

# 学生が作成した問題の出題パターンによる 類似度算出手法の提案と評価

高木 輝彦<sup>†</sup>高木 正則<sup>††</sup>勅使河原 可海<sup>†</sup>

本研究室では、2002年度から学生が協働的に作問可能なWBTシステム「CollabTest」を開発し、実験・評価を行ってきた。これまでに作成された問題数は6000問を超えた。しかし、作成された膨大な数の問題を管理することは困難であり、その多くが再利用されずに埋もれている。そこで、我々は過去の問題の有効利用を目的とし、類似問題群から動的に問題を出題するテスト出題方式を提案した。この出題方式を実現するためには、蓄積された問題を類似問題群へ分類する必要がある。そこで本論文では、まず、学生が作成した485問の問題を対象とし、問題を構成するテキスト情報の特徴を分析する。次に、この分析結果を基に蓄積された問題の類似度算出手法を検討し、問題の出題パターンによる類似度算出手法を提案する。提案手法では、問題を出題パターンにより分類し、パターンごとに語の抽出範囲を決定する。既存手法や既存システムとの比較実験の結果、類似度算出手法の有効性が検証された。

## A Proposal and Evaluation of Similar Level Calculation Method by Type of Quizzes Created by Students

Teruhiko TAKAGI<sup>†</sup>, Masanori TAKAGI<sup>††</sup>, Yoshimi TESHIGAWARA<sup>†</sup>

We have developed a Web-based on-line test system "CollabTest" which can create quizzes collaboratively by students. We have experimented and evaluated it from fiscal year 2002. The number of quizzes that have been made by students so far became 6000 or more. However, Many of those quizzes are buried without reusing. Therefore, in order to utilize past quizzes effectively we proposed the method to choose appropriate quizzes dynamically from the similar quizzes group. To achieve this method, we need to classify the past quizzes into the similar quizzes group. In this paper, we select 485 quizzes that were made by students and analyze the feature of texts that make up the quizzes. In addition, from the result of analyses we examine the method to calculate similarity level of past quizzes, and propose similarity level calculation method by type of quizzes. This proposal classifies the quizzes by type and decides the range of extraction of terms. The effectiveness of the method to calculate the similarity level is verified by comparing it with the existing method and system.

### 1. 研究の背景と目的

本研究室では、2002年度から高等教育におけるe-Learningコンテンツの不足解消と教師・学生間、学生同士のインタラクティブ性の向上を目的とした、学生が協働的に作問可能なWBTシステム「CollabTest」を開発し、実装・評価を行ってきた[1][2]。本システムでは学生が問題を作成し、作成した問題をグループ内で相互に評価（グループレビュー）する。また、教師はその問題を集めてオンラインテストを作成する。これまで過去5年間にわたり、本学の講義、本学通信教育部の講座、八王子市の小学校などで利用実験を実施し、作成された問題は6000問以上、レビューで投稿されたコメントは20000件以上となった（2008年6月時点）。この結果は、e-Learningコンテンツの不足の解消に寄与できると考えられ、また、教師・学生間、学生同士のインタラクティブ性の向上に有効であると考え

られる。

しかし、CollabTestの問題点の1つとして、作成された問題の管理・再利用が挙げられる。本システムでは学生が問題を作成するため、講義を重ねるごとに問題は増加し、それらの問題の種類や内容の把握・管理は困難になる。また、1つの講義内でオンラインテストに出題する問題数は多くても100問から200問程度であり、過去の講義で作成された多くの問題は再利用されずに埋もれている。過去の講義で作成された問題をいかに管理し、再利用するかが重要となる。

そこで、我々は過去の問題の有効利用を目的とし、類似問題群から動的に問題を出題するテスト出題方式の提案を行った[3]。図1に類似問題群からの動的なテスト出題方式を示す。学生がテストを要求すると（図1中①）、予め蓄積されている類似問題群から各問題の類似問題を自動で検索し（図1中②）、それらの問題を学生に提示する（図1中③）。

また、先行研究において、学生による類似問題の作成が可能なシステムの開発を行った。実験やアンケート

<sup>†</sup> 創価大学大学院工学研究科  
Graduate School of Engineering, Soka University

<sup>††</sup> 創価大学工学部  
Faculty of Engineering, Soka University

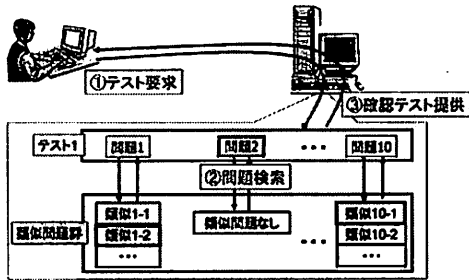


図1 類似問題群からの動的な出題

Fig. 1 Choice of appropriate quizzes from the similar quizzes group.

