

アイデア創出活動におけるディスカッション時の興味抽出手法の検討

井上正博 平島大志郎 勅使河原可海

創価大学大学院工学研究科情報システム学専攻

大学での研究活動などでは、新しいアイデアを創出する活動が行なわれる。その組織内の人は、今まで知らなかった情報に触れることでその活動をより促進することができる。新しい情報に触れるためには、その人にとって最適な情報が推薦されることによって加速される。そのためには、人の興味を抽出する必要がある。アイデア創出活動の場として発表とディスカッションを行なう会議があり、そこでは多くの興味が出現する。会議では発表者は、参加者による様々な観点からの意見をもらい、その意見によって新たな興味を持つ。しかし、発表者は知識不足などにより本来は有益な意見を、自分には関係ないとして見落としてしまうこともある。したがって、見落とした意見の内容に関して、発表者は興味を持っていない。そのため、アクセス履歴などから興味を抽出する既存研究では、発表者が持っていない興味を抽出できない。しかしながら、本来は有益な意見であるため、その内容に関する情報を推薦して欲しいという発表者の潜在的な要望があるはずである。そこで本研究では、発表資料と発言録を利用して、ディスカッションでの発表者にとって有益でかつ予想外な発言を自動抽出する。そして、抽出した発言からユーザの潜在的な要望を満たす興味抽出手法を検討する。

A Study on Extraction of Discussion Interests in New Idea Creation Environment

Masahiro INOUE Daishiro HIRASHIMA Yoshimi TESHIGAWARA

Graduate School of Engineering, Soka University

ABSTRACT: There are many occasions to create new ideas in research activities in universities as well as business activities in companies. Creating new ideas can be stimulated for the members of the organization by encountering with new information that they have never heard. In order to encounter such new information it can accelerate for each person to be recommended appropriate information. For that reason, each person's interests must be extracted. For the places to create new ideas, meetings that have presentation and discussion are quite effective, and a lot of interests can be emerged there. In meetings, presenters get feedbacks from various points of view, and become interested in new information from such feedbacks. However, presenters may not notice the useful information, because of deficient knowledge. Therefore, presenters do not draw their interests about the contents of unaware opinions and comments. An existing study to extract interests from access logs cannot extract the appropriate interests that presenters are not aware. However, presenters substantially want to have such information recommended, because it is essentially beneficial for them. In this research, we propose a method to automatically extract opinions and comments that are useful but unexpected for presenters by using meeting minutes and materials such as presentation slides. In addition, we discuss a method to extract person's individual interests that satisfy his or her latent desire from extracted opinions and comments by the proposed method.

1. はじめに

大学での研究活動などの中では、今までにないアイデアを創出する活動が存在する。その時期は、研究論文や書籍などの今まで知らなかった情

報が提示されることで、より新しいアイデアを創出することが可能になると考えられる。特に、近年急速に普及したことで、世の中の最新の情報や多くの人の考えや意見が集まっているインタ

インターネットには、多くの有益な情報が存在していると考えられる。しかし、インターネットには有益でない情報も多く存在するため、アイデア創出活動を行なう人にとって有益な Web ページを発見することは容易ではない。そのため、Web ページ推薦を行なう研究やシステム、サービスが多く存在している[1][2][3]。情報推薦を行なうためには、人の推薦して欲しい情報かどうかを判断するために、人の興味を抽出する必要がある。その興味を抽出する方法として、人に明示的に興味を入力させる手法やアクセス履歴や人の視点などから暗黙的に興味を抽出する手法がある[1]。また、ユーザのやり取りしているメールの内容に関連する Web ページを推薦する研究がある[2]。さらに、ユーザが投稿した blog エントリの内容に関連する Web ページを推薦する研究も行なわれている[3]。しかし、我々はこれらの研究では、アイデア創出活動を行う人の興味を完全に抽出することは難しいと考えた。その理由を以下に示す。

アイデア創出活動において人の興味を抽出を考えるうえで、知識活動と会議などのディスカッションとの関係を示した DRIP モデルを参考とした[2]。DRIP モデルを図 1 に示す。



