

# 多くの人の興味をひく研究発表タイトルの作成支援

## Title-Composing Support System for Reaching New Audiences

西原陽子\*1\*2      砂山渡\*3      谷内田正彦\*2  
Yoko NISHIHARA      Wataru SUNAYAMA      Masahiko YACHIDA

\*1 日本学術振興会      \*2 大阪大学大学院基礎工学研究科  
Japan Society for the Promotion of Science      Graduate School of Engineering Science, Osaka University

\*2 広島市立大学大学院情報科学研究科  
Graduate School of Information Sciences, Hiroshima City University

Researchers talk about their studies in research meetings, international conferences, and so on. Though it is important to be listened to their presentations by many audiences, the audiences tend to select and audit presentations that they are interested in. Therefore, researchers must compose good titles that bring to audiences' attention in order to obtain new audiences. However, there have not been systems to compose such titles. This paper proposes a support system for composing good titles for research papers in order to reach new audiences. Our system takes titles as input. The system evaluates title understandability and audience interest in the title. The system ranks titles and outputs a titles' list. Users are able to recompose their titles by referring to the list and each evaluation value. Experimental results showed that our system ranks titles in order of descending subjects' choices.

### 1. はじめに

国内国外を問わず、多くの研究会や会議があり、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。規模の大きい研究会では、全ての発表を聴講できないため、聴講者はタイトルやアブストラクトを見て、興味を持った発表に絞って聴講する。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。

### 2. 関連研究

#### 2.1 キャッチコピーの作成支援

販売されている商品には商品名だけでなく、商品の魅力を伝えるためのキャッチコピーがつけられていることがある。よく売れる商品につけられているキャッチコピーには、「キャッチコピーの中に商品のチャームポイントが書かれている」、「一瞬で分かる言葉が使われている」、「競合商品と差別化されている」の3つの要素が備わっているとされている[田村 05]が、現状では商品開発者が試行錯誤して作ることが多い。

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[松平 05]。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、ユーザが選択した単語から遺伝的操作によってキャッチコピーの候補を作成し、出力する。多くのキャッチコピーを得ることができる利点がある。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タイトルを作成するのは、人間とし、システムはその評価を行うだけとする。

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[西原 04]。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ものである。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用いて、タイトルを評価する。

#### 2.2 研究発表の支援

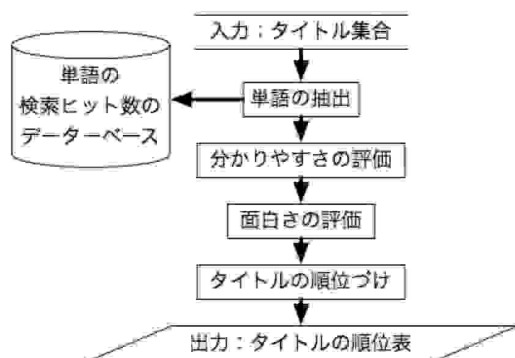
研究発表を申し込む際には、タイトルと抄録を添えることがある。分かりやすい文章を書くために、文章の推敲を支援するシステム[林 94]や文章の作成を支援するシステム[砂山 06]が研究されてきている。これらのシステムでは論文など、ある程度の量がある文章を対象としているが、本研究では論文よりも短いタイトルを扱い、その作成を支援することとする。

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でも可能である。実際の発表におけるスライドの構造化を支援するシステム[丸山]や、聴講者の興味を引くプレゼンテーションの手法やスライドを自動生成するシステム[高橋]も研究されている。これらの手法では、興味を持って会場に足を運んでくれ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全く興味がない人に興味を持たせ、聴講者の層を広げることが目標とする。そのために、人間がプログラムを眺めているときに、良さそう、面白そうと思い、選ぶタイトルを作成するシステムを構築する。

### 3. 提案システム構成

図3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。

連絡先: 西原陽子, 日本学術振興会/大阪大学大学院, 大阪府豊中市待兼山町 1-3, (tel) 06-6850-6363, (fax) 06-6850-6341, nisihara@gmail.com