トの結果から類似問題の出題によりテストを繰り返し解答することによる答えの丸暗記の防止が示唆された。また、学生が作成した類似問題は類似度合に偏りが生じ、学生が作成した新規問題をコンピュータで自動分類する必要性が示唆された[4]。

学生が作成した問題を類似問題群へ自動的に分類するためには、まず、学生が作成した問題の類似度を算出しなければならない。

本研究では類似問題を「問題で問われている知識や解決の中心となる知識が類似した問題」と定義する。この知識とは分野特有の概念や法則、人物、歴史など様々であり、問題によって1つ、または、複数存在する(以下、対象知識と呼ぶ)。

本論文では、対象知識を効率よく抽出するために、問題や問題を構成するテキスト情報の種類や性質を詳細に分析する。そして、分析結果から問題の出題パターンによる類似度算出手法を提案する。さらに、既存手法や既存システムとの比較実験により問題の出題パターン分類の有効性の検証や類似度算出手法の有効性を検証する。

## 2. 類似問題群への自動分類手順

図2に類似問題群への自動分類方式を示す。CollabTestに登録されている問題はカテゴリ・サブカテゴリの2階層により管理されている。まず、蓄積されている全問題の類似問題を検索する(図2中①)。検索する対象の問題群は検索元の問題が属するサブカテゴリ内の問題群とする。

次に、全問題の検索結果からサブカテゴリごとに類似問題群を生成する(図2中②)。また、新規に作成された問題は適切な類似問題群へ分類する(図2中③)。該当する類似問題群が存在しない場合は、新規に類似問題群を生成する。

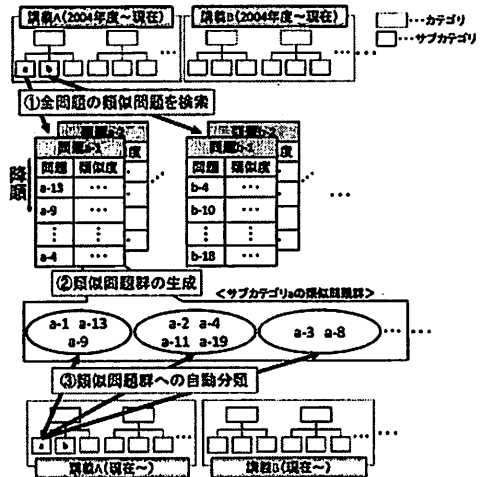


図2 類似問題群への自動分類方式

Fig. 2 Process to generate similar quizzes group.

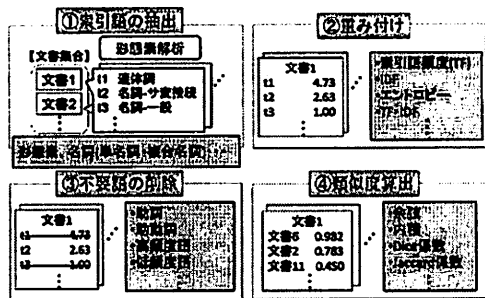


図3 一般的な類似度算出手順

Fig. 3 General process to calculate similar level.

## 3. 研究動向

近年、Web ページや文書などの類似性を判断する研究が盛んに行われている[5][6]。図3に一般的な類似度算出手順を示す。まず、各文書を形態素解析[7]により形態素に分解し、索引語を抽出する(図3中①)。形態素解析プログラムは奈良先端科学技術大学院大学で開発された「茶釜」[8]などがある。索引語の単位としては形態素や名詞などが挙げられる。

次に抽出した語に対して重み付けを行う(図3中②)。重み付け手法としては、索引語頻度(TF(Term Frequency))やIDF(Inverse Document Frequency)、それらを組み合わせたTFIDFがよく用いられる[7]。ある索引語頻度(TF)の重みを $tf$ 、総文書を $N$ 、その索引語を含む文書数を $df$ とするとTFIDFによる重み $w$ は次の式で表わされる。

$$w = tf \times \left( \log \frac{N}{df} + 1 \right) \quad (1)$$

さらに、類似度を算出する際にノイズとなる語を不要語として削除し(図3中③)、最後に、重みによりベクトルで表わされた各文書の類似度を算出する(図3中④)。類似度算出式としては、余弦や内積などがある[7]。\$x\_i, y\_i\$をそれぞれ文書 \$d\_x, d\_y\$ の索引語 \$i\$ に対する重みあるいは索引語の存在を表す2値変数、\$T\$を索引語の総数とすると余弦による類似度 \$\sigma(d\_x, d\_y)\$ は次のような式で表わされる。

$$\sigma(d_x, d_y) = \frac{\sum_{i=1}^T x_i \cdot y_i}{\sqrt{\sum_{i=1}^T x_i^2 \times \sum_{i=1}^T y_i^2}} \quad (2)$$

