数理解析研究所講究録
第1734巻 2011年 1-8

学習者個々の到達度を考慮した傾向別課題作成法の提案
A method for generating student assignments based on the learners' achievement

近畿大学　加島 智子　(Tomoko Kashima)
Kinki University

広島工業大学　松本　慎平　(Shimpei Matsumoto)
Hiroshima Institute of Technology

近畿大学　井原 辰彦　(Tatsuhiko Ihara)
Kinki University

関西学院大学　石井 博昭 (Hiroaki Ishii)
Kwansei Gakuin University

概要

「講義」，「e-Learning」，「プッシュ型問題配信」を組み合わせた学習環境を構築するため，本稿は数理的手法に注目し，学習者個々の理解度や達成度を定量的に評価した課題自動配信の可能性について検討した．著者らはこれまで，即時かつ能動的にメールを転送することが可能な携帯電話での学習を想定して，プッシュ型e-Learningの開発を進めてきた．プッシュ型e-Learningでは，利用者側からの問い合わせに関わらず短時間で回答可能な練習問題が自動的に配信されるため，身近で手軽な学習機会を創出することができるのではないかと期待している．特に，能動的な自学習慣の身に付いていない学習者に対して，継続的な学習時間の確保や学習の意識付けを中心とした時間外学習を支援できるのではないかと考えている．今日までの取り組みにより，e-Learningの基本機能の開発と実装が達成されている．今後の課題として，本研究では，e-Learningに蓄積された学習履歴情報の活用を計画している．具体的には，協調フィルタリングに基づく情報推薦処理を問題配信に応用することで，各学習者に対して，分野毎の習熟度や傾向に合わせた問題の提供を考えている．本稿は，問題配信アルゴリズムの構築に向け，テキストマイニング処理に基づいた内容ベースフィルタリングを取り上げ，分析結果を踏まえた考察と今後の可能性を述べる．

キーワード：e-Learning，協調フィルタリング，内容ベースフィルタリング，テキスト解析

1．はじめに

近年，学習者の価値観の多様化を背景にして，次世代の学校教育を目指す取り組みや議論が活発になってきた．情報技術の進展により，ディジタル教科書やe-Learningの可能性は広がり，Webの教育への活用も普及し企業や教育機関での利用が増えてきている．e-Learningの普及を中核とした情報技術の発展と共に，優れた学習支援環境の構築が進められている．しかし，従来構築されてきた一般的なe-Learningの基本機能だけでは，利用者が能動的に問い合わせなければ学習コンテンツを利用することができない．そのため，従来のe-Learningでは，時間経過と共に一部の学習者のみの利用に限られてしまうという問題が指摘されてお

り，本来支援すべき学習者の利用を促す仕組みの構築は急務の課題とされている．従来のe-Learningが抱える大きな要因として，本研究は，システムにログインするまでの敷居の高さにあるのではないかと考えている．通常，一般的なe-Learningの利用者は，公開されている各種サービスや学習教材を利用するためには，利用者側から能動的に問い合わせなければならない．したがって，既存のe-Learningが想定している利用者は，積極的に勉強する習慣が身に付いていると考えられる．そこで，本研究では，e-Learningで支援すべき対象として，e-Learningを継続的に利用できない学習者，すなわち，自学習慣が十分に身に付いていない学習者を想定したe-Learningの開発に着手した．システム利用者の選定と共に，本研究は学習形態の多様化に注目した．具体的には，現在，座学に象徴される指導者中心の従来の教育スタイルから，学習者個々が主体的に学習に取り組み，各自の到達度に応じたオーダーメイド教育スタイルへ変化が要求されていると考える．その解決策として，本研究は，全ての利用者に対して同様の内容の教材を提供している点に改善の余地があるのではないかと考える．以上2点を同時に解決可能な方法として，各学習者の到達度に応じて，システム側が問題を能動的に配信することが有効ではないかと考え，システムの開発とその機能の実装を進めた．その成果としてプッシュ型e-Learningを構築し，現在，基本機能の開発を完了させた．

我々が開発を進めているプッシュ型e-Learningは，携帯電話のメール機能の利用を前提としたものである．このことは，利用者側の端末を検討するに当たって，携帯電話の普及率の高さや身近で手軽なインターネット環境を提供可能なことを踏まえた結果による．近年，電子メールは我々にとって大変身近な存在であり，重要なコミュニケーションツールのひとつとして日常生活の様々な場面で用いられている．アイシェアが行った2009年の調査によると，最も利用した携帯電話の機能は52.2%で電子メールであると報告されている．また，ネットショップ等の受注メールや，メールマガジン，クーポン，ケータイ小説など，電子メールは個人対個人の対話支援ツールに留まらず，利用者の目的に合わせて多様化している．電子メールは日常生活に密着した最も身近な情報伝達媒体であると考えられるため，学習機会を手軽に創出可能であると考えた．

