

## Q&A サイトにおける ベストアンサー推定可能性の検証

石川 大介<sup>†1</sup> 栗山 和子<sup>†2</sup>  
関 洋平<sup>†3</sup> 神門 典子<sup>†1</sup>

Q&A サイトにおけるベストアンサーを計算機が推定可能か検証するために、人間の被験者によるベストアンサー推定実験を行った。ベストアンサー推定実験には Yahoo! 知恵袋データを利用し、「恋愛相談」と「パソコン」のカテゴリーから各 50 問を使用した。被験者二人による推定結果は「恋愛相談」では正解率 50%と 52%(ランダム推定結果は 34%)、「パソコン」では正解率 62%と 58%(ランダム推定結果は 38%)であり、いずれもランダム推定結果を上回った。また、被験者二人の  $\kappa$  係数は「恋愛相談」では 0.454(moderate)、「パソコン」では 0.613(substantial)であった。

### Investigation of Possibility of Best-Answer Estimation in Q&A Site

DAISUKE ISHIKAWA,<sup>†1</sup> KAZUKO KURIYAMA,<sup>†2</sup>  
YOHEI SEKI<sup>†3</sup> and NORIKO KANDO<sup>†1</sup>

In order to verify whether a computer can estimate the best answer on a Q&A site, a best-answer estimation experiment was carried out with human examinees. The Yahoo! Chiebukuro data was used for the best-answer estimation experiment, and 50 questions each were used for the categories of " Consultation of love " and " Personal computer. " The estimated result by two examinees for the correct answer rate was 50% and 52% (randomly estimated result: 34%) in " Consultation of love " and 62% and 58% (randomly estimated result: 38%) in " Personal computer "; therefore, each estimated result exceeded the randomly estimated result. Moreover, the kappa coefficient of the two examinees was 0.454 (moderate) in " Consultation of love " and 0.613 (substantial) in " Personal computer. "

### 1. はじめに

近年、Q&A サイト(「Yahoo!知恵袋」<sup>1)</sup> や「教えて!goo」<sup>2)</sup> など)を対象としたソーシャルメディア研究が活発になってきた<sup>3)4)</sup>。その一つの理由として、2007 年から「Yahoo!知恵袋」データ第一弾<sup>9)</sup> が研究利用のために配布可能となった点が挙げられる。

これら Q&A サイトには、最適な回答を選択するための枠組が備わっている。例えば、「Yahoo!知恵袋」では、最も納得・満足した回答を「ベストアンサー」として選択することができる。

このベストアンサーが選択される要因について、例えば Kim ら<sup>5)6)</sup> は「Yahoo!Answers」のベストアンサーに付けられたコメントを分析し、Socio-emotional(回答者の態度や感情面での支援)が最も多いことを報告している。

我々は、計算機がベストアンサー(最適な回答)を選択することが可能かどうかに着目している。この観点からの有用な報告はまだ見当たらない。そこで、本研究では、計算機によるベストアンサー推定可能性を検証するための予備実験として、人間の被験者によってベストアンサーを推定する実験を試みる。これにより、Kim らの結果とは異なる観点から、ベストアンサーが選択される要因を探る。

以下、本稿の構成を述べる。本節は 1 節である。2 節では、実験に使用する Yahoo!知恵袋データの概要と詳細を述べる。3 節では、被験者によるベストアンサー推定実験の方法を述べる。4 節ではその推定実験の結果を述べる。5 節では、4 節の実験結果に基づいてベストアンサーが選択される要因を考察する。そして 6 節にて本稿をまとめる。

### 2. 知恵袋データの概要

#### 2.1 ベストアンサーとは

「ベストアンサー」の説明を、Yahoo!知恵袋のヘルプ<sup>7)</sup> から抜粋したものを以下に示す。

- (1) 質問者は、もっとも納得、満足した回答を「ベストアンサー」に選べます。
- (2) また、自分でベストアンサーを選ばない場合は、ほかの利用者からの投票でベストア

<sup>†1</sup> 国立情報学研究所  
National Institute of Informatics

<sup>†2</sup> 白百合女子大学  
Shirayuri College

<sup>†3</sup> 豊橋技術科学大学  
Toyohashi University of Technology

ンサーを選ぶこともできます。

今回の実験で使用する知恵袋データ第一弾は、(1)の質問者本人によって選択されたベストアンサーだけが収録されている<sup>9)</sup>。なお、ベストアンサーの選ばれ方の詳細については図1に示す。