また、湯本らは専門用語を専門分野コーパスから自動抽出する方法を提案し実験的評価を行っている[9]。本研究における対象知識もほとんどが専門用語である。この研究では、専門用語の多くが複合語、とりわけ複合名詞であることが多いことに着目し、名詞(単名詞と複合名詞)を対象として専門用語を抽出する。

基本的なアイディアは単名詞のバイグラムから得られる単名詞の統計量を利用するという点である。バイグラムとは、扱う文字とその前後にある文字の2字の関係である。より具体的にいえば、ある単名詞が複合名詞を形成するために接続する名詞の頻度を用いる。この頻度を利用した数種類の複合名詞の重み付け手法を提案している。

単名詞 \$N\_1, N\_2, \dots, N\_L\$ がこの順で接続した複合名詞を \$CN\$ とすると、接続する頻度による重み付けの式は以下の \$LR(CN)\$ で定義される。

$$LR(CN) = \left( \prod_{i=1}^L (LN(N_i) + 1)(LR(N_i) + 1) \right)^{\frac{1}{2L}} \quad (3)$$

このとき、\$LN(N\_i)\$ は \$N\_i\$ の左方に接続した全単名詞数で \$LR(N\_i)\$ は \$N\_i\$ の右方に接続した全単名詞数である。\$CN\$ の出現頻度を \$f(CN)\$ とすると、最終的な \$CN\$ の重み \$FLR(CN)\$ は次式で表わされる。

$$FLR(CN) = f(CN) \times LR(CN) \quad (4)$$

ただし、\$f(CN)\$ は索引語頻度 (TF) とは違い、対象とする語がある複合名詞に含まれる場合はカウントしない。

この提案手法を基に「専門用語自動抽出システム」[10]やテキストマイニングツール「termmi」[11]が開発されている。termmi は各文書の語を式(4)で重み付けを行い、文書ごとに類似度を算出する。類似度算出式には余弦(式(2))を用いている。

#### 4. 本研究のアプローチ

我々は先行研究において、図3に示した一般的な類

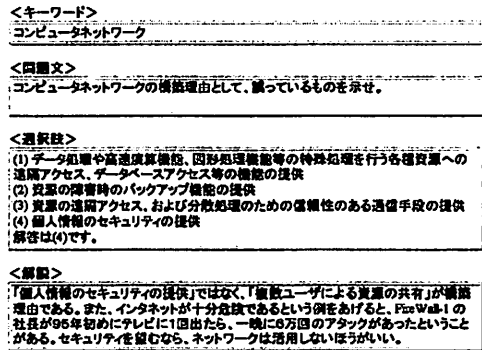


図4 CollabTestの問題情報

Fig.4 Contents of quiz on CollabTest.

似度算出手順に基づき、各ステップで最適な手法を選択し、既存手法を組み合わせた問題の類似度算出手法を提案した。また、既存システムとの比較実験により精度の向上を示した[12]。しかし、本研究で対象とするテキスト情報は学生が作成した問題であり、一般的なテキスト情報(web ページやシラバス文書など)とはその形式や特性が全く異なる。図4にCollabTestの問題入力画面を示す。学生が入力するテキスト情報は、キーワード、問題文、選択肢(正答・誤答)、解説の4つである(以下、問題情報と呼ぶ)。一連の文書とは違い、各問題情報の記入の仕方は様々であり、統一性がない。そのため、一般的な手法や既存手法では、最適な類似度算出手法を導くことは難しい。

類似度を算出する上で最も重要となるのは、対象知識をいかに抽出するかである。そこで、まず、対象知識の性質を分析し理解する必要がある。その上で、問題の類似度を算出する最適な手法を検討する。

#### 5. 学生が作成した問題の分析

##### 5.1 対象とした問題

表1に分析で対象とした問題を示す。2004年度からCollabTestを利用している講義「コンピュータネットワーク論I」の16個のサブカテゴリにおいて、2007年度までに作成された全問題485問を対象とした。この講義はCollabTestを一番長く利用している講義であり、類似問題が多数含まれると考えられたため分析の対象とした。

##### 5.2 対象知識の出現単位の分析

本節では、対象知識がどのような語の単位で出現しているかを分析する。分析方法は、1問ずつ対象知識を手動で選択・抽出し、単位ごとに分類する。表2に対象知識の単位とそれらの割合と例を示す。抽出した

表1 分析で対象とした問題

Table 1 Outline of quizzes targeted by analysis.