本稿は，これまでの成果を報告すると共に，対面式指導者中心の「講義」，学習者中心に行われる補助教材としての「e-Learning」，学習者個々の達成度を考慮し，手軽に学習機会を創出するための「プッシュ型メールシステム」を組み合わせた学習法を考える．e-Learningで行われる時間外課題や課題の成績データ，eメールプッシュ型システムに返信されたデータを分析することにより，各学習者の習熟度や解答傾向に合わせた問題を生成・配信することにより，学習者に対して，学習すべき分野の自学を継続的に促す．

2. e-Learning

近年，e-Learningの導入事例や有効性が報告されている [1]．e-Learning環境の構築や運用は，MoodleやNetCommonsといったオープンソースパッケージを利用することにより，大学を筆頭とした高等教育機関で数多く導入されており，多くの実践事例は様々な学術雑誌で報告されている．e-Learningは時間や場所といった物理的な環境制約を超越した学習環境を提供することから，その利用は年々増え続けており，Moodleの公式Webサイトによると，2010年11月の時点において，49,659サイトで利用されていると報告されている．教育支援を目的に構築された情報システムは，総称してe-Learningと呼ばれている[2]．e-Learning環境の構築や運用は，大学などの教育機関において既に不可欠な存在として認識されおり[3]，WebCTやBlackboardなどのソフトウェアや[4]，オープンソースパッケージを利用することで，サーバ構築に関する基礎知識を有してさえいれば誰でも簡単にe-Learningを構築することができる．教室内で行われていた講義をe-Learningによって代替することで，時間や場所といった物理的な環境の制約を受けずに学習可能となる．少人数グループでの学習に限定されてきた意見交換や学習成果の報告であっても，e-Learnig, twitterなどの最新のウェブサービスの利用や，Webカメラなどマルチメディア機器やSkypeなどのICTソフトウェアを組み合わせた方法により，インターネット環境での学習環境の提供とその支援が既に実現され始めている[5][6]

e-Learningを「ネットワークとPCを利用した学習形態」のみに限定せず，携帯電話・端末，学習用(ゲーム)ソフトウェア，衛星通信等を使用した学習形態を含めた場合，矢野経済研究所の2008年度の調査によると，国内e-Learning市場全体の市場規模は約1300億円と報告されている．ミック経済研究所の2007年の調査によると，ゲーム機向け学習ソフトウェア市場を除いた場合であっても，企業研修，予備校，学習塾，外国語会話，資格試験，生涯学習などの市場を総合した場合，全体で約5000億円の規模になると計上している．大学など教育機関のみならず，従業員教育や研修費用の抑制傾向にある企業にあっては，研修コスト削減のためにe-Learningへの切り替えが期待されている．新入社員研修や管理職研修，あるいは法規制などで実施が義務付けられているコンプライアンス研修などに対しては，e-Learningを適用する事例が増える傾向にある．IDC Japanが行った2009年国内法人向けe-Learning市場の市場予測によると，今後2008年から2013年まで年間平均成長率5.7%で市場規模が拡大すると予測されている．e-Learning普及の流れは，今後もより一層加速することが予測されている．

e-Learningの中でも，携帯情報端末を取り入れたユビキタスラーニングは，最近特に注目を集めている話題のひとつである．例えば，Nintendo DSでは数多くの学習用ソフトが販売されており，講義に導入された例は数多く報告されている[7]．携帯電話やPDAの通信機能を活用すれば，音声・映像配信やウェブシステ

ムの利用など，多様なサービスの提供が可能であるため，携帯端末を用いた学習の可能性とその有効性は複数の学術雑誌で報告され始めている[8].

特に，携帯電話のインターネット通信を利用したe-Learningは，その普及率の高さから積極的に構築されている．携帯電話を利用したインターネット通信について，Garbagenews.comが2008年12月に行った調査によると，6歳以上の年齢全体で見ると55.9%と非常に高く，13歳から49歳までの利用率は70%以上である．中でも20・30代に限っては，その利用率は80%を越えるなど，携帯電話によるインターネット利用率の高さが示されている．さらに，MMD研究所が2009年7月に実施した調査によると，2台以上の携帯電話保有者は23.8%であると報告されている．以上の背景からも，携帯電話を利用したe-Learningの今後の可能性の高さを知ることができる．その証拠に，携帯電話の利用を想定した多くのサービスが運営され，多機能で多様なコンテンツが提供されていることを確認できる．