### 3.1 タイトルからの単語の抽出

入力されたタイトルから単語（名詞に限定する）を抽出する。このとき、研究内容との関係が低いとし、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、単語  $w_a, w_b$  の相互情報量  $pmi$  の値が  $T^{*1}$  以上となる組を関連があるとし、タイトル  $t_i$  に含まれる全ての単語の組の数  $Ap(t_i)$  に対する、関連がある単語の組の数  $Rp(t_i)$  の割合  $R(t_i)$  を式 (1) によって求める。

$$pmi(w_a, w_b) = \frac{H(w_a) \times H(w_b)}{H(w_a \wedge w_b) \times H(ALL)}$$

$$R(t_i) = \frac{Rp(t_i)}{Ap(t_i)} \quad (1)$$

上の式において、 $H()$  は検索エンジン<sup>\*2</sup>における単語のヒット数、 $ALL$  は検索エンジン中の Web ページの数<sup>\*3</sup>とする。

また、分かりやすいタイトルならば、タイトルが表す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  $s(w_a)$  を求め、タイトルを理解するために必要な知識の量  $K(t_i)$  を、式 (2) を用いて評価する。

$$s(w_a) = 1 - \frac{H(w_a)}{ALL}$$

$$K(t_i) = MaxK - s(w_{max}) + \sum_{w_a \in W} s(w_a)P(w_a | \overline{w_{max}}) \quad (2)$$

式 (2) において、 $MaxK$  はタイトル集合における  $K(t_i)$  の最大値であり、 $w_{max}$  はタイトル  $t_i$  に含まれる単語の中で、評価値  $s$  が最も高い単語、 $W$  はタイトル  $t_i$  中の単語集合とする。

\*1 上から 25% に位置する値

\*2 提案システムでは Google を用いた。

\*3 提案システムでは  $10^{10}$  とした。

### 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を式 (3) によって求め、タイトルの面白さ  $N(t_i)$  を評価する。

$$N(t_i) = MaxN - \frac{2}{n_i^2 - n_i} \sum_{w_a, w_b \in W; w_a \neq w_b} pmi(w_a, w_b) \quad (3)$$

式 (3) において、 $MaxN$  はタイトル集合における  $N(t_i)$  の最大値、 $n_i$  はタイトル  $t_i$  中の単語の数とする。

また、面白いタイトルならば、タイトルでは見かけないが、多くの人を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  $F(t_i)$  を評価する。

$$F(t_i) = \max_{w_a \in W} \left\{ \frac{H(w_a)}{C(w_a)} \right\} \quad (4)$$

式 (4) において、 $C(w_a)$  は単語が含まれるタイトルの数とする。 $H(w_a)$  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、 $H(w_a)$  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つの評価式をまとめた。式 (5) によって、タイトルに評価値を与える。

$$V(t_i) = \alpha \times O(R(t_i)) + \beta \times O(K(t_i)) + \gamma \times O(N(t_i)) + \delta \times O(F(t_i)) \quad (5)$$

式 (5) において、 $O()$  は評価値の高い順に順位を与える関数、 $\alpha + \beta + \gamma + \delta = 1.0$  かつ  $0.0 \leq \alpha, \beta, \gamma, \delta \leq 1.0$  とする。式 (5) の評価値が低い順に、1 位、2 位... と順位をつける。

### 3.5 出力：タイトルの順位表

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。例えば、「発話意図を用いた Web における人同士の関与の強さ推定手法<sup>\*4</sup>」では、評価値の最大値が 122 であるときに、式 (1) の関連のある単語の割合の評価値が 112、式 (2) の必要な知識量の評価値が 71、式 (3) の単語の組合せの新しさの評価値が 13、式 (4) の評価値が 68 となる。最大値は 122 となっている。このタイトルは関連がない単語や「関与」といった難しい単語が用いられているために、分かりやすさの評価値が低いと分かる。ユーザは評価値を見て、より関連がある単語や分かりやすい単語に取り換えることで、多くの人に選ばれるタイトルへと改善していくことができる。