課題名	カテゴリ名	サブカテゴリ名	問題数
コンピュータネットワーク論I	第1章 序論	1.1 コミュニケーション	28
		1.2 コンピュータシステム	22
		1.3 分散システム	41
	第2章 コンピュータネットワーク応用技術	2.1 身近なコンピュータネットワークの応用例と家庭の情報化	33
		2.2 インターネット上で構築されている応用プログラム	28
		2.3 電子メール	28
		2.4 WWW	33
		2.5 インターネットの活用事例	24
	第3章 コンピュータネットワークとその基本機能	3.1 コンピュータネットワークの発展経緯	25
		3.2 コンピュータネットワークの基本的な考え方	45
		3.3 OSI参照モデルと基本機能	59
		3.4 TCP/IP参照モデルと基本機能	36
		3.5 ネットワークアーキテクチャの基本技術	20
	第4章 データリンク制御	4.1 誤り制御	22
	第6章 WAN技術	6.1 データ交換方式	22
		6.5 ルーティング	20

表2 対象知識の単位ごとの割合と例

Table 2 Objected knowledge type, percentage and example.

	単名詞	複合名詞	英語+日本語	英語
割合	13.6%	55.8%	5.5%	25.1%
例	ルータ	電子的コミュニケーション	OSI参照モデル	TCP

語は全部で669個であり、「単名詞」、「複合名詞」、「英語+日本語」、「英語」の全部で4種類に分類した。「単名詞」はそれ以上分割できない名詞、「複合名詞」は複数の単名詞によって構成される名詞である。本研究で扱う「英語」というのは、表2の例に示すような英単語の頭文字で表わされる専門用語や「Telnet」、「biadaptive方式」のような英単語を含む専門用語とする。また、「日本語」というのは、単名詞や複合名詞を含む日本語で表わされる名詞とする。

5.3 対象知識と出現箇所分析

本節では、対象知識が問題情報中のどこに出現しているかを分析する。解説は他の問題情報以上に記入の仕方に偏りがあるためここでは対象外とする。

しかし、対象知識はその問題の問われ方によって出現箇所が異なる。そこで、問題を出題パターンに分類した上で分析を行った。分析方法は、まず、全問題を出題パターンにより分類する。そして、出題パターンごとに対象知識が各問題情報に含まれる割合を算出する。

各出題パターンの問題数  $N_p$ 、対象知識が各問題情報に含まれていた問題数を  $N_t$  とすると、出題パターンごとに対象知識が各問題情報に含まれる割合  $P_t$  を次のような式で算出した。

表3 問題の出題パターンと割合

Table 3 Quiz types and percentage of quiz types.

パターンID	出題パターンと例	割合
Pa+	ある専門用語についての正しい例や説明などを選択する問題 例: モーリス番号の説明として正しいものを選ぶ	22.1%
Pa-	ある専門用語についての誤った例や説明などを選択する問題 例: リアルタイム処理分散システムの例として、適切でないものを選ぶ	8.0%
Pb+	ある専門用語についてその種類、属性と同じ専門用語を選択する問題 例: 以下の中で目視的コミュニケーションはどれか	2.9%
Pb-	ある専門用語についてその種類と属性と異なる専門用語を選択する問題 例: 次のうち目視的コミュニケーションでないものを選ぶ	6.6%
Pc+	ある専門用語に関連する事柄について正しい例や説明などを選択する問題 例: 分散環境におけるデータ管理に関する記述として、適切なものはどれか	4.1%
Pc-	ある専門用語に関連する事柄について誤った例や説明などを選択する問題 例: コンピュータネットワークの構築理由として誤っているものを示せ	6.0%
Pd+	ある専門用語に関連する事柄について関係する専門用語を選択する問題 例: 電話通話をするために、通しているものはどれか	5.6%
Pd-	ある専門用語に関連する事柄について関係しない専門用語を選択する問題 例: コンピュータネットワークの応用例として、銀行やコンビニエンスストアにないものはどれか	3.7%
Pe+	各選択肢に記述された専門用語とその説明の組み合わせのうち正しい組み合わせを選択する問題 例: OSI参照モデルのうちの4つの階層で、正しい説明がされているものを選ぶ	1.2%
Pe-	各選択肢に記述された専門用語とその説明の組み合わせのうち誤った組み合わせを選択する問題 例: 以下に示した語句の説明の中で、最も正しくないものはどれか	6.0%
Pf	ある専門用語についての例や説明からその専門用語を選択する問題 例: 単一のホストコンピュータを複数の端末から利用するホスト集中システムとは	26.2%
	その他	7.6%

$$P_t = \frac{N_t}{N_p} \tag{5}$$

問題の出題パターンへの分類は、対象知識の問われ方によって分類を行った。例や説明、種類や属性、関連する事柄などが問われているのか、また、正しいものを選ぶのか(+), 誤っているものを選ぶのか(-)によって細かく分類を行った。