2.1 MOODLEの機能

Moodleはオープンソースで開発されているe-Learningパッケージである．Moodleには様々な機能が標準で備わっており，具体的には，自動採点時間外課題，教材ページ作成・公開，フォーラム(BBS)，課題，成績などの機能が容易に利用可能である．その他にも，チャットやメッセージ送信，レッスン，投票，用語集など様々な機能が利用できる．アイテム分析の項目では，個々の問題に関するパフォーマンスの分析及び判定が自動に表示される．具体的には，個々の問題の難易度（ファシリティ指標），標準偏差の他，識別指数，判別係数も表示される．そこで，本研究では，学習者の学習達成度の指標であり，時間外課題の難易度を得ることにより学習者の分類，問題の分析に用いるためにアイテム分析に注目した．

アイテム分析

Moodelの問題分析表には，評価手段として，問題に関するパフォーマンスの分析や問題の難易度を判定することができ，学習者の時間外課題データが自動的に計算される．使用される統計変数は伝統的なテスト理論によって計算される．本稿では，統計変数の代表例として，正解，ファシリティ，標準偏差，識別指数，判別係数の計算法について説明する．

ファシリティ指標は，時間外課題の学習者にとっての問題の難易度を示す基準であり，全学習者の平均点を問題の最高点で割ることにより求められる．問題が正解/不正解のカテゴリに分けられる場合，この変数は正しく答えた学習者の割合を示す．標準偏差(*SD*)とは，変数は回答者の答えの広がり方を測定しており，全ての学習者が同じ答えを行った場合，*SD=0*となる．*SD*は各問題における一部のサンプル(到達/最大)の統計的標準偏差として計算する．識別指数(*DI*)とは，「優秀な学習者」vs.「優秀でない学習者」の各問題のパフォーマンスに関する大まかな指標を提示する．この変数は学習者を上位3分の1と下位3分の1に分割し，上位の学習者の獲得得点の合計から下位の学習者の獲得得点との差を全体の数で割ることで計算し，分析アイテムの平均点は上下学習者のグループのために計算され，平均グループは除外される．この変数は+1から-1の間の値を取り，もしこの指標が0.0以下になった場合，弱い学習者が強い学習者よりも正解数が多いことを示す．そのような問題は役に立たないと見なし，破棄すべき問題と考える．実際，このような問題は，時間外課題全体の評点の精度を下げる．判別係数 (*DC*)とは，優秀でない学習から優秀な学習者を分離するもう1つの指標である．判別係数は問題の点数および課題全体の点数の偏差値合計を問題の回答数，問題の部分的な点数の標準偏差と全体の点数の標準偏差の3つの値を掛け合わせた値で割ることにより求める．この変数は +1から -1の間の値を取る．正の値は優秀な学習者の方が答えることができたことを意味し，一方，負の値は優秀でない学習者がよく答えることができたことを意味する．判別係数が負の問題は優秀な学習者が間違って答えたことを意味し，従って最も優秀な学習者に対する不利な条件の問題となる．このような問題は避けるべきであると判断される．この問題に関して，すべての学習者が全く同じ評点の場合，問題の部分的な点数の標準偏差はゼロとなり，*DC*は未定義となる点に留意が必要である．この場合，*DC*は-999.00と表示される．識別指数に対して判別係数が優れている点は判別係数が極端な上位1/3や下位1/3の情報ではなく，学習者全体の情報を使用することである．従って，この変数はパフォーマンスに関して，より繊細に分析することになる．

3.開発システムの概要

プッシュ型e-Learningでは携帯電話の電子メールサービスの利用を想定しているため，練習問題を自動的に各学習者の携帯電話に配信することができる．各学習者は配信されたメールに対して，本文の指定箇所に解答を記述し，返信するだけで，採点結果を即座に受けることができる．また，専用サイトにアクセスすることで，解答履歴，各分野の達成度，自身の弱点を把握することができる．採点や解説の配信は，全て自動で行われる．全ての利用者の解答履歴や成績はデータベースに蓄積されており，教員は各利用者の理解度や達成度を適宜把握可能なため，講義の進行計画や個別指導の参考にできる．

データベースには，利用者の情報，過去問題の情報，難易度，分野，正解，解説などの情報が蓄積されている．本システムに利用者登録した学習者は，決められた日時に練習問題が決められた数だけ配信される．配信されるメールには一意なハッシュ値が件名に付与されており，学習者は受信した電子メールにハッシュ値をつけたまま本文に解答(記号)を添えて送信元のメールアドレスに返信することで，Webアプリケーションが採点を行い，正否，解説を即座に受けることができる．本機能は，我々が既に提案したメール利用支援インタフェースにより実現されている．メール利用支援インタフェースは，Javaにより開発されたサーバソフトウェアである．まず，メール利用支援インタフェースがメールサーバから定期的にメールデータを取得する．次に，データベース内のメールをJavaアプリケーションが非同期に採点することによって，採点及び自動返信機能が実装されている．採点の際，他利用者との正答率などの情報を用いて，問題の難易度の調整，学習者自身の正答率，達成度，弱点を把握できる．

