

語の影響度の計量による掲示板におけるキーワードの発見：
テキストマイニングのオンラインゲームへの応用

Discovering Keywords by Measuring Influence Rates in Bulletin
Board Services :

An Application of Textmining to Online Games

津田 耕平 Ruck Thawonmas 今野 優樹
立命館大学理工学研究科情報システム学専攻
知能エンターテインメント研究室
<http://www.ice.ci.ritsumeit.ac.jp/>

Kohei Tsuda Ruck Thawonmas Yuki Konno
The Graduate School of Science and Engineering,
Information Science and Systems Engineering,
Ritsumeikan University
Intelligent Computer Entertainment Lab.

概要：新しい市場調査の形として、掲示板などのオンラインコンテンツが注目を浴びている。販売後もコンテンツを変更しユーザーを留め続けなければならないオンラインゲームにおいても、ユーザーの要望を把握することは重要な課題となっており、そのために掲示板やチャットの文章が利用されることがある。本論文ではコメントの返信関係に注目することにより、電子掲示板の文章中から話題を構成するキーワードを発見する新たな手法を提案する。

キーワード：オンラインゲーム、語の影響度の計量、単語の頻度、キーワード、返信関係、ツリー型掲示板

Summary : Online contents such as bulletin board services (BBSs) have recently gained a lot of attention as new tools for market analysis. In online games, where the contents must be accordingly updated after their releases in order to have the users addicted to the games, it is important to grasp the users' demands, and for this task BBSs and chat messages have been used. In this paper, we focus on the relations among comment replies, and propose a new method for discovering keywords by measuring influence rates in BBSs.

Keywords: Online Games, Influence Rates, Occurrence Frequencies, Keywords, Reply Relations, Tree-style BBS

1 はじめに

現在, 商品販売後の追跡調査として掲示板などの調査が行われることがある [1]. 販売後も利用者の要望に合わせてコンテンツを変更していかなければならないオンラインゲーム [2] においても, ユーザーの要望を把握したり, システムのエラーを発見するために掲示板やチャットの文章が利用されることがある [3]. また電子掲示板の内容を把握しようという研究も行われている [4].

そこで本論文では, 単語の重要度を出現頻度以外の指標を加えて評価するためにコメントの返信関係を考慮したモデル [5] を参考にし, 電子掲示板の文章中から影響のあるキーワードを発見する新たな手法を提案する.

以下では, 2章で語の影響度を計量することによりキーワードを発見する手法, 3章で実験条件, 4章で実際にツリー構造を持つ電子掲示板に本手法を適用した例を示し, 5章でまとめを述べる.

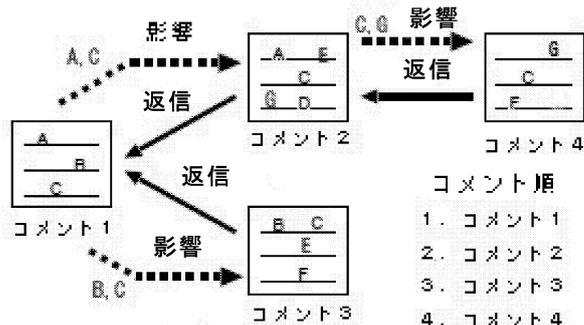


図 1: コメントチェーンの影響の例

2 語の影響度の計量

本手法は, 返信関係に注目することによりコメント中に出現する語の影響度を計量し, 文章中に存在する影響のあるキーワードを発見する.

影響のあるキーワードとは多くのコメントに影響を与えた語のことであり, あるキーワードがコメント中で用いられる際には, 記述時に参照されたコメントから影響を受けていると考える. また語の影響度とは, コメント中におけるある単語の他のコメントからの影響の度合いである.

以下に語の影響度の定義, コメント間の関係の定義, 語の影響度の計量の方法, 及び具体例を示す.

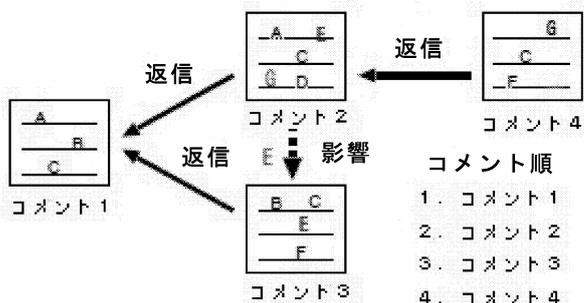


図 2: パラレルチェーンの影響の例

2.1 語の影響度

あるコメントがなされたとき, そのコメントは記述時に参照された, 他のコメントからの影響を受けている. 他のコメントからの影響は, ある文書の引用や, 話題の継承という形で現れると仮定すると, コメントからコメントへの影響は, そのコメント間で継承される語により計量可能である. 本稿では, 語の継承によりもたらされるコメント間の影響を, 語の影響度と定義し, 掲示板全体で影響度の大きい語を抽出することを目的とする. 具体的には, コメント間の関係に注目し, その関係ごとに語の影響度を計量する.