## 4. 評価実験

提案システムによって、多くの人を選ぶタイトルに高い順位がつくことを確認した。

### 4.1 実験概要

実験では、提案システムを用いてタイトルを順位付けし、順位付けされたタイトルと被験者が選んだタイトルを比較した。

\*4 JSAI2006 で第一著者が発表した研究のタイトルになる

表 1: システム出力の例 (JSAI2006 のタイトル .  $\alpha = 0.3$  ,  $\beta = 0.1$  ,  $\gamma = 0.5$  ,  $\delta = 0.1$ )

順位	タイトル	V	R	K	N	F
1	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
2	分子ミュージック	646	32	6	1	17
3	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
4	仮想実験可能な惑星系シミュレータ	1372	8	8	31	69
5	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
6	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
7	ボクシングにおけるスキル習熟過程について	1680	5	12	30	75
8	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
9	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
10	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
-	-	-	-	-	-	-
91	発話意図を用いた Web における人同士の関与の強さ推定手法	1757392	112	71	13	68

実験で用いたタイトルは人工知能学会全国大会の 2005 年のタイトル (JSAI2005) と 2006 年のタイトル (JSAI2006) , 情報科学技術フォーラムの 2005 年のタイトル (FIT2005) と 2006 年のタイトル (FIT2006) , 言語処理学会年次大会の 2005 年のタイトル (NLP2005) であり, 英語のタイトルを除き, 122 個プログラムでの掲載順に JSAI2006 は上位 122 個, 他は上位 120 個を抜粋し, 用いた .

被験者には 1 行に 1 つのタイトルが書かれたリストを示し, 「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 . この意味は, 直感で何となく選んでもらわれるタイトルを提案システムが抽出できるかを確認するためである . 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 . 被験者は大学生・大学院生とし, JSAI2006 は情報科学を専攻する大学生 3 年生 58 名, 他のタイトル集合では工学を専攻する大学生 33 名を割り当てた . 全体の 30 % に当たる被験者によって選ばれたタイトルを正解とし, システム出力と比較した .

#### 4.2 実験結果

JSAI2006 のタイトル集合から, 被験者によって選ばれた上位のタイトル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 .

システム出力の上位 30 % に含まれた正解タイトルの割合を表 3 に示す . 表 3 では, 上位 30 % の中に正解の半分程度が含まれた . このことより, 提案システムは多くの人に選ばれるタイトルに高い評価値を与えられると分かった .

表 3 では, 式 (5) の重みはデータによって異なるが, 分かりやすさが 0.4 程度, 面白さが 0.6 程度に配分されている . 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った .

表 4 に正解順位とシステム出力の順位の順位相関係数を示す . 順位相関係数はケンドールの順位相関係数とした . 相関係数は 0.4 ~ 0.6 程度の値であれば, 中程度の相関があるが, 全てのデータで中程度の相関が得られた . このことから, 提案システムによって与えられる順位は人間が感じる順位と近いと分かった .

表 2: 選んだ被験者数が多かった上位のタイトルとその数 (JSAI2006 のタイトルを用いた)

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向けて	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

表 3: 出力上位 30 % において含まれた正解タイトルの割合と式 (5) の重み

データ	割合				
JSAI2006	0.66	0.3	0.1	0.5	0.1
JSAI2005	0.49	0.2	0.1	0.5	0.2
FIT2005	0.57	0.2	0.2	0.4	0.2
FIT2006	0.49	0.4	0.1	0.4	0.1
NLP2005	0.55	0.3	0.1	0.5	0.1

表 4: システム出力の順位と正解順位の順位相関 (式 5 の重みは表 3 の通り)

