, 方法, 考慮, 付き, 比較, 利用, 試作, 研究, 考察, 実現

#### 3.2 タイトルの分かりやすさの評価

タイトルの分かりやすさを2つの方法で評価する。分かりやすいタイトルならば、タイトル中に互いに関連がある単語が多数含まれていると考えられる。そこで、

\*連絡先: 大阪大学大学院基礎工学研究科システム創成専攻  
〒560-8531 大阪府豊中市待兼山町1-3  
E-mail: yoko@yachi-lab.sys.es.osaka-u.ac.jp

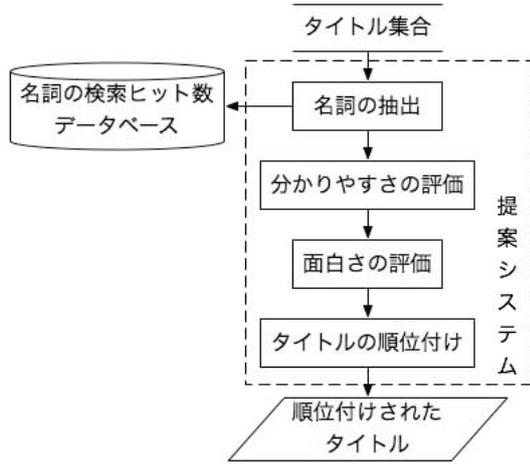


図 1: 研究発表タイトルの作成支援システム構成

タイトル  $t_{ll_i}$  に含まれる全ての名詞の組の数に対する、関連がある名詞の組の数の割合  $R(t_{ll_i})$  を式 (3) によって求める。

$$P(t_j) = \frac{H(t_j)}{ALL} \quad (1)$$

$$Rp(t_{ll_i}) = \sum_{0 \leq j < k \leq n_i} \begin{cases} 1 & \text{if } \frac{P(t_j)P(t_k)}{P(t_j \wedge t_k)} \geq T \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$R(t_{ll_i}) = \frac{Rp(t_{ll_i})}{Ap(t_{ll_i})} \quad (3)$$

式 (3) において、 $Ap(t_{ll_i})$  はタイトル中の名詞の組の数とする。式 (2) において、 $n_i$  はタイトル中の名詞の数、 $T$  は閾値とする。式 (1) において、 $H$  は名詞  $t_j$  の検索ヒット数、 $ALL$  は検索エンジン中の Web ページの数で  $10^{10}$  とする。

また、分かりやすいタイトルならば、タイトルが表示する研究内容を理解するために必要な知識の量が少ないと考えられる。そこで、従来手法 [5] を元に、タイトル中の単語の難しさを式 (4) によって求め、タイトルを理解するために必要な知識の量  $K(t_{ll_i})$  を、式 (5) を用いて評価する。

$$S(t_j, t_{ll_i}) = 1 - \frac{H(t_j)}{ALL} \quad (4)$$

$$K(t_{ll_i}) = MaxK - S(t_{max}, t_{ll_i}) + \sum_{0 \leq j \leq n_i} S(t_j, t_{ll_i}) P(t_j | \overline{t_{max}}) \quad (5)$$

式 (5) において、 $MaxK$  はタイトル集合における  $K(t_{ll_i})$  の最大値であり、 $t_{max}$  はタイトル  $t_{ll_i}$  に含まれる名詞の中で、式 (4) の評価値が最も高い名詞とする。

表 1: JSAI2006 データにおけるシステム出力のタイトル上位 10 件と評価値の順位 ( $\alpha = \beta = \gamma = \delta = 0.25$ )

タイトル	Val	Rtrm	Cmp	Nov	Fam
架空名義入札者発見に基づく不正入札に頑健なオークションメカニズム	416	11	93	2	2
分子ミュージック	646	32	6	0	17
ディスプレイロボットを利用した物体の擬人化	1260	6	11	42	44
仮想実験可能な惑星系シミュレータ	1372	8	8	31	69
知識検索サイトにおける有害情報のフィルタリング知識の表出化	1524	25	104	13	1
マルチロボットシステムにおける自発的な協調方法を用いた動的役割割当	1672	2	22	47	31
ボクシングにおけるスキル習熟過程について	1680	5	12	39	75
電子的コミュニケーションに用いる顔の抽象性と脳反応	2496	39	41	14	20
ヒューマノイド型ロボットの動作制御コード分析による身体知表出実験	2838	14	74	29	6
大学の授業を支援するウェブログネットワーク	3008	1	48	44	22

### 3.3 タイトルの面白さの評価

タイトルの面白さを 2 つの方法で評価する。面白いタイトルならば、タイトルに含まれる単語の組合せがありふれたものではなく、目新しいものが多いと考えられる。そこで、タイトルにおける単語の組合せの新しさの平均を式 (6) によって求め、タイトルの面白さ  $N(t_{ll_i})$  を評価する。

$$N(t_{ll_i}) = MaxN - \frac{2}{n_i(n_i - 1)} \sum_{0 \leq j < k \leq n_i} \frac{P(t_j)P(t_k)}{P(t_j \wedge t_k)} \quad (6)$$