2.2 コメント間の関係

返信関係に注目することによりコメント間の関係を三つに分類し, 定義する. ここで図 1 から図 3 は以下に述べるコメントの関係における影響の例を表した図である. 図においてコメントはコメント 1, 2, 3, 4 の順で投稿され, A から G は単語を表している. また, 例における太矢印は返信関係を表しており, 点線矢印は他のコメントからの影響を各チェーンごとに表している. なお, 後述する各チェーンからの影響の範囲は $\alpha = 1$ としたときである.

[コメントチェーン]

コメントチェーン (図 1) とは, 返信の元となったコメントからの影響である. これは返信のコメントでは返信の元となったコメントと同じ語が使われることが多いという点 [5] を考慮したときの, 返信関係からの影響である.

ここで返信の元となったコメント, さらにその元となったコメントからの影響というように, どこまで遡って影響の範囲とするかという問題 ($\alpha = 1$ のときは返信の元となったコメントのみを影響の範囲とする) がある.

図 1 の例で示すと, コメントチェーンの影響を受ける語はコメント 1 の時点では存在せず, コメント 2 の時点では A, C. コメント 3 の時点では B, C. コメント 4 の時点では C, G である.

[パラレルチェーン]

次にパラレルチェーン (図 2) は同じ返信元を持つコメントの中で, 以前になされた返信からの影響である. これは返信の元となったコメントに対する以前になされた返信から影響を受けていることが多いという点を考慮したときの, 以前に派生した話題からの影響である.

ただしコメントチェーンの場合と同様に先行して行われたコメントが複数ある場合は, それらの影響をどこまで遡って影響の範囲とするかが問題になる. また, コメントチェーンの影響を受けている語は除く. なぜならコメントの投稿者が意識的に返信した元のコメント中の語からの影響を受けた可能性が高いからである.

図 2 の例で示すと, パラレルチェーンの影響を受ける語はコメント 1, 2 の時点では存在せず, コメント 3 の時点では E (C はコメントチェーンの影響を受けているので含まな

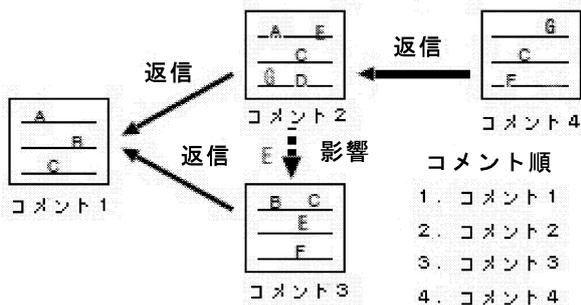


図 3: クロスチェーンの影響の例

い). コメント 4 の時点では存在しない.

[クロスチェーン]

最後にクロスチェーン (図 3) とは、直前のコメントからの影響である。これはコメントを行うものが直前のいくつかのコメントを見ていることが多いという点を考慮したときの、投稿順による影響である。

ただし、コメントチェーン、パラレルチェーンの場合と同様にどこまで遡って影響の範囲とするかという問題がある。また、コメントチェーンやパラレルチェーンの影響を受けている語を除く。なぜなら、投稿の順序よりも、意識的に行われた返信関係のほうが影響を与えた可能性が高いからである。

図 3 の例で示すと、クロスチェーンの影響を受けている語はコメント 1, 2, 3 の時点では存在せず、コメント 4 の時点では F(C はコメントチェーンの影響を受けているので含まない)。

[影響の範囲]

各チェーンの定義で述べたようにコメントチェーン、パラレルチェーン、クロスチェーンをどこまで遡って影響の範囲とするかという問題がある。

このどこまで遡るかという値を α とすると、例えば $\alpha = 3$ のときは、コメントチェーンは返信の元のコメントと、返信の元の元のコメント、さらに返信の元の元の元のコメントまでが影響の範囲となる。またパラレルチェーンは、返信の元となったコメントに対する直前の 3 つの返信のコメントまで、クロスチェーンは直前の 3 つのコメントまでがそれぞれの影響の範囲となる。

ここでパラレルチェーンの影響の元となるコメントと、コメントチェーンの影響の元となるコメントは重複することはない。またクロスチェーンの影響の範囲においては、コメントチェーン・パラレルチェーンの影響の元となるコメントは除かれる。なぜならば、クロスチェーンの定義でこれらの影響を受けた語を除く場合と同様に、意識的に行われた返信関係による影響のほうが投稿順による影響より大きいと考えられるからである。例えば直前の 3 つのうち 2 つがコメントチェーン又はパラレルチェーンの影響の元となるコメントであれば、クロスチェーンの影響の元となるコメントはひとつのみである。