データ	JSAI2006	JSAI2005	FIT2005	FIT2006	NLP2005
相関係数	0.66	0.67	0.61	0.62	0.63

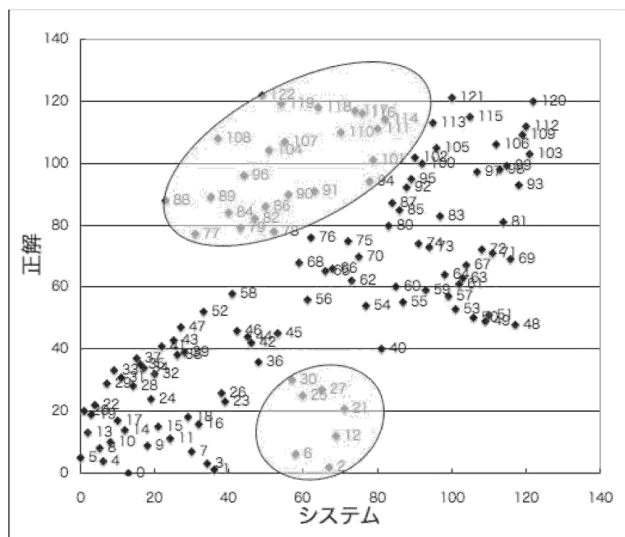


図 1: JSAI2006 データでの正解順位とシステム出力順位の相関

表 5: 正解とは異なり、高い順位になったタイトルの例 (太字は式 (4) の値が最大となった単語)

タイトル	R	K	N	F
日常体験記録利用のための情報抽出に関する研究	111	1	50	7
エリアスキャナによる人物属性識別システム	106	22	52	13
文献情報に基づく分野間ネットワーク分析	82	13	69	29
講義情報を用いた教材配信制御システム	118	24	60	15

図 1 に、JSAI2006 のタイトルにおける正解とシステム出力の順位相関を示す。図 1 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った (図 1 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で適切な評価値とならなかったため、正解順位とシステム出力の順位が異なった。これは、表 5 に示す「記録」「エリア」「分野」「教材」など、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である。過去のタイトル集合も考慮することによって、これらの単語を除くことができると考えられる。また、下の楕円で囲まれた部分に関しては、被験者は最初の方に面白そうな単語があるかどうかを判断しているのに、それを評価できなかったことが原因です。表 6 に示す「競」「ペット」「ゲーム」などタイトルでは見かけないが、有名な単語に高い評価値がつかず、正解よりも低い順位がついた。以上より、式 (4) のタイトルでは見かけないが有名な単語を取る方法を改善し、面白さの評価値が高いタイトルを優先的に抽出して、ランキングの方法を改善する必要があると分かった。

最後に、被験者によって書かれ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 7 に示す。表 7 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。この結果から、人は自分の興味にあったモノをまず選び、続いて、分かりやすいモノ、面白そうなモノを選ぶ傾向があると分かった。したがって、タイトルの分かりやすさと面白さでタイトルを評

表 6: 正解とは異なり、低い順位になったタイトルの例 (太字は式 (4) で評価されなかった単語)

タイトル	R	K	N	F
競による人間とペットロボットの関係の改善	36	90	17	99
身体スキル獲得プロセスにおける身体部位への意識の変遷	75	49	80	38
対話ゲーム木のための非決定加算的信念強化	92	72	54	52

表 7: アンケートで記述された被験者がタイトルを選んだ基準 (JSAI2006 のデータで回答した被験者のみ)

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
役に立つ・実用的	9
見た目が良い	5
自分の身近に存在する	3

価する提案システムが妥当と確認できた。

## 5. むすび

多くの人の興味をひくタイトルを作成する支援システムを提案した。提案システムは、タイトルの分かりやすさと面白さを評価し、4 つの評価値からタイトルを順位づけし、出力する。実験によって、提案システムによって、多くの人が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。

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の改善と、面白いタイトルを優先的に表示するタイトルのランキング手法を改善することがある。

## 参考文献

- [林 94] 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.
- [丸山] CoffeeMaker:(URL)<http://www.ipa.go.jp/jinzai/esp/15youth/mdata/99-18.html>
- [松平 05] 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.
- [西原 04] 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.
- [西原 06] 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.
- [砂山 06] 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，知能と情報，Vol.18, No.2, pp.280-289, 2006.
- [高橋] 高橋 メ ソ ッ ド 風 プ レ ゼ ン ジ ェ ネ レ ー タ : (URL)<http://www.madin.jp/pregen/>
- [田村 05] 田村：たった 1 行で！売る，実業之日本社，2005.