図 1 DRIP モデル

Discussion (発表・議論) フェーズでは、テーマに関するアイデアや知識を他者に対して発表し、共有することで多角的な視点から議論を行う。Rumination (整理) フェーズにおいて議論内容の整理を行い、Investigation (調査・考察) フェーズにおいて、発表中に指摘された箇所の調査やアドバイスに基づいて考察などを行う。これにより、新しいアイデアを創出し、知識として

理論化・具体化していく。様々なプロセスを通じて発生したアイデアや知識を基に Preparation (まとめ) フェーズでは、効率的にアイデアや知識を共有するために発表資料を作成する。そのため発表資料は、Discussion フェーズまでのアイデアをまとめたものということができる。

これらのフェーズを完全に行なうことで、アイデア創出活動を行なう組織内の人がディスカッションでもらった意見の中から Rumination フェーズで整理し、今後の興味を確定する。そのため、このフェーズ後に整理した内容のメールや blog エントリを作成した場合は、興味を抽出することも可能であると考えられる。また、この興味に基づき Investigation フェーズにおいてインターネットを利用して調査を行なった場合は、その人の調査時に閲覧した Web ページのアクセス履歴の傾向などから抽出することができる。

しかし、Rumination フェーズにおいてすべての人が完全に整理することは容易ではない。特に、アイデア創出活動の経験の少ない人の場合はなおさらである。例えば、ディスカッションでもらった重要な意見を知識不足や見逃しにより、重要でないと判断してしまうことがある。この時は、その人は異なった、もしくは完全でない興味を基に行動を行なうため、その行動から抽出された興味をもとに推薦される情報も、また異なった、もしくは完璧でない情報が推薦される。また、既存研究は、メール・blog エントリの作成や、インターネットでの調査を行わない場合は興味を抽出することができない。このように、アイデア創出活動を行なう人が見逃しているディスカッションでの重要な意見である潜在的な興味を抽出することを本研究の目的とする。

2. ディスカッション時の興味抽出手法

潜在的な興味を抽出のために、ディスカッションの意見・質問にその回答がまとめられている発言録から興味を抽出を行なう。潜在的な興味を抽出できることで、ユーザに新たな興味を持たせることができる。さらに、見落としした興味の情報で

あるため、意外性のある情報を提示できる可能性が高くなる。

また、違った観点や視野からの意見を得られることでユーザの興味に変化し易いディスカッションは、アイデア創出活動において必ず行なわれるため、確実に興味を抽出することができる。

具体的には、有益でかつ予想外な発言の抽出とユーザの興味を表すキーワード抽出の2つのステップで構成されている。

2.1. 有益でかつ予想外な発言の抽出

本節では、1つ目の有益でかつ予想外な発言を抽出するステップについて述べていく。ディスカッションでの発表者の興味の変化を抽出することは、ディスカッションで起こった新しい興味を抽出することであると言い換えることができる。そこで、ディスカッション時の発表者にとって予想外な発言の中でも有益な発言を抽出する2つのステップについて述べる。ここで述べる予想外な発言とは、発表者がディスカッション以前に質問されると予想できなかった発言である。有益な予想外発言とは、その中でアイデア創出を行なう上で助言になった発言のことである。

(1) 予想外な発言の抽出

専門用語(キーワード)自動抽出サービス「言選Web」[3]を利用して発表資料と発言録からそれぞれキーワードを抽出する。ただし、人の興味を示すには重要でないキーワードを省くために、抽出されたキーワードの中で「質」などの1文字のキーワードと、「10問」などの数値を含むキーワード、重みが最低値の1のキーワード、研究室の学生の名前は除いた。図2に発言録の一部を例として示す。図3にその発言録から抽出されたキーワードと重みの一部を示す。

```

A: ↓
P15 表はどのような基準で類似度を算出しましたか?
↓
B: ↓
これは最初に使った問題で算出して、類似問題を自分で色分けしました
↓
C: ↓
問題1に対してやった、問題2に対してやったっていうわけ方?
↓
B: ↓
そうです。
↓
C: ↓
キーワードを抜きにしたら、文系も理系も同じ問題が起きる思う。キーワードを使わない場合はどうなの?
↓
B: ↓
キーワードを使わないとダメでした。文系は特に。
↓
D: ↓
最終目標が様々な授業形態に対応できるもので、文系に対応したいのはかる。てるがやりたいのは再利用しないかと。まずは前提条件を設けて理系問題でしっかりやってから文系をやるって課題のほうがいいかも。
↓
B: ↓
理系ではある程度うまくできたけど、文系と理系では問題形式が違うのどちらも総合的に考慮していかないといけないです。
↓
D: ↓
文系も理系の人が作ってるように、それはB君がやることなの?
↓
B: ↓
...学生の意欲を重視してるのでくると思う