[コメント k における各影響度を表す式]

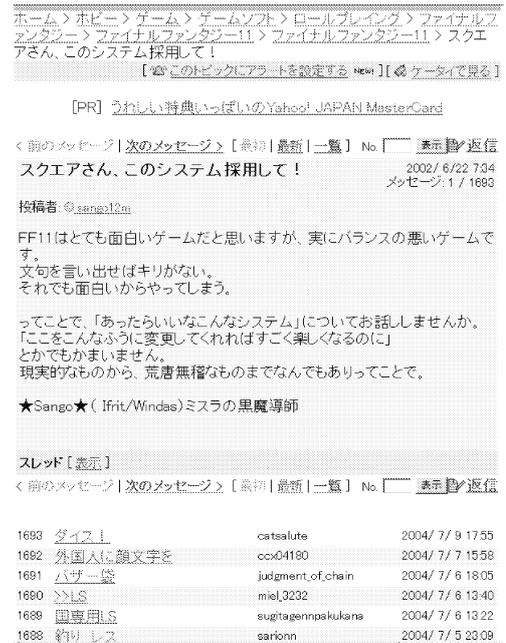


図 4: 実験に用いた Yahoo!掲示板

各チェーンの違いを考慮しないとした場合、他のコメントから影響を受けるすべての単語の影響度が同じであり、語 t のコメント k における影響度 $i_{t,k}$ は

$$i_{t,k} = \frac{1}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \quad (1)$$

となる。ここでコメント k 中の単語のうち、コメントチェーンの影響を受ける語の集合を $w_{c,k}$ 、パラレルチェーンの影響を受ける語の集合を $w_{p,k}$ 、クロスチェーンの影響を受ける語の集合を $w_{x,k}$ とする。しかし、実際には単語によりどのチェーンからの影響を受けているかは異なるのでそれぞれ異なった重みをつけることが必要になる。

よって、コメント k におけるコメントチェーン、パラレルチェーン、クロスチェーンの影響を受ける語の影響度をそれぞれ、 $i_{c,k}$ 、 $i_{p,k}$ 、 $i_{x,k}$ とし、各チェーンによる影響がそのコメントに占める割合を重みとして用いると、各影響度の値は次のように表される。

$$i_{c,k} = i_{t,k} \cdot \frac{|w_{c,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \quad (2)$$

$$i_{p,k} = i_{t,k} \cdot \frac{|w_{p,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \quad (3)$$

$$i_{x,k} = i_{t,k} \cdot \frac{|w_{x,k}|}{|w_{c,k}| + |w_{p,k}| + |w_{x,k}|} \quad (4)$$

[各影響度の大小関係]

$|w_{c,k}|$, $|w_{p,k}|$, $|w_{x,k}|$ の大小関係に注目すると、

$$|w_{c,k}| > |w_{p,k}| > |w_{x,k}| \quad (5)$$

であることが期待される。なぜならば返信の元となったコメントチェーンから最も多くの単語が用いられ、派生した話題であるパラレルチェーンの単語が次に多く用いられると考えられるからである。

表 1: $|w_{c,k}|, |w_{p,k}|, |w_{x,k}|$ の平均値 (コメント数 1697)

チェーン	影響の範囲		
	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
$ w_{c,k} $	2.6706	3.2033	3.4013
$ w_{p,k} $	0.4532	0.6694	0.8497
$ w_{x,k} $	0.2764	0.2876	0.3023

表 1 は今回の実験に用いた掲示板 (図 4) において、先程述べたことを裏付ける結果である。ただし α の値が大きくなると影響とみなす範囲が広くなり、 unnecessary 単語を抜き出してしまう可能性もある。よって掲示板ごとに重要な語を発見できる適切な値を設定しなければならない。

よって式 (5) より、式 (2), (3), (4) における $i_{t,k}$ 及び $\frac{1}{|w_{c,k}|+|w_{p,k}|+|w_{x,k}|}$ の部分は各式において共通部分であるから、

$$i_{c,k} > i_{p,k} > i_{x,k} \quad (6)$$

と期待できる。

[語 t のコメント k における影響度を表す式]

各単語はいずれかひとつの影響のみを受ける。よって語 t のコメント k における影響度 $j_{t,k}$ を表す式は

$$j_{t,k} = \begin{cases} i_{c,k} & \text{コメントチェーンの影響のとき} \\ i_{p,k} & \text{パラレルチェーンの影響のとき} \\ i_{x,k} & \text{クロスチェーンの影響のとき} \end{cases} \quad (7)$$

となる。

[語 t のコメント n までの影響度の総量]