式 (6) において、 $MaxN$  はタイトル集合における  $N(t_{ll_i})$  の最大値とする。

また、面白いタイトルならば、タイトルでは見かけないが、多くの人知っている有名な単語が含まれていると考えられる。そこで、単語が含まれるタイトル数に対する単語の検索ヒット数の割合から、タイトルの面白さ  $F(t_{ll_i})$  を評価する。

$$F(t_{ll_i}) = \max_{0 \leq j \leq n_i} \frac{H(t_j)}{C(t_j)} \quad (7)$$

式 (7) において、 $C(t_j)$  は単語が含まれるタイトルの数とする。 $H(t_j)$  の値が高すぎる単語は有名ではなく一般的なありふれた単語と考え、 $H(t_j)$  の値が 5 千万以上の単語は評価しない。

### 3.4 タイトルの順位付け

4つの評価式からタイトルに評価値を与え、順序を付ける。式(8)によって、タイトルに評価値を与える。

$$V(ttl_i) = \alpha \times O(R(ttl_i)) + \beta \times O(K(ttl_i)) + \gamma \times O(N(ttl_i)) + \delta \times O(F(ttl_i)) \quad (8)$$

式(8)において、 $O()$ は評価値の高い順に順位を与える関数とする。また、 $\alpha + \beta + \gamma + \delta = 1.0$ とし、 $\alpha, \beta, \gamma, \delta$ は0以上1以下の値とする。式(8)の評価値が低い順に、1位、2位...と順位をつける。

システムの出力例を表1に示す。ユーザは自分が作成したタイトルの順位と評価値を参照し、自らのタイトルに欠けている要素を補っていく。

## 4 実験1

本実験の目的は、提案システムによって多くの人が選ぶタイトルに高い順位がつくことの確認である。

### 4.1 実験手順

実験ではシステムを用いてタイトル集合を順位付けし、正解順位との相関を求めた。

正解を作成するために、1行に1つのタイトルが書かれたリストを被験者に示し、「良さそう、面白そう」と思うタイトルを1つ2秒程度で判断し、選んでもらった。また、タイトルを選び終わった後、選んだ基準を被験者に記述してもらった。

実験にはJSAI2006における研究発表のタイトルから英語のタイトルを除いた、120件を用いた。被験者は情報科学を専攻する大学生・大学院生58名である。

### 4.2 実験結果

被験者が選んだタイトルの上位10件と選んだ被験者の数を表2に示す。表2では「サッカー」「ペット」「ボクシング」など、一般に知られている単語が使われているタイトルが多い。また、「デザイン転写」「身体知」「人間乱数」など、既存の単語を組合せて作った新しい用語を含むタイトルも多い。

式(8)の最適な重みは $\alpha = 0.3, \beta = 0.1, \gamma = 0.5, \delta = 0.1$ であり、このとき順序相関係数が0.66であった。重みは分かりやすさが0.4、面白さが0.6に配分されたが、これは人間が良さそう面白そうと選ぶタイトルでは、分かりやすいよりも面白さを重視されることを意味する。また、全ての重みが0.1以上であったことから、人間

表2: JSAI2006 データにおける被験者が選んだ上位10件のタイトルと被験者数

タイトル	被験者数
サッカーシミュレータ環境における模倣による行動の学習	35
音楽のデザイン転写 技術の開発にむけて	31
心理状態認識を用いたペットロボットの行動選択手法の提案	29
画像情報を用いた自律移動ロボットの位置姿勢計測	28
ボクシングにおけるスキル習熟過程について	28
架空名義入札者発見に基づく不正入札に頑健なオークションメカニズム	27
注意機構にもとづくロボットの自律的な行動選択手法の提案	26
迷路問題における難易度指標の導入と実時間探索アルゴリズムの性能解析	25
マルチロボットシステムにおける自発的な協調方法を用いた動的役割割当	25
赤外線センサーネットワークによる人物追跡	24
ヒューマノイド型ロボットの動作制御コード分析による身体知 表出実験	24
人間乱数の分析	23