```

図2 発言録の例

理系,13.00,	技術,4.00,	サブカテゴリ,3.46,
文系,13.00,	最終目標,3.46,	類似問題群,3.32,
問題,12.25,	汎用性,3.46,	問題形式,3.13,
手法,6.93,	問題数,3.46,	学生,3.00,
カテゴリ,6.93,	応用先,3.46,	自動,3.00,
文章,5.00,	分類手法,3.46,	研究,3.00,

図3 抽出したキーワードとその重みの例

次に、発表資料には、発表者の今までの考えやアイデアがまとめられているため、その中に含まれないキーワードは、発表者にとって予想外なキーワードである可能性が高い。そのため、発表資料に含まれていない発言録のキーワードのみを新出語として抽出する。そして、その新出語を含む発言を発表者にとって予想外の発言であると判断する。

(2) 有益でかつ予想外な発言の抽出

新出語には、発表者にとって予想外なキーワードも含まれているが、それ以上に興味とは全く関係のないキーワードも含まれている。そこで、新出語の中からそれらのキーワードの省いていく必要がある。そのために、発言の中で重要なキーワードのみを2つの条件で抽出する。そして、重要新出語を含む発言を有益でかつ予想外な発言と判断する。これらの具体的な処理の例を図4に示す。

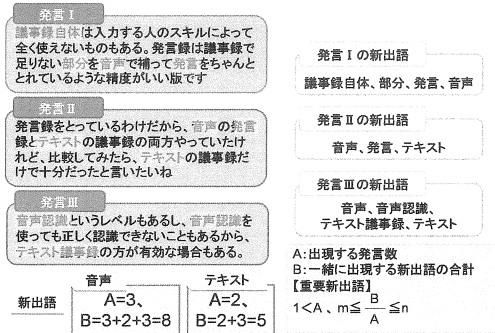


図 4 重要新出語抽出例

まず、発言録では、質問と回答のやり取りが多く行なわれている。そのため、質問の発言で出現したキーワードが重要なキーワードである場合は、回答の発言でも出現する可能性が高い。そこで、発言録の中で新出語が出現する出現数 A を算出する。そして、出現数 A が 2 以上を重要新出語とする。

次に、発言の中で重要なキーワードは、発言録の発言の長さに関係なく出現しているはずである。新出語と一緒に同じ発言に出現する他の新出語を共起語とし、発言録の中の共起語の合計数 B を求める。次に、B を A で割って、出現する発言における共起語数の平均の値を算出する。平均値が高くなると、多くの新出語と一緒に出現しているため、その発言は長文ということができる。そこで、平均の最適な範囲である m と n を予備実験で求めている。

2.2. 人の興味を表すキーワード抽出

まず、本研究では既存研究と同じ興味表現をすることで汎用性を持たせるために、人の興味をベクトル空間によって表現する。しかし、既存研究と本研究のベクトル空間での重みづけが違うため、今回は重み重視しないこととする。そのため、今回は人の興味をキーワードで表現する。そして、今回はキーワードの重みをキーワード抽出で利用した「言選 Web」により付けられた重みを利用する。重みとは、図 3 の中の数値の部分である。次に、人のアイデア創出活動に関する興味を抽

出する手法について述べていく。本研究では、人の興味を示すキーワード群の抽出手法として、以下の 2 つを提案し、最も実験結果が良い手法を採用する。それぞれの提案手法の例を図 5 に示す。

提案手法 1

発言録の中で予想外で有益な発言を抽出するために、用いた予想外でかつ発言の中で重要なキーワードである重要新出語を人の興味のあるキーワードとする。

提案手法 2

抽出した発言の中で最も重要なキーワードを人の興味として抽出したい。そのため、発言録からキーワードを抽出するときに用いた言選 Web の重みが高いものを重要なキーワードと仮定した。そこで、抽出した発言の中で重みが最も高い新出語 1 つずつを人の興味のあるキーワードとして抽出する。