語 t のコメント n までの影響度の総量 $I_{t,n}$ は、語 t の各コメントにおける影響度を足したものである。よって次のように表すことができる。

$$I_{t,n} = \sum_{k=1}^n j_{t,k} \quad (8)$$

2.3 語の影響度の計量の方法

語の影響度の計量の方法を順を追って述べる。

- まず、最初のコメントがなされた時点では単語は評価されない。
- 次のコメントがなされた時点でコメントチェーンの影響を受けている単語を取り出す。
- 次にパラレルチェーンの単語を取り出す。
- 最後にクロスチェーンの影響を受けている単語を取り出す。
- それぞれの単語の影響度を計量する (式 (7) を参照)。
- 2~5 を繰り返す。

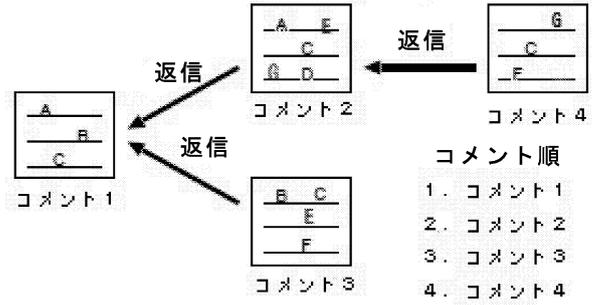


図 5: 4 件までのコメントの例

2.4 具体例

図 5 の 4 件までのコメントの例を用いて、以下に具体的な計算例を示す。

ここで図 5 における条件は図 1, 2, 3 と同様である。よって影響の範囲は $\alpha = 1$ としたときであり、各チェーンに対応する単語は 2.2 節の各チェーンの定義において挙げた例と同様である。

- コメント 1 の時点ではどの単語も評価されない。
- コメント 2 において、コメントチェーンの影響を受けている単語は A, C である。
よって $|w_{c,2}| = 2$
- コメント 2 において、パラレルチェーンの影響を受けている単語は存在しない。
よって $|w_{p,2}| = 0$
- コメント 2 において、クロスチェーンの影響を受けている単語は存在しない。
よって $|w_{x,2}| = 0$

- コメント 2 において、単語 A, C の影響度を計量する。

$$|w_{c,2}| + |w_{p,2}| + |w_{x,2}| = 2 \text{ であるから,}$$

$$\text{単語 A の影響度 } j_{A,2} = i_{c,2} = \frac{1}{2} \cdot \frac{2}{2} = \frac{1}{2}$$

$$\text{単語 C の影響度 } j_{C,2} = i_{c,2} = \frac{1}{2} \cdot \frac{2}{2} = \frac{1}{2}$$

- コメント 3, 4 においても順に 2~5 を行う。

コメント 3 において、

コメントチェーンは B, C. パラレルチェーンは E.

$$|w_{c,3}| = 2, |w_{p,3}| = 1, |w_{x,3}| = 0$$

$$|w_{c,3}| + |w_{p,3}| + |w_{x,3}| = 3 \text{ であるから}$$

$$\text{単語 B の影響度 } j_{B,3} = i_{c,3} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

$$\text{単語 C の影響度 } j_{C,3} = i_{c,3} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

$$\text{単語 E の影響度 } j_{E,3} = i_{p,3} = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

コメント 4 において、

コメントチェーンは C, G. クロスチェーンは F.

$$|w_{c,4}| = 2, |w_{p,4}| = 0, |w_{x,4}| = 1$$

$$|w_{c,4}| + |w_{p,4}| + |w_{x,4}| = 3 \text{ であるから}$$

$$\text{単語 C の影響度 } j_{C,4} = i_{c,4} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

$$\text{単語 G の影響度 } j_{G,4} = i_{c,4} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

$$\text{単語 F の影響度 } j_{F,4} = i_{x,4} = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

3 実験条件

[対象]

ツリー構造を持つ Yahoo!掲示板 (図 4) において, MMO-RPG(大規模多人数参加型オンラインロールプレイングゲーム)のプレイヤーのゲームに対する要望が題材であるトピックを選択し, 2002/6/22~2004/8/15 までの 1697 件までのコメントを対象に語の影響度 (本手法), 出現頻度, 媒介影響量 [5], TFIDF 法 [6] の各手法を適用し, 実験を行った.

[前処理]

前処理は, 文章に対し形態素解析 [7] を行い単語に区切った後, 単語のうち名詞のみ (非自立語, 数, 代名詞, 接尾語などの単体では意味のない語を除く) に注目した. 話題となるキーワードは名詞であることがほとんどであるからである.

[掲示板の内容]

掲示板では, 主に新たに導入して欲しいアイデアやシステム上の不満点が述べられている.