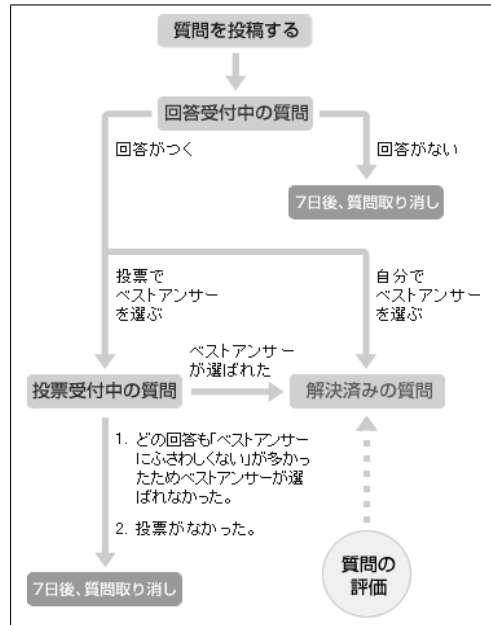


図1 ベストアンサーの選ばれ方 (Yahoo!知恵袋のヘルプ<sup>8)</sup>より引用)

## 2.2 データの詳細

Yahoo!知恵袋第一弾データの詳細は表1の通りである。以後、本稿では表1の「解決済みの質問」を質問、「ベストアンサー」と「その他の回答」とを合わせて回答と呼ぶこととする。また、Yahoo!知恵袋データ第一弾は、知恵袋データと呼ぶこととする。

## 2.3 カテゴリー毎の質問件数

知恵袋データのカテゴリー毎の質問件数を集計した。質問件数の多い順にソートしたものを表2に示す。

表1 Yahoo!知恵袋データの詳細 (仕様<sup>9)</sup>より抜粋)

収録期間	2004年4月1日～2005年10月31日
解決済みの質問	3116009件 (約916 MB)
ベストアンサー	3116008件 (約935 MB)
その他の回答	10361777件 (約2.3 GB)

表2から、質問は各カテゴリーに分散している傾向が示される。例えば、最も質問件数の多いカテゴリー「恋愛相談、人間関係の悩み」においても、総質問件数に対する質問件数の割合は6.74%である。

一方、第1位から第4位のカテゴリーにおける質問件数は、他のカテゴリーに比べてそれぞれ10万件(5%)を超えている。よってYahoo!知恵袋では、これらカテゴリーは他のカテゴリーに比べてより活発に質問が投稿されていることが推察される。

表2 カテゴリー毎の質問件数

順位	カテゴリー名 (カテゴリー ID)	質問件数 (総質問件数に対する割合)
1	恋愛相談、人間関係の悩み (2078297875)	210105 (6.74%)
2	Yahoo!知恵袋 (2078297310)	208793 (6.70%)
3	Yahoo!オークション (2078297287)	190432 (6.11%)
4	パソコン、周辺機器 (2078297424)	171848 (5.51%)
5	病気、症状、ヘルスケア (2078297872)	84394 (2.70%)
6	政治、社会問題 (2078297787)	78777 (2.52%)
7	テレビ、ラジオ (2078297526)	75194 (2.41%)
8	インターネット (2078297392)	70529 (2.26%)
9	言葉、語学 (2078297830)	68513 (2.19%)
10	動物、植物、ペット (2078297833)	61541 (1.97%)
	(全カテゴリーの質問件数の合計)	3116009 (100%)

## 2.4 平均回答数

次に、表2の各カテゴリーにおける質問に付いた回答数の平均を集計した。この結果を、表2と同じく質問件数の多い順にソートして表3に示す。

全質問における平均回答数は4.32である。これは、一つの質問に対して、(ベストアンサーを含めて)平均約4件の回答が付いていることを表している。

質問件数が第1位の「恋愛相談、人間関係の悩み」では平均回答数は6.74であり、全質問における平均回答数(4.32)よりも多い。一方、質問件数が第4位の「パソコン、周辺機