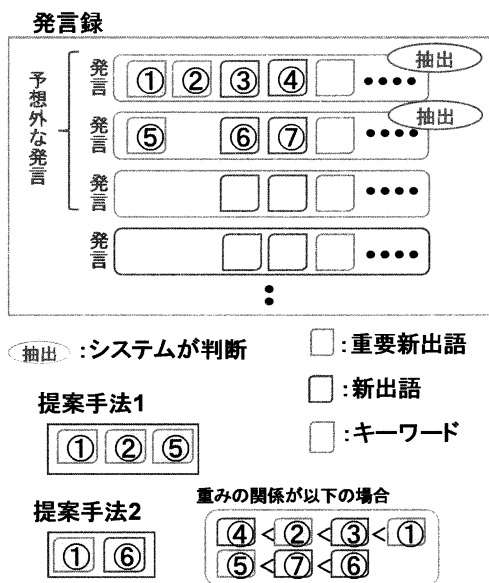


図 5 各提案手法の例

3. 評価実験・考察

本実験は、アイデア創出時期の可能性が高い学部 4 年生が中心の 7 名が発表した 8 回の研究室全体ゼミとグループゼミを対象とした。

実験 1 は、有益でかつ予想外な発言を抽出できているかを評価していくための実験である。実験手法は、発表者にゼミが終わって十分な期間において発言録の中の発言から有益でかつ予想外の発言のみを選択してもらった。そして、提案手法を用いて抽出した発言を比較して精度を評価した。精度を評価するために、利用する再現率と適合率のイメージ図を図 6 に示す。

|U|: 発言録の全発言数

|N|: システムが抽出した発言数

|C|: 発表者が選択した「有益でかつ予想外な発言」の数

|R|: システムが抽出した「有益でかつ予想外な発言」の数

$$\text{再現率 (網羅性)} = \frac{|R|}{|C|}$$

$$\text{適合率 (正確性)} = \frac{|R|}{|N|}$$

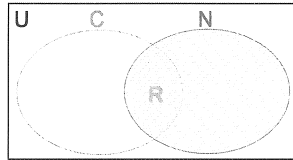


図 6 再現率と適合率

まず、提案手法では、閾値の m と n を求める必要があるため、最初の 4 回のゼミを利用して呼び実験を行なった。その結果を表 1 に示す。

予備実験では、発表者が選択した「有益でかつ予想外な発言」を全部抽出できるように、再現率が 100% になる閾値を求めた。今回は $m=1$, $n=4$ となった。

表 1 予備実験のデータ

	ゼミ名			
	A	B	C	D
再現率	100%	100%	100%	100%
適合率	55%	17%	24%	29%

次に、残りの 4 回のゼミでも同様の閾値で抽出してみた。その結果を表 2 に示す。ゼミ E だけ再現率が 100% でなかった。これは、出現数 A が 1 だったため 100% にならなかった。先ほど決定した m と n の閾値は問題ないことを示している。

適合率に関しては、表 1 と表 2 の平均は 32% となった。これは、発言録の中に有益でかつ予想外な発言が 8% しか含まれていない中から抽出しているため、多くの対象外の発言を省くことができている。しかし、まだ精度を向上させる必要がある。原因の 1 つとして、どのゼミでも出現していた重要でないキーワードを重要新出語していたことである。例えば、「情報」「目的」など研究活動でよく出てくるが、情報を推薦するためには重要でないキーワードである。これらを事前登録や学習機能などにより、取り除く仕組みを付加させることでより精度は上がると考えられる。

表 2 実験 1 の結果

	ゼミ名			
	E	F	G	H
再現率	75%	100%	100%	100%
適合率	18%	31%	56%	31%

次に、抽出したキーワードから発表者の興味を示すキーワードを抽出する手法を評価するための実験 2 について述べていく。実験 1 の 7 名が発表した 8 回のゼミのうち、進捗報告などアイデアを創出した会議を除いた 5 回のゼミを対象としている。実験 2 の手法は、対象の発表者に発言録から抽出したキーワードそれぞれを以下の 5 種類に分類してもらった。