サービスが開始された直後は, 新たに導入して欲しいという要望が多かったが, レベルの高いユーザーが増え新しいユーザーが入ってきにくい点や, レベルの上がりすぎたユーザーからの何か新しいものがないと飽きるといった不満が次第に増えてきている.

掲示板に対する書き込みは新規にゲームを始めた人から, ゲームを長期に渡り続けている人まで多岐に渡っている. オンラインゲームはユーザー数の多さにゲームの運用会社の利益が比例する [8] ので, 様々なユーザーの意見に対応しなければならない.

4 実験

掲示板に前処理を行った結果, 1697 件までのコメントにおいて 4906 語が存在していた. 以下, 語の影響度, 出現頻度, 媒介影響量, TFIDF 法の各手法を掲示板に対して適用した結果を示す.

4.1 語の影響度

表 2 は本手法を掲示板に対し適用した際の, 1697 件の時点における影響度の総量の上位 10 位までの語である. また, $\alpha = 1, 2, 3$ と影響の範囲を変えて実験を行った.

表 3 は, 影響度の総量の上位に出現する語を影響の範囲の各値により分類した結果をまとめたものである. 以下に, 影響の範囲を変化させたときに上位に出現する語のうち, 異なつて出現する語について順に述べる.

「アイテム」

$\alpha = 2, 3$ において, 「アイテム」という語が出現していることが分かる.

この語は「バランス」とともに $\alpha = 1$ から $\alpha = 2$ に影響の範囲を広げたときに出現している. その後, さらに $\alpha = 3$ に広げた場合には, 「アイテム」は 9 位に下がっているため, $\alpha = 2$ のとき最も重要とみなされていることが分かる.

表 2: 影響度の総量の 1697 件の時点における上位 10 位までの語

順位	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
1	システム	システム	人
2	人	人	システム
3	レベル	レベル	レベル
4	ゲーム	ゲーム	ゲーム
5	装備	バランス	装備
6	経験	アイテム	魔法
7	戦士	装備	経験
8	魔法	経験	自分
9	変更	自分	アイテム
10	挑発	魔法	戦士

表 3: 表 2 に出現する語を α により分類した結果

α の値	出現する語
$\alpha = 1, 2, 3$ のすべて	システム 人 レベル ゲーム 装備 経験 魔法
$\alpha = 1, 2$ のみ	なし
$\alpha = 2, 3$ のみ	アイテム 自分
$\alpha = 1, 3$ のみ	戦士
$\alpha = 1$ のみ	変更 挑発
$\alpha = 2$ のみ	バランス
$\alpha = 3$ のみ	なし

この点については考察において後述する。コメント中では、このような「アイテム」がほしいなどのユーザーの要望があり、以下の例のように使われる重要な語であることが分かる。

[例]—————
打ち上げ花火とか信号弾とか…危険信号が送れるようなアイテムがあるといいよね、というか見てみたいw
—————

「自分」

$\alpha = 2, 3$ において、「自分」という語が出現していることが分かる。この語は影響の範囲を広げるほど上位になっている。「自分」は一般的に用いられる語とも考えられるが、「自分」が掲示板中でどのように使われているかに注目すると

[例]—————
:
自分の顔 (48)
自分のアップ (53)
:
自分の名前 (102)
自分の名前 (142)
:
※ () 内はコメントの番号

というように、他の語と一体になって同一の話題について用いられる例も数多く存在した。よって「自分」という語が影響の範囲を広げると出現したのは、ある程度まとまった範囲でこのような話題が集中しているためである。しかし、それと同時に「自分も」という同意の表現や、「自分は」のように人称代名詞として自分を指す表現として、偶然集中して用いられたためにパラレルチェーンやクロスチェーンに含まれる例も存在する。

よってこのような一般的に用いられる可能性のある語は、事前に登録することにより削除することが望ましい。今回は、影響の範囲を変化させることにより取り出す語と取り出さない語に注目するために、このような処理は行わなかった。

「戦士」

$\alpha = 1, 3$ において、「戦士」という語が出現していることが分かる。「戦士」はジョブ (職業) のひとつであり、たびたび単独で用いられている重要な語である。掲示板の中では、「戦士」というジョブについての問題点が集中して議論されていた。

「変更」

$\alpha = 1$ において、「変更」という語が出現していることが分かる。この掲示板ではゲームのさまざまな部分の変更を要求するコメントが多く、単なる不満でなく、この部分をこのように変更して欲しいと具体的な要望を求めるときに集中して用いられる重要な語である。

「挑発」

$\alpha = 1$ において、「挑発」という語が出現していることが分かる。この語は、先程述べた「戦士」のアビリティ (ジョブの持つ能力) のひとつである。よって「挑発」は「戦士」