表 3 カテゴリー毎の平均回答数

質問件数の順位	カテゴリ名 (カテゴリ ID)	平均回答数
1	恋愛相談、人間関係の悩み (2078297875)	6.74
2	Yahoo!知恵袋 (2078297310)	5.43
3	Yahoo!オークション (2078297287)	6.52
4	パソコン、周辺機器 (2078297424)	2.76
5	病気、症状、ヘルスケア (2078297872)	3.43
6	政治、社会問題 (2078297787)	5.11
7	テレビ、ラジオ (2078297526)	3.25
8	インターネット (2078297392)	2.45
9	言葉、語学 (2078297830)	3.67
10	動物、植物、ペット (2078297833)	3.91
	(全質問の平均回答数)	4.32

器」では平均解答数は 2.76 であり、全質問における平均回答数 (4.32) よりも少ない。この理由の一つに、「恋愛相談、人間関係の悩み」では、誰もが自分の経験から回答することが可能であるに対して、「パソコン、周辺機器」では専らパソコンの専門知識を持つユーザが回答することができるためと考えられる。

さらに、質問件数が多い順の 100 位までの平均回答数について、平均回答数の上位 10 件と下位 10 件のカテゴリーを表 4 に示す。表 4 から、「恋愛相談、人間関係の悩み」は平均回答数の上位グループに属し、「パソコン、周辺機器」は平均回答数の下位グループに属していることが分かる。

次節では、これら「恋愛相談、人間関係の悩み」と「パソコン、周辺機器」のカテゴリーを用いて実験を行う。

### 2.5 回答数に対する質問件数

回答数に対する質問件数を表 5 に示す。例えば、表 5 の最上位の回答数が 1 とは、回答が 1 件のみ付いている質問件数の合計が 539291 件 (17.3%) であることを示す。なお、回答数が 1 個だけの質問とは、ベストアンサーしか付いていない質問のことである。

次に、総質問件数に対する累積割合に着目する。回答数が 3 件の総質問件数に対する累積割合は 50% 弱である。つまり、知恵袋データの全質問は、回答数が 1 件のもの、2 件のもの、3 件のものが、全体の半数を占めていることを指す。回答数が 6 件のものまで含めると、累積割合は約 80% である。

表 4 平均回答数の上位 10 件と下位 10 件

平均回答数の上位 10 件 (降順)			
平均回答数	質問件数の順位	カテゴリ名 (カテゴリ ID)	質問件数 (割合)
6.74	1	恋愛相談、人間関係の悩み (2078297875)	210105(6.74%)
6.52	3	Yahoo!オークション (2078297287)	190432(6.11%)
6.38	40	マナー (2078297813)	23693 (0.76%)
6.22	81	家計、節約 (2078297796)	6659 (0.21%)
6.20	24	子育ての悩み (2078297890)	35153 (1.12%)
6.18	64	ストレス (2078297874)	13274 (0.42%)
6.08	65	幼児教育、幼稚園、保育園 (2078297888)	12944(0.41%)
5.88	47	結婚 (2078297821)	20468 (0.65%)
5.51	32	小・中学校、高校 (2078297881)	26823 (0.86%)
5.50	50	新車 (2078297781)	19225 (0.61%)
平均回答数の下位 10 件 (降順)			
平均回答数	質問件数の順位	カテゴリ名 (カテゴリ ID)	質問件数 (割合)
2.37	87	これ、探してます (2078297946)	6087 (0.19%)
2.45	8	インターネット (2078297392)	70529 (2.26%)
2.54	54	Yahoo!サービス (2078297295)	15744(0.50%)
2.60	90	企業と経営 (2078297793)	5918 (0.18%)
2.62	25	携帯電話、モバイル (2078297430)	35001 (1.12%)
2.66	89	資格 (2078297903)	5944 (0.19%)
2.68	49	スポーツ (2078297757)	19322 (0.62%)
2.69	71	CM (2078297527)	10014 (0.32%)
2.70	28	ゲーム (2078297514)	30085 (0.96%)
2.76	4	パソコン、周辺機器 (2078297424)	171848 (5.51%)

## 3. ベストアンサー推定実験の方法

### 3.1 実験対象のカテゴリーの選定

今回の実験では、質問件数が上位のカテゴリーから、「(1 位) 恋愛相談、人間関係の悩み」と「(4 位) パソコン、周辺機器」とを推定実験に用いる。「恋愛相談、人間関係の悩み」と「パソコン、周辺機器」のカテゴリーは、それぞれ「社会調査型」と「情報検索型」の質問タイプに分類されると考えられる<sup>10)</sup>。