学習者の理解度に関する情報は，教員に配信されるようになっている．なお，配信日時・回数などの各種項目は，電子メールを決められた形式で送信することで設定できる．本機能は，メールの文面を解析して，データベースへの情報登録や更新することで実装されている．以上の機能はMoodleと連携可能である．

3.1傾向別課題作成の推薦アルゴリズム

Resnick らによると，推薦システムの実行過程にはOIP(output-Input-Process)と呼ばれる3つのステップがある[9].

Step1：データ入力

通常，マーケティングなどで利用される場合，「食べたい」，「欲しい」，「好き」などという情報をシステムの利用者が入力するステップとなっており，この情報獲得には，「明示的獲得(直接学習者に尋ねる)」方法と，「暗黙的獲得（web上のデータなどから予測）」の2つの方法がある．本研究のような学習データの場合，明示的獲得では，学習者本人に対して，直接得手不得手を尋ねることで情報を収集することになる．明示的獲得の場合，データ内容の正確性が保証され，同時に，学習者自身にとっても，収集された情報は納得できるものであることが多い．しかし，質問に逐一解答することは学習者にとって面倒な作業であるため，大規模な情報収集は極めて困難であると予測される．そこで，本研究は大量のデータを収集することが可能な暗黙的獲得の方法を用いる．本研究では，Moodleによりe-Learning環境を構築することで，時間外課題の課題結果から「得意」，「苦手」を判別し，学習者の能動的な手間を削減する．

Step2：得意度の推測方法

問題推薦を行う対象とする学習者Aの成績データとその他の学習者の成績データより，対象学習者Aの学習の得意度，今後獲得する点数を予測する．推薦を行う手法には大きく分けて2つの方法があり，具体的に「内容ベースフィルタリング（問題の特徴を利用するもの）」と「協調フィルタリング（他人の解答データを利用するもの）」である．商品の嗜好の推測とは異なり，学習データにはセレンディピティ（単に得手不得手の問題を推薦するだけでなく，問題を忘れた頃に出題するような意外性のある推薦）も考慮が必要であると考える．この場合，協調フィルタリングの方が，本人の過去の履歴を見る内容ベースフィルタリングよりもセレンディピティを起こしやすいと考えられる．しかし，過去の学習履歴が少ない場合，どの学習者と嗜好データが似ているか予測が難しいため，本人による情報の入力による内容ベースフィルタリングの方が有利であると考えられる．よって，今回は両者を混合したハイブリッド型を用いる．問題の特徴を利用し，更に同じ分野が得意としている学習者，苦手としている学習者の情報を用いて対象とする学習者Aが得意とするか苦手とするか決定する．

内容ベースフィルタリングによる問題の分類

ここではMoodleの問題分類を行う．問題の属性として，ファシリティ，標準偏差，識別指数，判別係数を用い，クラスタ分析により分類を行う．クラスタ分析は，似ているデータ同士は同じ振る舞いをするという前提のもとに，似ているデータは同じクラスタに，似ていないデータは別なクラスタにデータをグループ化する分析である．クラスタは，そのクラス内の他のデータとは似ているが，違うクラスタ内のデータとは似ていないようなデータの集合である．クラスタ分析では，通常，データを多次元空間内の点とみなし，距離を定義し，距離の近いものを似ているとする．本稿では，K-Means法を利用する．K-Means法はPartitioning Methodと呼ばれるクラスタリングの一種である．データを与えられたk個のクラスタに分類し，クラスタの中心値をそのクラスタを代表する値とする．クラスタの中心値との距離を計算することで，データがどのクラスタに属するかを判断する．この際，最も近いクラスタにデータを配分する．全てのデータに対してクラスタにデータを配分し終わったあと，クラスタの中心値を更新する．クラスタの中心値はすべて点と平均値である．以上の操作を中心値との距離の合計が最小になるまで繰り返し分類を行う．計算式を以下に示す．