の議論の中で、より詳細な議論に用いられる重要な語であるといえる。

「バランス」

$\alpha = 2$ において、「バランス」という語が出現していることが分かる。掲示板中では、ゲームの中の「バランス」が悪いという議論がたびたび交わされており、重要な語であるといえる。またこの語は、「システム」、「ゲーム」とともに掲示板全体において重要な語である。この点については 4.2 節において後述する。

【影響の範囲を変化させた結果に基づく考察】

表 2 をみると影響の範囲を広げていくと「人」、「自分」といった語がより上位に出現していることが分かる。「自分」という語は先程述べたように、集中して同じ話題について議論がされる際に用いられる。しかし、そうでない一般的に用いられる際にも影響の範囲を広げたときには評価されることになる。よって、一般的に用いる例が評価されないためには影響の範囲をあまり大きくしすぎないか、事前に前処理で取り除く必要がある。

また「人」については一般的に用いられることがほとんどであり、その他の手法で上位でないならば、本手法の信頼性が疑わしいと思われる。この点については 4.2 節において述べる。さらに $\alpha = 2$ のときに「バランス」や「アイテム」といった語が上位にきていることに注目する。この掲示板のテーマはゲームに対する要望であり、「アイテム」はゲームにおいて重要な概念であり、「バランス」はたびたび議論されるゲームの問題点である。このことより、これらの語は掲示板において話題の中心となる重要な語であることが分かる。

最後に $\alpha = 1, 3$ 及び $\alpha = 1$ でのみ出現する語は、「戦士」「変更」「挑発」のように「バランス」や「アイテム」といった語よりも詳細な議論で用いられる語である。2 章の冒頭で述べたように、本手法で定義する影響のあるキーワードとは多くのコメントに影響を与えた語である。よってより大きな議論で用いられる「バランス」、「アイテム」といった語が出現している $\alpha = 2$ のときに適切な影響の範囲の値であるといえる。この値が今回の実験で適切であったのは、対象とした掲示板においてコメントを投稿する際に、およそこの範囲で以前のコメントを見ているためだと思われる。実際にツリー掲示板にも様々な種類が存在することから、 α はコメントの表示方法や訪問者の特徴などによって最も適した値が異なると考えられる。

4.2 出現頻度、媒介影響量との比較

表 4 は出現頻度と媒介影響量 [5] を掲示板に対し適用した際の、1697 件の時点における上位 10 位までの語である。

表 5 は各手法で上位に出現する語を比較した結果をまとめたものである。以下に各手法で共通して出現する語、異なっており出現する語について順に述べる。

表 4: 出現頻度と媒介影響量の 1697 件の時点における上位 10 位までの語

順位	出現頻度	媒介影響量
1	人 (710)	システム (8.5836)
2	レベル (662)	人 (8.4234)
3	魔法 (561)	レベル (7.2879)
4	戦士 (523)	ゲーム (5.5790)
5	装備 (373)	魔法 (5.4448)
6	経験 (357)	挑発 (5.0806)
7	赤 (353)	戦士 (4.9205)
8	攻撃 (351)	経験 (4.7472)
9	白 (348)	装備 (3.9118)
10	挑発 (332)	黒 (3.5483)

※ () 内は各手法の値

表 5: 各手法に出現する語の比較

手法	出現する語
語の影響度 ($\alpha = 1, 2, 3$) 出現頻度 媒介影響量のすべて	人 レベル 装備 経験 魔法
語の影響度 ($\alpha = 1, 3$) 出現頻度 媒介影響量のみ	戦士
語の影響度 ($\alpha = 1$) 出現頻度 媒介影響量のみ	挑発
語の影響度 ($\alpha = 1, 2, 3$) 媒介影響量のみ	システム ゲーム
出現頻度のみ	赤 攻撃 白
媒介影響量のみ	黒
語の影響度 ($\alpha = 1$) のみ	変更
語の影響度 ($\alpha = 2$) のみ	バランス
語の影響度 ($\alpha = 2, 3$) のみ	アイテム 自分

「語の影響度, 出現頻度, 媒介影響量で共通する語」

「人」

「人」はどの手法でも取り出されている単語であるが、キーワードとはいえない。頻度が突出して多いためだと思われる。

「レベル」

「レベル」とは、キャラクターの強さを表す値である。掲示板中では、「レベル」についての問題点が議論されており、重要な語であるといえる。以下はその一例である。

[例]

経験値補正して !!

<中略>

ザコ倒しても経験値が入るよう補正してよ!

必死で「丁度」や「同じ」を倒しても規定の経験値しか貰えないんじゃない、レベル上がれば上がるほど、レベル上げがキツくなるだけじゃないか!