以下、本稿では「恋愛相談、人間関係の悩み」は「恋愛相談」、「パソコン、周辺機器」は「パソコン」と表記することとする。

### 3.2 推定実験用の質問の作成方法

「恋愛相談」と「パソコン」のカテゴリーから、それぞれ回答数が 2 件から 4 件付いて

表 5 回答数に対する質問件数

回答数	質問件数	総質問件数に対する割合	総質問件数に対する累積割合
1	539291	17.3 %	17.3 %
2	568982	18.2 %	35.5 %
3	502319	16.1 %	51.6 %
4	392856	12.6 %	64.2 %
5	294957	9.4 %	73.7 %
6	216797	6.9 %	80.7 %
7	159345	5.1 %	85.8 %
8	115540	3.7 %	89.5 %
9	84611	2.7 %	92.2 %
10	62028	1.9 %	94.2 %

いる質問をランダムに 50 件抽出した。つまり、回答数が 1 件のみ (ベストアンサーのみ) の質問と、回答数が 5 件以上付いた質問は、今回の実験では対象外とした。

抽出した 50 問の質問に付いた回答群はシャッフルし、質問と回答のセットを構築した。これら質問と回答のセット、各 50 問を被験者らに配布した。

### 3.3 被験者

今回の実験では、被験者は二名 (第一著者と第二著者) であり、共に情報学分野の研究者である。この被験者らには、上記方法で作成した 50 件の質問に対する回答 (2~4 個) の中から、最もベストアンサーに相応しいと考えられる回答の一つを選択してもらった。

また、比較のために、被験者らに配布した 50 問と同じものを、計算機を用いてランダムに推定するプログラムを作成した。このプログラムで推定を 10000 回行い、その平均からランダムによる推定結果を算出した。

## 4. ベストアンサー推定実験の結果

### 4.1 推定実験に要した時間

- 被験者らによる推定: (50 問の推定時間) 約 1 時間弱
- ランダムによる推定\*1: (10000 回の試行による推定時間) 約 1.27sec(99%CPU)

### 4.2 ベストアンサー推定実験の結果

ベストアンサー推定結果を表 6 に示す。表 6 の結果から、以下の点が示される。

- 被験者らの正解率は、どちらの категория においてもランダムによる正解率を上回った。
- 被験者らの正解率は、「恋愛相談」では 50% と 52%、「パソコン」では 62% と 58% であり、被験者間の差はほとんどなかった。
- むしろ、正解率の差は category による違いに生じている («恋愛相談」は約 50%、「パソコン」は約 60%)
- 一方、ランダムによる正解率はそれぞれ 34% と 38% であり、category に対してはほとんど差がなかった。

表 6 ベストアンサーの推定結果

被験者	恋愛相談	パソコン
	正解数 (正解率)	正解数 (正解率)
被験者 A	25/50 (50%)	31/50 (62%)
被験者 B	26/50 (52%)	29/50 (58%)
ランダム	17.0/50 (34%)	19.0/50 (38%)

### 4.3 被験者間の一致度

次に、表 6 の結果を用いて、被験者間の  $\kappa$  係数を算出した。 $\kappa$  係数の算出方法は以下の通りである<sup>11)</sup>。

$$\kappa \text{ 係数} = \frac{P_o - P_c}{1 - P_c} \quad (P_o: \text{観測された一致率}, P_c: \text{偶然による一致率})$$

上記の計算式によって算出された  $\kappa$  係数の判定基準は表 7 の通りである。

表 7  $\kappa$  係数の判定基準<sup>12)</sup>

$\kappa$ 係数	判定基準
0.0 ~ 0.2	slight
0.21 ~ 0.4	fair
0.41 ~ 0.6	moderate
0.61 ~ 0.8	substantial
0.81 ~ 1.0	almost perfect

$\kappa$  係数の計算式と判定基準を用いて、被験者 A と被験者 B とベストアンサーにおける二者間の値を算出した。これら結果を表 8 に示す。

表 8 から、被験者 A と被験者 B の  $\kappa$  係数は、「恋愛相談」では 0.454(moderate, 中程

\*1 使用した計算機のスペックは以下の通り。

CPU: Xeon 2.0GHz Quad Core, Mem: 16GB, OS: CentOS 5.3 64bit

度の一致)、「パソコン」では 0.613(substantial, かなりの一致)を示した。どちらのカテゴリにおいても、被験者らが選んだベストアンサーは、中程度もしくはかなりの一致が認められることが示される。