$$\sum_{i=i}^{K}\sum_{x\in C_i}||x-c||^2 \quad (1)$$

本稿では，5つのクラスタに分類を行いその後，テキスト解析により各クラスの問題文にどのような言葉がどれだけ含まれているかを見ることによりテキスト全体の傾向を分析する．全体の傾向を把握するため，テキストマイニングの頻度分析を行う．基礎無機化学の問題はすべて英語で記述しており，英語の問題文を分析するにあたり文章を統一する．文頭は，大文字は小文字に統一，短縮形を展開，活用語を原型に統一，そして複数形を単数に統一する．統一を行った後，単語（名詞）の出現回数をカウントする．頻度の高い単語がこの分類によるキーワードであり問題のキーとなる単語と予測する．ここではどの文章にも必要と考えられるbe動詞や，followingなど問題内容とは関係のない単語は省略する．また，問題文の係り受け状態も分析する．係り受け関係とは日本語で「私は2個のリンゴを食べる」という文章で係り受けを抽出したときと同様，「私→食べる」「リンゴ→食べる」「2個→リンゴ」という係り受け関係とする．先ほどの頻出単語から最も多い3つの単語に注目をし，他のどのような単語と同時に出題されているか，どのような係り受け表現の中で用いられているか分析を行う．テキストに出現する単語のすべての組のうち頻出単語に合致するものwの集合を$W = \{w_i\}$とする．単語が出現した問題番号を$n(w)$とし，単語w_iと単語w_jが同時に出現した行の数を$n(w_i, w_j)$とするとき，次の5つの手順に従って単語の共起情報を抽出する．まず，問題に出現する単語の全ての組について信頼度P_{ij}，共起ルール数C_{ij}を求める．

$$P_{ij} = \frac{n(w_i, w_j)}{n(w_i)} \tag{2}$$

$$C_{ij} = n(w_i, w_j) \tag{3}$$

ここで，信頼度とは，ある行に単語w_iが出現したとき，単語w_jが同一行中に出現していた確率を表している．また，w_iは前提，w_jは結論であり，共起ルール数は，単語w_iとw_jが同時に出現した行の数である．(w_i, w_j)のうち，信頼度P_{ij}が下限値$\underline{P}$以上であり，かつ，共起ルール数C_{ij}が下限値$\underline{C}$以上であるという条件を満たしながら，さらに，w_i, w_jの少なくとも一方が注目語であるか，又はw_i, w_jのいずれかが注目語である(w_i, w_j)の組に含まれるものを，特別に共起ルールと呼ぶ．

w_i, w_jの少なくとも一方が注目語であるという条件を満たす(w_i, w_j)の組の中で，前提と結論の両方が注目語の集合W_Nに含まれるものをT_u，前提のみが注目語であるものをT_{tf}，結論w_jのみが注目語であるものをT_{ft}とすると，次式のTで表される．

$$T = T_{tt} \cup T_{tf} \cup T_{ft} \tag{4}$$

w_i, w_jのいずれかが注目語である(w_i, w_j)の組に含まれるという条件を満たす(w_i, w_j)の組の中で，T_{tf}，T_{ft}に関して，注目語でない方の単語をそれぞれW_{tf}, W_{ft}とする．

$$W_{tf} = \{w_j | (w_i, w_j) \in T_{tf}\} \tag{5}$$

$$W_{ft} = \{w_i | (w_i, w_j) \in T_{ft}\} \tag{6}$$

注目語を含む(w_i, w_j)の組に出現する注目語以外の語W'は，$W' = W'_{tf} \cup W'_{ft}$となり，条件を満たす$(\mathrm{w_i}, \mathrm{w_j})$は$T'$で表すことができる．

$$T' = \{(w_i, w_j) | P_{ij} \geq \underline{P}, C_{ij} \geq \underline{C}, w_i \in W', w_j \in W'\} \tag{7}$$

$T \cup T'$を，その信頼度の降順に共起単語対として出力する．結果には，信頼度と共起ルール数に加え，テキスト中の全行のうち単語w_iとw_jが同時に出現した行の割合であるサポートS_{ij}の値も出力する．全問題数をNとして，S_{ij}は次式で示される．

$$S_{ij} = \frac{n(w_i, w_j)}{N} \tag{8}$$

協調フィルタリングによる学習者の成績分類

Moodleに登録されている問題を分類する．特定の学習者と相関の高い理解度の学習者を見付けることで，今後の推薦問題を予測する．ここでは，複数の学習者の学習履歴データから，ある特定の学習者の理解度を推定することを考える．本稿では，協調フィルタリングの代表的手法である相関係数法を用いた[9]．相関係数法は相関性に基づいたものであるあり，学習者の理解度を他人の理解度の線形和で予測し，その係数として学習者間の相関係数を採用したものである．具体的には，学習者i番目の人がxという問題に対する理解の度合いをM_{ix}とするとき，相関係数は次式で与えられる．

$$M_{ix} = M_i + \sum_j C_{ji}(M_{jx} - M_j) / \sum_j |C_{ji}| \tag{9}$$

ここで，全ての和は欠損値以外で取られるものとし，M_iはM_{ix}のxに関する平均，C_{ij}はi行j行の相関係数を表している．なお，C_{ij}は次式で与えられる．

$$C_{ij} = \sum_x (M_{jx} - M_i)(M_{jx} - M_j) \times 1/(\sum_x (M_{jx} - M_i) \sum_X (M_{jx} - M_j)^2)^{\frac{1}{2}} \tag{10}$$