「装備」

「装備」とは、プレイヤーの能力を上げるためのアイテムのことであり、ゲームに関する要望で用いられる重要な語である。以下は掲示板中で交わされていた「装備」に関する議論の一部である。

[例]

装備品について スクウェアに要望メール出してみようと思ってるんですが、

<中略>

全ての装備品を装備可能にして、その上でジョブによる差別化を図って欲しい

装備品の事なのですが、現在 同じジョブ、同レベル帯だと装備品がほとんど同じ状況ですよね。もちろん、装備品の種類も増やして欲しいですが…

「経験」

「経験」はレベルを上げるために必要な「経験値」という語で用いられ、重要な語である。

「魔法」

「魔法」はゲームのキャラクターのアビリティ(能力)のひとつであり、単独で用いられる以外にも、他の語とともに用いられる重要な語である。以下にその例となるコメントのタイトルを挙げる。

[例]

白系魔法の合体魔法ほか (1126)

神聖魔法 (1166)

新移動魔法 (1560)

※ () 内はコメントの番号

「戦士」「挑発」について

「戦士」、「挑発」は 4. 1 節で述べたように、詳細な議論で用いられる重要な語である。

[語の影響度、媒介影響量で共通する語]

「システム」「ゲーム」について

表5を見ると、語の影響度や媒介影響量の項目では出現している「システム」や「ゲーム」という単語が出現頻度の項目では上位に入っていないことがわかる。以下は、この掲示板におけるトピックエディターのコメントである。トピックエディターとは、そのスレッドのテーマを決定する重要なコメントである。

[トピックエディターのコメント]—————

スクエアさん、このシステム採用して！

FF 11 はとても面白いゲームだと思いますが、実にバランスの悪いゲームです。

<中略>

ってことで、「あったらいいなこんなシステム」についてお話ししませんか。

上記のコメントから、「システム」や「ゲーム」は、この掲示板の話題の中心となる語であることが分かる。また「バランスの悪い」とここで指摘されているように「バランス」という語が重要な語であることが、ここからも読み取ることができる。これにより、出現頻度では影響のあるキーワードを抽出することは難しく、返信関係を用いて評価した本手法や媒介影響量がより正確にキーワードを取り出せていたことがわかる。

[出現頻度又は媒介影響量でのみ出現する語]

「赤」「白」「黒」「攻撃」

「赤」、「白」、「黒」はそれぞれ赤魔(法/道士)・白魔(法/道士)・黒魔(法/道士)の略で使われているが「魔法」というアビリティに関する周辺的な話題である。また「攻撃」は「攻撃間隔」「特殊攻撃」「遠距離攻撃」などの付属的な語がつく語であり、中心となっている話題に対し、周辺的に用いられる語である。

[語の影響度でのみ出現する語]

「バランス」「アイテム」「変更」「自分」

4. 1 節で述べたように、これらの語はある一定の場所で集中して用いられる語であり、コメントチェーンに加え、パラレルチェーンやクロスチェーンといった影響を導入したときに初めて出現した語である。

[出現頻度・媒介影響量の結果との比較に基づく考察]

以上の結果より、出現頻度・媒介影響度で取り出している語よりも、語の影響度($\alpha = 2$)で上位に位置している「バランス」や「アイテム」といった語のほうがより中心的な話題であり、出現頻度や媒介影響量よりも話題の中心となるキーワードを取り出しているといえる。これらの語はパラレルチェーンやクロスチェーンという概念により初めて取り出すことのできた、掲示板において重要な語であるといえる。

4.3 TFIDF 法との比較

表6はTFIDF法[6]を掲示板に対し適用した際の、1697件の時点における上位10位までの語である。結果を見ると、どれも掲示板という自由発話において個人に依存して

表6: TFIDF法の1697件の時点における10位までの語

順位	単語	tf-idf
1	ちゃち	7.4366
2	三輪車	4.9577
3	ケラ	4.4620
4	転換	3.9370
5	味	3.7183
5	あらあら	3.7183
5	お客様	3.7183
5	緊急	3.7183
5	戦記	3.7183
5	トヨタ	3.7183

いる単語を取り出している、掲示板で中心に議論されている語を抜き出せていない。以下に1位の「ちゃち」の例に触れて示す。

「ちゃち」

[例]—————
冷やかして、ちゃち入れんじゃないよ^^ (1474)
※ ()内はコメントの番号

TFIDF法の結果で、1位の「ちゃち」が出現するのは上記のコメントだけである。内容からも重要単語でないことは明らかである。

[TFIDF法の結果との比較に基づく考察]

TFIDF法は各文章を特徴づける単語が抽出される。しかし掲示板におけるコメントのように短い区切りをひとつの文章とみなすと、そのコメントだけに用いられている単語の値が極端に大きくなってしまふ。そのため今回の実験で用いたような自由発話に基づく掲示板に対しては、有効な結果は得ることができなかった。よって、本手法のほうがより重要な語を取り出しているといえる。