一方、被験者 A とベストアンサーおよび被験者 B とベストアンサーの  $\kappa$  係数は、いずれも 0.24 ~ 0.38 (fair) である。被験者間よりも著しく低い  $\kappa$  係数が得られた原因については、今後さらに分析する必要である。

表 8 被験者 A と被験者 B とベストアンサーにおける二者間の一致率

被験者 A と 被験者 B との一致率		
項目名	恋愛相談	パソコン
被験者間の一致数	32/50	38/50
一致率 (偶然による一致を含む)	0.64	0.76
偶然による一致率	0.34	0.38
$\kappa$ 係数	$(0.64-0.34)/(1-0.34)$ = 0.454...(moderate)	$(0.76-0.38)/(1-0.38)$ = 0.613...(substantial)
被験者 A とベストアンサーとの一致率		
項目名	恋愛相談	パソコン
被験者 A とベストアンサーの一致数	25/50	31/50
一致率 (偶然による一致を含む)	0.5	0.62
偶然による一致率	0.34	0.38
$\kappa$ 係数	$(0.5-0.34)/(1-0.34)$ = 0.242...(fair)	$(0.62-0.38)/(1-0.38)$ = 0.387...(fair)
被験者 B とベストアンサーとの一致率		
項目名	恋愛相談	パソコン
被験者 B とベストアンサーの一致数	26/50	29/50
一致率 (偶然による一致を含む)	0.52	0.58
偶然による一致率	0.34	0.38
$\kappa$ 係数	$(0.52-0.34)/(1-0.34)$ = 0.272...(fair)	$(0.58-0.38)/(1-0.38)$ = 0.322...(fair)

## 5. 考 察

本節では、被験者 A、被験者 B、ベストアンサーらの三者の回答が一致した結果を中心に考察する。

### 5.1 三者間の一致・不一致

まず、三者間の一致・不一致を集計したものを表 9 に示す。表 9 の太字の項目は、被験

表 9 三者間 (被験者 A, 被験者 B, ベストアンサー) の一致・不一致の集計

項目名	恋愛相談	パソコン
被験者 A と被験者 B と BA が一致	<b>17/50</b>	<b>24/50</b>
被験者 A と被験者 B と BA が全て不一致	1/50	0/50
被験者 A と被験者 B が一致 (BA とは不一致)	15/50	14/50
被験者 A と BA が一致 (被験者 B とは不一致)	8/50	7/50
被験者 B と BA が一致 (被験者 A とは不一致)	9/50	5/50
(合計)	50/50	50/50

者 A, 被験者 B, ベストアンサーが全て一致した件数である。以下、これらを三者一致と呼ぶこととし、「恋愛相談」の 17 件と、「パソコン」の 24 件を対象に分析した。

### 5.2 ベストアンサーを選ぶ要因

三者一致を分析するために、まずベストアンサーを選択した要因についてまとめる。4 節の推定実験を行った結果、ベストアンサーを選ぶための要因と考えられるものがいくつか挙げることができる。それらを以下にリストアップする。

- (1) 【分かりやすい】記述が分かりやすいもの
- (2) 【詳しい】詳しい説明があるもの：詳しいけど、説明がよくわからないものは BA としない。場合によっては、簡潔に書いてあるものを BA にすることもある。
- (3) 【根拠】情報の根拠となる出展 (URL) があるもの：自分の体験に基づいているものなど (自分の体験 > 他人の体験 > 体験なし, 推察)
- (4) 【丁寧】日本語が丁寧であるもの：日本語が丁寧な回答を BA に選ぶことが多かった。
- (5) 【ポジティブ】質問者にポジティブなコメントを含むもの：例えば、「がんばってください」などの表現があるもの。逆に、質問者にネガティブなコメント (罵声など) があるものは BA にしづらかった。
- (6) 【的確】的確で説得力があるもの (ただし根拠は示されていない)

### 5.3 三者一致における要因分析

前述のベストアンサーを選ぶ要因の項目を用いて、表 9 における三者一致の結果を分析した。抽出された要因を表 10 にまとめた。さらに、表 10 の結果を要因毎に集計したものを表 11 に示す。