表3に問題を出題パターンで分類した結果とそれらの例、全問題中の各出題パターンの問題の割合を示す。分類された問題はすべて専門用語に関する問題であった。「その他」というのは、計算問題やある事柄の順序問題などの問題群であり、類似問題を特定することが困難であるためここでは対象外とする。

表4に出題パターンごとに対象知識が含まれる各問題情報の割合を示す。表からも分かるように、出題パターンごとに対象知識が確実に含まれる(表4中で100%となる)問題情報を特定することができる。Pd+

表 4 出題パターンと対象知識が各問題情報に含まれる割合

Table 4 Quiz types and percentage of quizzes that objective knowledge is included in each contents of quiz.

パターンID	キーワード	問題文	正答	誤答
Pa+	48.6%	100%	1.9%	1.9%
Pa-	61.5%	100%	2.6%	2.6%
Pb+	35.7%	100%	0%	0%
Pb-	59.4%	100%	0%	0%
Pc+	5.0%	100%	0%	0%
Pc-	13.8%	96.9%	3.4%	3.4%
Pd+	7.4%	70.4%	29.6%	3.7%
Pd-	27.8%	100%	0%	0%
Pe+	0%	0%	100%	100%
Pe-	6.9%	0%	100%	100%
Pf	22.8%	0.8%	100%	0%

表 5 出題パターンごとの対象知識の出現箇所

Table 5 Contents of quiz used in each quiz types.

出題パターン	問題情報
Pa+, Pa-, Pb+, Pb-, Pc+, Pc-, Pd-	問題文
Pd+	問題文, 正答
Pe+, Pe-	正答, 誤答
Pf	正答

の問題群では問題文と正答によって 100%を満たしていた。また、Pc-の問題群では1問を除いたすべての問題で問題文に対象知識が含まれていた。

表 5 に出題パターンごとの対象知識の出現箇所を示す。以上より、出題パターンと対象知識の出現箇所に関連性があることが分かった。

#### 5.4 対象知識の出現頻度の分析

本節では、対象知識が問題情報内でどのくらい出現するのかを分析する。分析方法は、1問ずつ対象知識を手動で選択・抽出し、問題情報内での出現する頻度を調査する。

表 6 に分析結果の一例として、問題ごとに対象知識とその出現頻度を示す。表 6 から分かるように、1回しか出現しない語もあれば、複数回出現する語もあり、対象知識の出現頻度が必ずしも高くなるとは限らなかった。従って、対象知識と出現頻度に関連性はないことが分かった。

表 6 対象知識と出現頻度

Table 6 Objective knowledge and term frequency.

問題	対象知識	出現頻度	問題	対象知識	出現頻度	問題	対象知識	出現頻度
1	電子メール	3	11	メールアドレス	1	19	電子メール	1
2	電子メール	2	12	メールアドレス	2	20	電子メール	1
3	電子メール	2	13	メールアドレス	3	21	電子メール	4
4	電子メール	1	14	メールアドレス	1	22	電子メール	4
5	電子メール	2	15	メールアドレス	4		電子メール	2
6	電子メール	3	16	メールアドレス	3		電子メール	2
7	電子メール	2	17	メールアドレス	2		電子メール	2
8	電子メール	3	18	メールアドレス	2		電子メール	2
9	電子メール	2						
10	電子メール	4						

## 6. 学生が作成した問題の出題パターンによる類似度算出手法の提案

### 6.1 類似度算出手順

5.4 節の分析結果から、頻度情報を用いて索引語を抽出する既存手法では、対象知識の判断が困難であると考えられる。対象知識と頻度情報には関連性がないため、重み付け手法や不要語の削除が困難である。また、5.3 節の分析結果から、出題パターンごとに問題を分類することで対象知識の出現箇所を特定することができることが分かった。そこで、問題を出題パターンに分類したうえで類似度を算出する手法を提案する。図 5 に本研究で提案する類似度算出手順を示す。まず、各問題を出題パターンにより分類する(図 5 中①)。5.3 節で述べたように全部で 11 パターンに分類する。次に、問題の出題パターンごとに対象知識抽出に使用する問題情報を決定する(図 5 中②)。各出題パターンと問題情報の対応は表 5 の通りである。そして、問題ごとに決定された問題情報を形態素解析し、5.2 節で述べた 4 種類の単位で語を抽出する(図 5 中③)。