類似度が0.8以上であるとき，類似精度が特に高いものとして理解できる．同時に，予測精度も高いものと理解できるため，本稿では，類似度が0.8以上の学習者のデータを用いて相乗平均にて予測を行った．

Step3：推薦問題の提示：

最終的に，得られた推薦問題を各学習者に提示する．

4.運用実験

達成度を考慮した傾向別課題作成法の有効性を検証する．Moodleにより構築されたe-learning環境上で，時間外課題7回分で使われた384の問題についてK-Means法により分類を行った[表1]．表1に示したとおり，ファシリティに関して分類3は73.89という最も高い値を得たことから，学習者の多くは分類3の問題群に属する問題を理解していると考えられる．同様に，識別指数も高い値であったことから，学習者にとって適切な問題であったのではないかと考えられる．分類3ほどではないものの，半分以上の学習者は分類5に属する問題に正解していることから，学習内容をおおよそ理解していると考えられる．一方，分類１に関しては，半分以下の学習者しか正解できていなかったことから，分類1に属する問題は学習者にとってやや難しい問題であったと考えられる．同様に考えると，分類2に属する問題は学習者にとってやや難しい問題であることがわかる．分類4に属する2つの問題は，ファシリティ，識別子数共に0であった．判別係数もマイナスであったことを踏まえた場合，不適当な問題であったか，あるいは誰も問題を解いていなかったかの2通りの解釈ができる．このことから，分類4を分析対象から除外した．

本研究は，学習者に対して，分野毎の得手不得手を問題配信に利用することを考えている．そこで，各問題はどのような分野に属するかを明らかにするため，テキストマイニング法により単語頻度の分析を行い，各分類のキーとなる単語を抽出した．得られた結果を表2-5に示す．頻度の高い単語は，時間外課題の中で重要なキーワードであると考えられる．また，他の問題には出てきていない頻度の高い単語は，今回の時間外課題で特別に出てきた単語であることから，キーとなる可能性がある．なお，この単語の意味を理解していないならば，この回の時間外課題を解くことが難しくなることから，特に重要な単語であると考えられる．

抽出されたキーワード同士の関係性を確認するため，上位3つの頻出単語は他のどのような単語と同時に出現しているのかを調査した．得られた係り受け関係を図1－4に示す．これらを視覚的に表示したことにより，特に分類2の問題の複雑性が確認された．複雑な関係性から，多くの単語や知識が必要な問題であると理解できる．一方，分類3に関しては多くの関わりが確認されなかったことから，分類3に属する問題は容易であると評価されていたことがわかる．以上の結果を踏まえると，質問中に多くの語彙があるということが問題の難易度を定義する最も大きな要因であると考えられる．本稿で分析対象とした問題については，多くの語彙が含まれることと豊富な知識を要することが直接的に関係しており，こうした問題全体の傾向が各問題の難易度を決定付けていたのではないかと考えられる．

表１．K-Means法による分類

クラスID	size	ファシリティ	識別指数	判別係数	residual
1	87	44.37	0.55	0.61	294.84
2	86	28.29	0.33	0.51	451.72
3	104	73.89	0.87	0.62	494.66
4	2	0.01	0.00	-998.69	0.61
5	105	57.58	0.73	0.62	384.11

表2．分類１における頻出単語

1	Electron	16
2	Atom	12
3	Element	12
4	Bond	11
5	Compound	9
6	Octet	9
7	Structure	9
8	Energy	7
9	Lewis	7
10	Pair	7

表3．分類２における頻出単語

1	Atom	18
2	Bond	18
3	Electron	14
4	Compound	12
5	Covalent	12
6	Energy	12
7	Ion	11
8	Lewis	9
9	Structure	9
10	Pair	8

表4．分類３における特徴

1	Element	31
2	Electron	19
3	Atom	16
4	Reference	16
5	Configuration	15
6	Color	11
7	Group	11
8	Style	8
9	Consider	8
10	Div	8