表 11 から、どちらのカテゴリにおいても、(2)【詳しい】と、(4)【丁寧】の要因が多く (半数以上)、次いで多い要因は、「恋愛相談」では (6)【的確】、「パソコン」では (3)【根拠】である。

表 10 三者一致 (被験者 A, 被験者 B,BA) の要因分析

質問番号	要因番号
Q3	4,6
Q4	1
Q6	2,6
Q8	2,4,6
Q24	2,6
Q25	4,6
Q28	2,4,5
Q29	1,3(知人),4
Q31	2,4,6
Q32	2,4,6
Q37	2
Q39	4
Q40	2,3(うちの主人),4
Q41	1
Q45	2,4
Q48	2,5,6
Q49	3(自分)

一方, (5)【ポジティブ】は, どちらの категория においても, ほとんど見られない要因である。むしろ, 質問者に対して苦言などが含まれている回答の方が, ベストアンサーに選択されていた。この結果は Kim らの研究<sup>5)6)</sup> による結果と異なる。ただし, Kim らによる研究と本研究とでは, 要因の分類の仕方が異なるため, 両者の違いを検証するには, 今後より詳細な分析が必要となる。

## 6. おわりに

本稿では, 被験者を用いたベストアンサー推定実験の結果を報告した。本研究で示された被験者によるベストアンサー推定結果の正解率とランダムによる推定結果の正解率は, 計算機による推定を行う際の目安になると考えられる。

今後は, 本研究におけるベストアンサーを選ぶ要因の分析結果に基づいて, 計算機による

表 11 三者一致 (被験者 A, 被験者 B,BA) の要因分析の集計結果

要因	恋愛相談	パソコン
(1)【分かりやすい】	3/17	2/24
(2)【詳しい】	10/17	20/24
(3)【根拠】	3/17	7/24
(4)【丁寧】	10/17	21/24
(5)【ポジティブ】	2/17	0/24
(6)【的確】	8/17	2/24

ベストアンサー推定システムの作成を試みる。

謝辞 本研究の実施にあたって, ヤフー株式会社が国立情報学研究所に提供した「Yahoo! 知恵袋データ (第 1 弾)」を利用いたしました。

## 参 考 文 献

- 1) Yahoo!JAPAN: Yahoo!知恵袋,  
<http://chiebukuro.yahoo.co.jp/>
- 2) OKWave: 教えて!goo,  
<http://oshiete.goo.ne.jp/>
- 3) 第 1 回 知識共有コミュニティワークショップ,  
[http://www.infosocio.org/cfp\\_workshop\\_a2008.html](http://www.infosocio.org/cfp_workshop_a2008.html)
- 4) 第 2 回 知識共有コミュニティワークショップ,  
[http://infosocio.org/cfp\\_workshop\\_a2009.html](http://infosocio.org/cfp_workshop_a2009.html)
- 5) Kim, S., Oh, J. and Oh, S., Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective, *American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting*, Milwaukee, Wisconsin, ASIS&T, 2007.
- 6) Kim, S. and Oh, S., User's Relevance Criteria for Evaluating Answers in Social Q&A Site, *Journal of the American Society for Information Science and Technology*, Vol.60, No.4, pp.716-727, 2009.
- 7) Yahoo!知恵袋ヘルプ - ベストアンサーとは,  
<http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-22.html>
- 8) Yahoo!知恵袋ヘルプ - ベストアンサーを選ぶ,  
<http://help.yahoo.co.jp/help/jp/chiebukuro/qa/qa-23.html>
- 9) Yahoo!JAPAN 「Yahoo!知恵袋」データの提供について,  
<http://research.nii.ac.jp/tdc/chiebukuro.html>  
( 研究機関用データは国立情報学研究所 IDR 事務局から配布されている )

- 10) 栗山 和子, 神門 典子: Q&A サイトにおける質問と回答の分析, 情報処理学会 第 95 回 情報学基礎研究会, 2009-FI-95, 2009.
- 11) Sim, J. and Wright, C.C., The kappa statistic in reliability studies: use, interpretation, and sample size requirements, *Physical Therapy*, Vol.85, No.3, pp.257-268, 2005.
- 12) Teufel, S. and Moens, M., Summarizing scientific articles: experiments with relevance and rhetorical status, *Computational Linguistics*, Vol.28, No.4, pp.409-445, MIT Press, 2002.