続いて、抽出された語に対して重み付けを行う(図 5 中④)。重み付け手法は、現在検討段階であるが、現時点では頻度情報に基づかない 2 進重みを採用している[13]。これは、抽出されたすべての語に対して重み 1 を付与する。また、キーワードは対象知識を入力するように学生に指示し、登録時の必須項目になっているので、抽出された語の中でキーワードにも含まれる語が存在すればさらに重み 1 を加える。

さらに、不要語の削除を行う(図 5 中⑤)。不要語の対象としては、上位概念を表す語と問題の決まり文句に含まれる語としている。この例の場合、電子メールに対してメールは上位概念を表すので削除する。最後に余弦(式(2))により類似度を算出する(図 5 中⑥)。

### 6.2 本提案手法の特徴

一般的な類似度算出手法は、すべてのテキスト情報

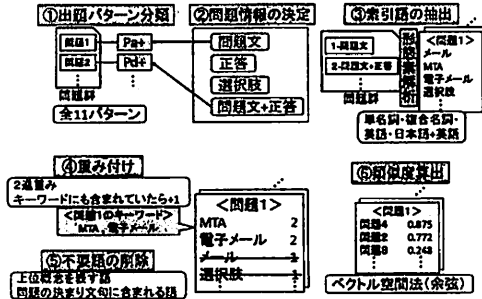


図5 出題パターンを考慮した類似度算出手順  
Fig. 5 Proposed calculating similar level method.

から語を抽出し、抽出された語の頻度情報に応じて重み付けを行い、類似度を算出する。それに対して、本提案手法では、対象となるテキスト情報（問題情報）を必要最小限のテキスト情報に絞り込んだうえで語を抽出し、頻度情報に依存しない重み付けを行い、類似度を算出する。つまり、テキスト情報（問題情報）の絞り込みにより精度の向上を図るというアプローチが本提案手法の最大の特徴となる。

## 7. 実験・評価

### 7.1 実験概要

本章では、対象知識抽出方法（図5中①～④）の有効性を検証するために実施した実験（以下、実験1）と出題パターンを考慮した問題の類似度算出手法（図5中①～⑥）の有効性を検証するために実施した実験（以下、実験2）について述べる。実験1では、出題パターンによる分類が対象知識の抽出にどのような影響を与えるかを既存の重み付け手法と比較し検証した。比較対象としては3章で述べた、TFIDF、接続頻度と出現頻度による重み付け手法を用いた。

実験2では、3章で述べた既存システム *termmi* や4章で述べた先行研究において我々が提案した類似度算出手法と比較し、全体的な類似度算出手法の精度の向上を検証する。評価尺度としてはそれぞれの実験において、再現率、適合率のマイクロ平均[7]を用いる。

実験では5章で対象とした講義とは別の講義「コンピュータネットワーク論II」で行った。この講義では2004年度からCollabTestを利用しており、2007年度までに作成された問題のうち「電子メールの運用」について作成された22問を対象とした。また、表7にそれらの問題を出題パターンに分類した結果を示す。

### 7.2 実験1：対象知識抽出方法の比較実験

実験1の実験手順を以下に示す。なお、各手順を1問ずつ全22問に対して行う。

表7 各問題と出題パターン

Table 7 Each quizzes and quiz types.

問題1	Pe-	問題7	Pf	問題13	Pc+	問題19	Pc-
問題2	Pf	問題8	Pc-	問題14	Pd+	問題20	Pa-
問題3	Pa-	問題9	Pa+	問題15	Pa-	問題21	Pe-
問題4	Pa+	問題10	Pd+	問題16	Pd+	問題22	Pb-
問題5	Pa+	問題11	Pa-	問題17	Pa-		
問題6	Pe-	問題12	Pe-	問題18	Pa-		

(1) 教員が手動で対象知識を抽出

情報系の科目の知識を十分に持っている教員（共著者）に手動で対象知識を抽出してもらった。なお、抽出する前に対象知識の説明を行った。

(2) 提案手法による語の抽出・重み付け

図5中①～④に示した語の抽出と重み付けを行う。また、抽出された語は重みの大きい順に整列する。

(3) 既存手法による語の抽出・重み付け

全問題情報から語を抽出し、式(1)と式(4)によりそれぞれ重み付けを行う。抽出する語の単位は提案手法と同じとした。また、抽出された語は重みの大きい順に整列する。

(4) 教員が抽出した語の数を閾値とし、手順(2)と手順(3)の結果から語を抽出

閾値を設けるのは教員が選択した語が上位に抽出されているかどうかを評価するためである。教員が抽出した語が2つに対し、各手法で抽出した語の2番目と3番目が同じ重みであれば3番目も含めて抽出することとした。