表5.分類５における特徴

1	Element	21
2	Bond	16
3	Compound	14
4	Electron	14
5	Energy	12
6	Covalent	9
7	Molecule	9
8	Atom	8
9	Ion	7
10	Color	6

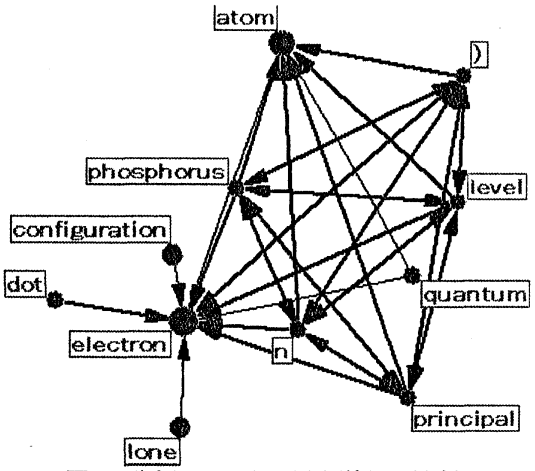

図1. 分類1における頻出単語の関係

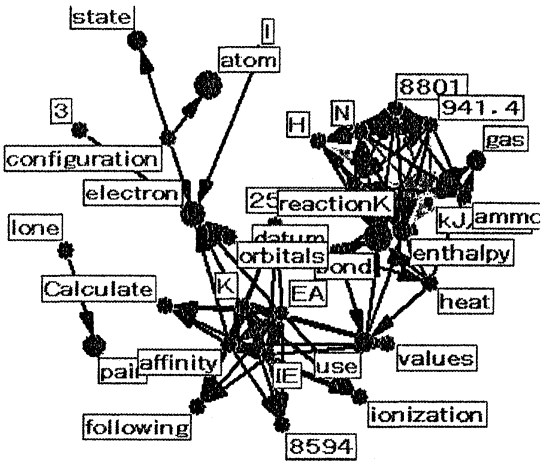

図2. 分類2における頻出単語の関係

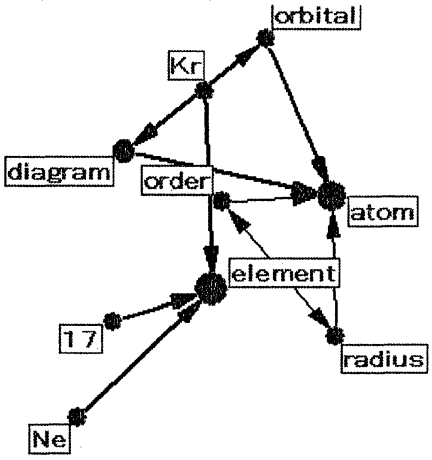

図3. 分類3における頻出単語の関係

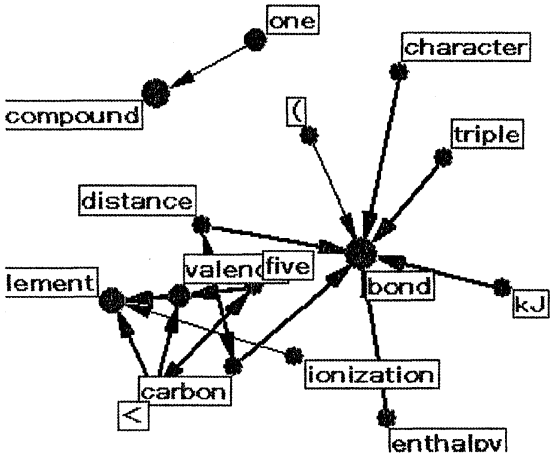

図4. 分類4における頻出単語の関係

次に，各学習者の成績データを用いて，協調フィルタリングにより分析を行った．分析結果を表6に示す．本稿では，95人の学習者のうち，1人の学習者に問題を推薦していく過程を明らかにした．まず，提案法に従って，各学習者との相関関係を分析した．相関係数が0.8以上の場合に予測精度が高いことを踏まえて，相関係数が0.8以上の学習者を対象とした．対象者は26名であり，対象の学習者は，1回目の時間外課題があまり良い結果ではないものの，その他の時間外課題では良い結果であるという特徴を持っていた．1回目の時間外課題で分類2の問題が多く出題されていたことから，対象の学習者と相関の強い学習者のグループは，問題分類の分類2のような問題を苦手とする学習者であることがわかる．また，相関の強い学習者の過去の成績データをもとに相乗平均を出すことによって，表6に示した予測点数を得た．表6から，学習者は8番目の問題で良い点数を取れていないと予想される．よって，対象の学習者26人に対しては，難易度に直結すると考えられるファシリティが低い問題で，かつ分類2に属する問題の復習が必要と考えられる．8番目の問題に関して，実際には，対象の学習者は予測通り5点という結果となった．11番目の問題の場合，予測点数3.81点に反して，対象の学習者はとても良い得点である9点を得ている．問題11のヒストグラムを図5に示す．11番目の問題は，標準偏差は低く，また，平均点3.8点と難易度の高い問題とされていた．したがって，標準偏差の低い問題では，予測はやや難しいことが分かる．