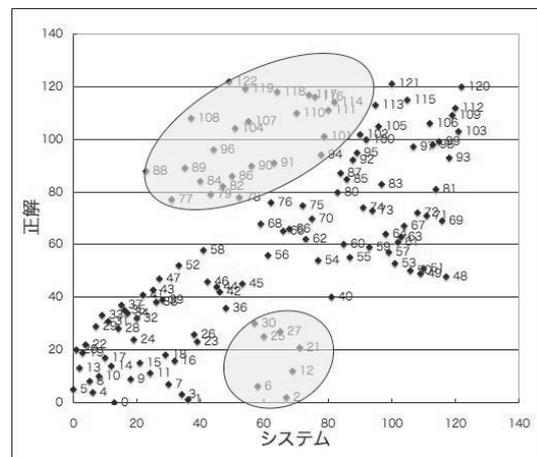


図2: JSAI2006のタイトルにおける正解とシステム出力の順序の相関

が良さそう、面白そうと選ぶタイトルの評価には4つの評価式が必要と分かる。

図2に、JSAI2006のタイトルにおける正解とシステム出力の順序相関を示す。図2では、1位から40位までは正解とシステム出力の順位がおよそ一致しているが、40位から80位までに順位が異なるタイトルがある(図2の楕円で囲まれた2つの部分)。上の楕円で囲まれた部分に関しては、タイトルでは見かけないが、有名な単語を評価する式において、有名ではなく一般的な単語に高い評価値が与えられたため、正解とシステム出力の順位が異なった。下の楕円で囲まれた部分に関しては、タイトル中のフレーズの組合せは新しいが、単語の組合せは新しくないため、正解よりも低い順位がついた。相関を上げるためにはこの2点を改善する必要がある。

表 3: 被験者が記述したタイトルを選んだ基準と記述した被験者数

タイトルを選んだ基準	被験者数
自分の興味・関心に合う	40
何となく面白そう	16
意味が分かる	12
自分の身近に存在する	3
見た目が良い	5
役に立つ・実用的	9

表 4: 正解とシステム出力の順序相関係数と式 (8) の最適な重み

学会名	順序相関係数	$\alpha$	$\beta$	$\gamma$	$\delta$
NLP2005	0.63	0.3	0.1	0.5	0.1
JSAI2005	0.67	0.2	0.1	0.5	0.2
FIT2005	0.61	0.2	0.2	0.4	0.2
FIT2006	0.62	0.4	0.1	0.4	0.1

続いて、被験者が書いたタイトルを選んだ基準を分類したものと記述した被験者の数を表 3 に示す。表 3 では、記述した人数が最も多かった基準は「自分の興味・関心に合う」であった。続いて人数が多かった基準は、「何となく面白そう」で、次に多かった基準は「意味が分かる」であった。このことから、タイトルの分かりやすさと面白さでタイトルを評価する提案システムが妥当なものと分かる。

## 5 実験 2

実験 1 と異なるタイトル集合を用いて、提案システムの有効性を確認する。

### 5.1 実験手順

実験 1 とは異なる 4 つのタイトル集合に対して、式 (8) の最適な重みを見つけ、この時の順序相関係数を求めた。実験では、被験者にタイトルのリストを示し、良さそう、面白そうと思うものを 1 つ 2 秒程度で判断し選んでもらった。

実験では英語のタイトルを除き、NLP2005 (242 件)、JSAI2005 (220 件) FIT2005 (260 件) FIT2006 (254 件) を用いた。被験者は情報科学を専攻する大学生・大学院生 16 人とし、1 つのタイトルを 8 人が評価した。

### 5.2 実験結果

表 4 に 4 つのタイトル集合における最適な重み、および正解とシステム出力の順序相関係数を示す。相関は 0.6 以上であり、システムは正解に近い順序でタイ

トルを順位付けできた。このことから提案システムで得られる評価値は妥当と分かった。

また、4 つのデータでは、JSAI2006 とは異なる重みが与えられた。だが、重みの配分から表 4 においても、分かりやすさよりも面白さが重視されることが分かる。すなわち、扱うタイトル集合によって最適な重みは若干異なるが、分かりやすさよりも面白さを重視し重みの配分をすることで、タイトルに意味のある評価値が与えられると分かった。

## 6 むすび

多くの人に興味を持つタイトルを作成する支援システムを提案した。実験によって、提案システムによって、多くの人を選ぶタイトルに妥当な評価値を与えられることを確認した。ユーザは提案システムの出力を参照することで、自分が作ったタイトルに欠けている要素を知り、より良いタイトルを作ることが可能になると思われる。今後の課題にはタイトルでは見かけないが有名な単語を取り出す方法と、フレーズの組合せの新しさを評価する方法を考えることがある。

## 参考文献

- [1] CoffeeMaker:(URL)<http://www.ipa.go.jp/jinzai/esp/15youth/mdata/99-18.html>
- [2] 林：技術文章向けの日本文推敲支援システムの実現と評価，信学論，Vol.J77-D-II, No.6, pp.1124-1134 (1994).
- [3] 松平，萩原：対話型遺伝的プログラミングと電子化辞書を用いたキャッチコピー作成支援システム，電学論 C，Vol.125, No.4, pp.616-622 (2005).
- [4] 西原，砂山，谷内田：有効な組合せの発見による創造活動支援，信学論，Vol.J87-D-I, No.10, pp.939-949 (2004).
- [5] 西原，砂山，谷内田：Web ページの難易度と学習順序に基づく情報理解支援システム，信学論，Vol.J89-D, No.9, pp.1963-1975 (2006).
- [6] 砂山，橘：サブトピックモデルに基づく文章の流れの評価指標の提案，日本知能情報ファジィ学会誌，Vol.18, No.2, pp.280-289 (2006).
- [7] 高橋メソッド風プレゼンジェネレータ：(URL)<http://www.madin.jp/pregen/>