(5) 再現率、適合率のマイクロ平均の算出

$i$  を問題番号とし、問題  $i$  から教員が抽出した語の数を  $A_i$ 、問題  $i$  から各手法で抽出した語の数を  $B_i$ 、問題  $i$  から各手法で抽出された語のうち教員が抽出した語と同じ語の数を  $C_i$  とすると、再現率、適合率のマイクロ平均  $\bar{R}$ 、 $\bar{P}$  の式は以下のようになる。

$$\bar{R} = \frac{\sum_{i=1}^{22} C_i}{\sum_{i=1}^{22} A_i} \quad (6)$$

$$\bar{P} = \frac{\sum_{i=1}^{22} C_i}{\sum_{i=1}^{22} B_i} \quad (7)$$

この場合、再現率  $\bar{R}$  は教員が対象知識と判断し、抽出した語に適合する語をまったく抽出しているかを表し、適合率  $\bar{P}$  は教員が対象知識と判断し、抽出した語に適合する語だけを抽出しているかを表す。

以上の実験手順から得られた評価結果を表9に示す。

表9 実験1の結果

Table 9 Result of experiment 1.

重み付け手法	再現率	適合率
本提案手法	0.77	0.42
接続頻度と出現頻度	0.12	0.12
TFIDF	0.19	0.19

本提案手法は接続頻度と出現頻度による重み付けと比較して、再現率、適合率の値がそれぞれ約6.4倍、3.5倍向上した。また、TFIDFと比較しても、それぞれの値が約4.1倍、2.2倍向上した。

### 7.3 実験2：類似度算出手法の比較実験

実験2の実験手順を以下に示す。

#### (1) 実験1と同じ教員による類似問題群への手動分類

実験1と同じ教員が、22問の問題を類似問題群へ手動で分類した。分類の結果、教師の判断による類似問題を含む問題は全部で11問存在し、5つのグループに分類された。表10に教員による分類結果を示す。

#### (2) 各手法による類似問題の検索

本提案手法、先行研究の手法、termmiによる手法で、類似問題を検索する。手順(1)で分類された11問の類似問題を1問ずつ検索する。類似問題を検索する対象は、実験で対象としている問題22問となる。検索された問題は類似度の高い順に整列する。

#### (3) 問題の抽出

教員が選択した類似問題の数を閾値とし、手順(2)の検索結果から問題を抽出する。閾値を設けるのは、教員が類似問題であると判断した問題が上位に検索されているかどうかを評価するためである。例えば、表10の類似問題群A(問題1, 問題6, 問題21)の閾値はそれぞれ2となる。この場合、教員が選択した問題が2つに対し、もし、2番目と3番目が同じ類似度であれば3番目も含めて抽出する。

(4) 再現率、適合率、F尺度のマイクロ平均を算出  
問題*i*から教員が選択した類似問題の数を $A_i$ 、問題*i*から各手法により抽出された問題の数を $B_i$ 、問題*i*から抽出された問題のうち教員が選択した問題と同じ問題の数を $C_i$ とし、式(6)と式(7)に適用させた。この場合、再現率 $\bar{R}$ は教員が類似問題と判断し、選択した問題に適合する問題をもれなく抽出しているかを表し、適合率 $\bar{P}$ は教員が類似問題と判断し、選択した問題に適合する問題だけを抽出しているかを表す。

この実験では、教員が判断した類似問題を含む問題は全部で11問であるため、その他の問題の $A_i$   $B_i$   $C_i$

表10 教員による分類結果

Table 10 Classified result by teacher.

A	B	C	D	E
問題1	問題2	問題4	問題5	問題15
問題6	問題11	問題8	問題14	問題19
問題21				

表11 実験2の結果

Table 11 Result of experiment 2.

類似度算出手法	再現率	適合率
本提案手法	0.71	0.71
先行研究における提案手法	0.36	0.31
termmi	0.07	0.07

の値は全て0となる。

以上の実験手順から得られた実験結果を表10に示す。本提案手法は先行研究における提案手法と比較して、再現率、適合率の値がそれぞれ約2.0倍、2.3倍向上した。また、termmiと比較しても、それぞれの値が約10倍向上した。

### 7.4 実験の考察

#### 実験1:

実験結果から、本提案手法は2進重みによるほぼ均等な重み付け手法を採用したにもかかわらず精度が向上した。このことから、出題パターン分類による語の抽出が非常に効果的であったと考えられる。また、出題パターン分類により抽出する語を必要最小限に絞り込むことができ、複雑な重み付けを行う必要がなくなったと考えられる。