5 まとめ

掲示板においては、返信の元から同じ語が用いられるという特徴に加えて、一連の議論は集中して行われている。そこでそのような議論で用いられる語を評価するために、パラレルチェーン・クロスチェーンという新たなコメント間の関係を定義することにした。本論文では、それらの定義を使い、ツリー型掲示板において話題となるキーワードを発見するための新たな手法を提案した。

実験の結果から以下の考察を得ることができた。

- [5]でも述べられている通り、返信されたコメントに用いられる単語は、返信の元となったコメントからの

影響を受けている。よってこれらの単語に影響を与えた語として評価することにより、掲示板において話題の中心となっている語を取り出すことが可能である。

- さらに返信の元となったコメントに対し以前に派生した別の話題のコメントからの影響や、投稿者が直前のコメントを見ていることが多いという点に注目したときの投稿順による影響も考えられる。そのためこれらの影響を考慮することにより、出現頻度や返信の親子関係だけに注目した場合には取り出せなかったキーワードを取り出すことができる。
- また 2. 2 節で定義した影響の範囲 (α の値) の適切な値は、この値を広げすぎると不必要な単語も取り出してしまふことから、掲示板により重要な語を取り出すための適切な値を設定することが必要である。

今後は対象とする掲示板の分野における単語 (例: ゲームに関するもの) や同義語を必要に応じてあらかじめ辞書に登録しておいたり, "賛成", "反対" といった返信元に対する関係を付与することなどを考えていきたい。

謝辞

本研究は一部、平成 16 年度科学研究費補助金 (基盤研究 (C) 16500091)、及び立命館大学 21 世紀 COE プロジェクト「京都アート・エンタテインメント創成研究」一による援助を得て実施された。また、査読者の先生方から有益な御教示、御助言を頂いた。ここに記して謝意を表する。

参考文献

- [1] Ward Hanson(原著), 上原征彦, 長谷川真実(訳): インターネット・マーケティングの原理と戦略, 日本経済新聞社, 2001.
- [2] 細井浩一: ゲーム的創造力の拡張と応用, game++3, 2001 年 11 月 10 日, 京都リサーチパーク, 2001 年 11 月 10 日.
- [3] Ruck THAWONMAS: ゲームマイニングの野望: チャットルーム・掲示板から話題を特定する技術, エンタテインメントコンピューティング 2003, IPSJ Symposium Series Vol. 2003, No. 1, pp. 19-23, 2003 年 1 月 13 日~15 日.
- [4] 松尾豊, 大澤幸生, 石塚満: 電子掲示板における会話からのトピックの発見と要約, 人工知能学会全国大会, 3D1-07, 2002 年 5 月.
- [5] 松村真宏, 大澤幸生, 石塚満: テキストによるコミュニケーションにおける影響の普及モデル, 人工知能学会論文誌 第 17 巻 3 号, pp. 259-267, 2002.
- [6] G. Salton and M.J. McGill: Introduction to Modern Information Retrieval, McGraw Hill Computer Science Series, 1983.
- [7] 形態素解析システム 茶筌.
<http://chasen.naist.jp/hiki/ChaSen/>
- [8] 高韶霞 (Kasumi Kao): 急成長を遂げる台湾のオンラインゲーム事業, 知的資産創造, 2002 年 1 月号.
<http://www.nri.co.jp/opinion/chitekishisan/2002/pdf/cs20020114.pdf>

津田 耕平 2004 年立命館大学情報学科卒業。同年, 同大学理工学研究科情報システム学専攻博士課程前期課程入学。現在に至る。blog やソーシャルネットワークといった新しいコミュニケーションから得られる情報に興味がある。

Ruck Thawonmas 1987 年タイ国 Chulalongkorn 大学電気学科卒業。1990 年茨城大学大学院工学研究科情報工学専攻修士課程修了。1994 年東北大学大学院工学研究科情報工学専攻博士課程修了。博士 (工学)。同年より日立製作所, 理化学研究所, 会津大学, 高知工科大学を経て、2004 年 4 月より立命館大学情報理工学部知能情報学科教授。仮想空間における知識獲得, 双方向物語の自動生成, 教育用オンラインゲームの開発論に興味をもつ。

今野 優樹 2005 年立命館大学情報学科卒業。同年, 同大学理工学研究科情報システム学専攻博士課程前期課程入学。現在に至る。オンラインコミュニティの開発に興味がある。

連絡先: 氏名: Ruck THAWONMAS 住所: 〒 525-8577 所属: 立命館大学情報理工学部知能情報学科 電話: 077-561-5048 FAX: 077-561-5203 電子メール: ruck@ci.ritsumeai.ac.jp