- i) 研究に関係しないキーワード
- ii) 研究に関係するか判断できないキーワード
- iii) 研究に関する Web ページを検索する時にほぼ毎回利用するキーワード
- iv) ゼミ後に研究に関する Web ページを検索する時に利用したキーワード
- v) 研究に関係するが Web ページ検索の時に利用したことがないキーワード

そして、全対象発表者がそれぞれ分類したキーワードの合計数と、その中から 2 つの提案手法で抽出したキーワードの合計数を表 3 と表 4 に示す。

表 3 実験結果の全員の合計数 (提案手法 1)

	提案手法 1 (採用)					
	A	B	E	F	G	平均
i	17%	0%	30%	5%	0%	11%
ii	43%	27%	46%	8%	6%	21%
iii	18%	9%	0%	7%	7%	9%
iv	19%	20%	0%	0%	17%	18%
v	0%	10%	0%	22%	33%	14%

表 4 実験結果の全員の合計数 (提案手法 2)

	提案手法 2					
	A	B	E	F	G	平均
i	0%	25%	20%	14%	50%	16%
ii	14%	18%	31%	8%	0%	12%
iii	18%	9%	0%	7%	14%	11%
iv	19%	40%	0%	0%	0%	18%
v	0%	10%	0%	22%	17%	11%

本研究では、表 3 と表 4 の人の興味は全くない i のキーワードの抽出率が低く、ディスカッションによって起こった新たな興味である iv のキーワードと潜在的な興味である v のキーワードの抽出率が高い提案手法 1 を採用する。アイデア創出活動において中心となる興味の iii のキーワードは、興味を中心であるため既存研究でも抽出できる興味である。そのため、iii のキーワードの抽出率をここでは議論しない。

本研究でもっとも抽出したいキーワードは潜在的な興味のキーワードであるため、v のキーワード抽出率が高くなる必要がある。今回は 14%であった。抽出率をより高めるために、同じ発表者の連続した会議や他の発表者の抽出したキーワードなどと学習機能または、比較機能を新たに検討し、提案する必要がある。また、ゼミ A とゼミ E に関しては、v のキーワード抽出率が 0%であったため、その会議の特徴を検討する必要がある。

次に、抽出した発言からでなく、発表者が選択した「有益でかつ予想外な発言」のみで同様の抽出率を評価した。その結果を表 5 に示す。

表 5 有益発言のみと抽出した発言の比較

	i	ii	iii	iv	v
抽出した発言	11%	21%	9%	18%	14%
選択した発言	9%	14%	8%	4%	11%

表 5 からは、抽出した発言より選択した発言の方が興味のない i のキーワードが少ないが、本研究で抽出すべき iv と v のキーワード抽出率も低くなっている。そのため、発表者が選択した発言のみでは精度がよくないことがわかった。そのため、表 2 の抽出率が低いことよりも、キーワード抽出の精度をあげる必要があることがわかった。さらに、発表者の選択した発言のみであってもキーワードの抽出率は低いため、キーワード抽出手法を検討する必要があることがわかった。

4. まとめと今後の課題

組織内のアイデア創出活動を支援する Web ページ推薦のために、人の興味をディスカッションより抽出する新たな手法を検討した。そして、人の潜在的な興味を抽出することができたことを述べた。今後、提案手法の長期的な実験を行い、本提案に基づいた Web ページ推薦システムを実装し、評価を行う。

参考文献

- [1] 土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロファイリング技術，人工知能学会誌 *Journal of Japanese Society for Artificial Intelligence*，Vol.19, No.3(20040501) pp. 365-372
- [2] 斎藤真理，山本則行：電子メールからの興味抽出手法と意外な Web ページと出会うきっかけを与えるエージェントシステム，日本認知科学会 *Cognitive Studies*，11(3),252-261.(Sep.2004)
- [3] 竹原幹人，大島裕明，田中克己：Blog における書き手の興味を考慮した意見情報の提示手法，情報処理学会 Vol.2005, No.67(20050713) pp. 39-45 2005-DBS-137(1)-(6)
- [4] 土田貴裕，友部博教，大平茂輝，長尾確：議事録に基づく知識活動サイクルの活性化，The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006, 3B3-1
- [5] 「言選 Web」：http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html