#### 実験2:

実験結果から、本提案手法は先行研究において検討した手法より精度が向上しており、既存手法の組み合わせから最適な手法を導くことはできないことが分かった。特に、先行研究の手法では、重み付けや不要語の削除において頻度情報を用いた手法を採用していたため、問題の類似度算出において頻度情報用いることは有効的でないと考えられる。

また、本提案手法では教員が選択した類似問題が上位に検索されていたことから、次のステップである類似問題群の生成が容易になると考えられる。従って、類似度算出手法の有効性を示すことができたといえる。

## 8. 今後の課題

重み付け、不要語、類似度算出式に関して更なる検討を行う。出題パターンごとに使用する問題情報を特定することができたので、それらの問題情報の分析を

行い、出題パターンごとに最適な手法を提案する。また、他講義の問題も対象とし、出題パターンによる問題情報や索引語の単位の妥当性を検証する。

本提案手法の問題点として、出題パターンによる分類を自動で行うことが困難であることが予想される。そこで、自動化の検討を行うと共に、出題パターンによる分類を行わない場合の類似度算出手法も同時に検討する。さらに対象知識を詳細に分析し、重み付け手法や不要語の削除方法を検討する。

また、次のステップである類似問題群生成 (図 2 中②) の検討を行う。既存のクラスタリング手法の調査や我々が提案する類似度算出手法による検索結果の特性を分析し、最適な手法を検討する。

## 9. まとめ

本論文では、類似問題群へ自動的に分類するうえで重要となる、学生が作成した問題の類似度算出手法を検討・提案し、実験・評価を行った。検討するにあたって、4年間で蓄積された学生作成問題 485 問を分析し、分析結果から問題の出題パターンによる類似度算出手法の提案を行った。

対象知識抽出方法の比較実験 (実験 1) から、出題パターンによる分類の有効性を示し、出題パターンにより複雑な重み付けを行わなくてもよいことが分かった。

類似度算出手法の比較実験 (実験 2) から、問題の出題パターンによる類似度算出手法の有効性を示すことができた。また、次のステップである類似問題群生成が容易になったことが分かった。さらに、問題を対象とした類似度を算出する際に問題を対象とした場合、語の頻度情報を使用することができないことを再確認した。

## 10. 参考文献

- [1] 高木正則, 田中充, 勅使河原可海: 学生による問題作成およびその相互評価を可能とする協調学習型 WBT システム, 情報処理学会論文誌, Vol.48, No3, pp.1532-1545, 2007.3.
- [2] 高木正則, 田中充, 勅使河原可海: 協働的に作問する過程で競争可能なオンラインテストシステムの実装と評価, 教育システム情報学会誌, Vol.24, No1, pp.13-25, 2007.
- [3] 高木正則, 田中充, 勅使河原可海: 学生が作成した類似問題を利用するオンラインテスト機能の実装・評価, 情報処理学会情報教育シンポジウム (SSS2003), 同シンポジウム論文集, pp.27-34, 2003.
- [4] Masanori TAKAGI and Yoshimi TESHIGAWARA: A WBT System Enabling to Create New or Similar Quizzes Collaboratively by Students, The Second IASTED International Conference on Education and Technology(ICET2006), Calgary(Canada), Proceedings of ICET2006, pp.263-268, 2006.
- [5] 八太絵美, 福本徹, 横山節雄, 赤堀侃司: “文書間の類似度に基づく論文検索システムの開発と評価”, 日本教育工学会研究報告集, pp.91-96(2002).
- [6] 野沢孝之, 井田正明, 芳鐘冬樹, 宮崎和光, 喜多一: シラバスの文書クラスタリングに基づくカリキュラム分析システムの構築, 情報処理学会論文誌, Vol.46, No1, pp.289-300, 2005.1.
- [7] 徳永健伸: 情報検索と言語処理, 東京大学出版会 (1999).
- [8] 松本祐治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 形態素解析システム『茶釜』version 2.0 使用説明書 第二版, NAIST Technical Report, NAIST-IS-TR99012, 奈良先端科学技術大学院大学 (1999).
- [9] 湯本紘彰, 森辰則, 中川裕志: 出現頻度と接続頻度に基づく専門用語抽出, 自然言語処理, Vol.10, No.1, pp.27-45, 2003.1.
- [10] 東京大学中川研究室・横浜国立大学森研究室: 専門用語自動抽出システム.
- [11] Windows 用テキストマイニングツール “termmi”: <http://gensen.dl.itc.u-tokyo.ac.jp/termmi.html>.
- [12] 高木輝彦, 高木正則, 勅使河原可海: 学生が作成した問題の類似性に基づいた自動分類方式の提案, 情報処理学会第 70 回全国大会, 同大会論文集, Vol.4, pp.687-688, 2008.3.
- [13] 北研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版 (2002).