表6. 時間外課題の予測点数と実際の点数

時間外課題番号	6	7	8	9	10	11
実際の点数	10	9	5	10	10	9
標準偏差	4.05	2.15	1.99	4.43	4.07	2.96
予測（0.8以上）	7.02	8.36	5.21	7.86	8.11	3.81

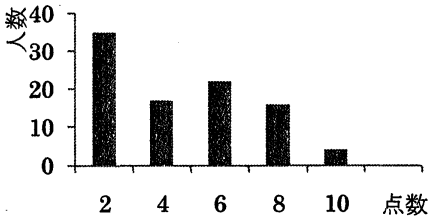

図5. 問題11のヒストグラム

表7. eメールプッシュ型システムの利用率

被験者	A	B	C	D	E	F	G
経験値	342	501	207	1879	551	68	373
問題発行数	1	1	3	5	1	1	1
メール送信総数	160	137	122	122	123	122	124
メール返信数	133	122	32	62	49	30	46
返信率	83%	89%	26%	50%	39%	24%	37%

図6. 返却率の推移

今回，開発システム単独での利用率の推移に関する検証を行っている．具体的には，開発者と数名の協力者による運用実験により，システムの動作，約3ヶ月間配信を行い返信数により利用可能であるかを確認した結果，表7，図5に示したとおり，平均50%の返信率を得た．

5.終わりに

本稿では，学習者個々の達成度を考慮した傾向別の課題作成法を提案し，講義，e-Learning，プッシュ型システムを連携させる運用法について述べた．個々の到達度を考慮した問題配信を行うため，e-Learning利用者に対して問題分類及び相関関係の強い学習者の分類を行い，学習者の苦手とする問題を抽出することが可能となった．具体的には，学習者の学習達成度の相関が高い人は理解度が似ていることが示され，協調フィルタリングに基づく情報推薦処理を問題配信に応用することで，各学習者に対して，分野毎の習熟度や傾向に合わせた問題の提供を示した．また，問題配信アルゴリズムの構築に向け，テキスト解析に基づいた内容ベースフィルタリングを取り上げ，問題のクラスタリング，難易度の分類などを行うことが可能となった．今後，e-Learningの分析結果に基づくeメールプッシュ型システムによる問題配信を連携し自動で処理を行うことにより，指導者を補助しながら，学習者の学習効果が上げることが期待される．同時に，更なる実験を繰り返し，学習効果や理解度がどの程度向上するかを検証する予定である．

参考文献

[1] ジョシュ バーシン，ブレンディッドラーニングの戦略—e-Learningを活用した人材育成，東京電機大学出版局，2006.

[2] 最新 ITソリューションマップ(第9回), eラーニングソリューション, Business communication, Vol.43(1), 503, pp,137-139, 2006.

[3] 清水康敬, 大学におけるeラーニングの現状と推進の視点, 電気学会誌, Vol.129, No.9, pp.596-599, 2009.

[4] 長谷川聡，小橋一秀，長谷川旭, Webベース教育システムについて-名古屋文理大学における学習・教育支援の実践と提案-，名古屋文理大学紀要，第4号，pp.65-70，2004.

[5] 大西英雄, 網島ひづる, 金井秀作 他, 大学院教育におけるWebカメラを用いた遠隔地e-Learningシステム構築の評価, 県立広島大学保健福祉学部誌, Vol.9(1), pp.111-120, 2009.

[6] 蜂屋絵美里, Skypeが描く世界語会話教室 - 非透明人間と不透明人間の協働実現を目指して, 野村総合研究所学習者小論文コンテスト2009.

[7] 渡部清, 和田剛樹, 吉本直樹 他, 個人携帯端末としてのニンテンドーDSの可能性, 日本科学教育学会研究会研究報告, Vol.21(6), pp.51-54, 2007.

[8] 九里徳泰, 携帯電話によるEラーニングを活用した大学多人数講義での運用実験，メディア教育研究, Vol.1, No.2, pp.145-153, 2005.

[9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstom and J.Riedl, GroupLens: An Open Architechture for Collaborative Filtering of Netnews. Pro.c. of ACM Conf. on Computer Supported Cooperativ Work (CSCW94) pp.175-186, 1994.