

SQL 実習支援システムのための教授戦略生成法

松本 拓也^{†1} 岡田 信一郎^{†2}
石井 智佳^{†2} 柿沼 有希^{†2}

著者らは、SQL 実習支援システムの開発を行っている。出題に必要な教授戦略を従来は人手で作成していた。しかし、人手による教授戦略は作成者の主観に偏る可能性がある。本研究では、書籍などから収集した SQL 文をもとに、より客観的な教授戦略を自動生成する手法について提案する。さらに、本手法で生成した教授戦略を実習支援システムに適用し、実用評価を行った結果を合わせて報告する。

A Generation Method of Teaching Strategies for SQL Exercise Support System

TAKUYA MATSUMOTO,^{†1} SHIN-ICHIROU OKADA,^{†2}
CHIKA ISHII^{†2} and YUKI KAKINUMA^{†2}

We have been developed SQL exercise support system. we had made teaching strategies, which is required for question-making, by hand work previously. However, human-made teaching strategies tend to be affected by subjective thought of creator. We propose a method to generate more objective teaching strategies based on SQL sentences collected from books. In addition, we applied generated teaching strategies to the system and evaluated the system by operating that. We also report evaluation results.

^{†1} 茨城大学大学院 理工学研究科

Graduate School of Science and Engineering, Ibaraki University

^{†2} 茨城大学 工学部

Faculty of Engineering, Ibaraki University

1. はじめに

データベース操作言語 SQL は様々な情報システムで利用されており、情報分野に携わる者にとってその習得は必須であると言える。SQL を学ぶ上で、特に初心者の学習においては、単純な問題を反復して解くことが有効である。しかし、数多くの問題を出題・採点することは指導者にとって大きな負担となる。そこで著者らは、基本的な SQL の問題の出題・採点を自動的に行う「SQL 実習支援システム」¹⁾の開発を行っている。

本システムでは、個々の学習者の理解状況に応じた学習項目を決定し、その項目の問題を作成し出題する。学習項目の決定のためには、学習者の理解状況を記録する学習者モデルと、出題順を規定する教授戦略が必要となる。従来、教授戦略は人手で作成していたが、人手による教授戦略は作成者の主観に偏る可能性がある。

そこで著者らはまず、教授戦略の基礎となる学習項目を自動生成する手法を提案した。²⁾これは収集した SQL 文から関連の強い予約語を抽出することで、客観的で実利用に即した学習項目を作成するものである。本研究ではこれを改良・発展させた、客観的な教授戦略を自動生成する手法について提案する。さらに、実用評価として、生成した教授戦略を適用したシステムを本学の講義で実際に運用しアンケート調査を行った。この評価結果についても合わせて報告する。

2. 関連研究

学習者適応型の出題を行うシステムとしては 3) がある。これは、学習の順序などを規定した講義シナリオをもとに各分野の解説や問題の出題を行うものである。講義シナリオでは学習者の理解状況を参照し、未習得の分野があればその解説や問題の出題を繰り返す。しかし、あらかじめ用意された問題を出題するため、場合によっては同一の問題が繰り返し出題される可能性がある。

4) では問題文テンプレートと語彙情報をもとに選択問題を生成する。出題した問題と学習者の理解度をもとに次に出题すべき問題の難易度を計算し、選択肢の組み合わせにより問題の難易度を変更している。学習の順序を規定するものは無いため、初心者が順序立てて学習することには向かない。

5) では、本研究と同様に SQL の実習システムを開発している。これは、学習者に問題のジャンルを選択させ、そのジャンルの問題を自動生成して出題するものである。そのジャンルを理解した学習者向けに、より難易度の高い問題を出題する機能も持つ。あらかじめ用意

された SQL のテーブルデータから問題を自動生成する点では類似しているが、本研究では学習する項目とその順序を教授戦略をもとに決定する点で異なる。

3. SQL 実習支援システム

3.1 システムの概要

本システムの概要について説明する。本システムは SQL を初めて学習する者を対象とし、個々の学習者の理解状況に見合った問題を自動生成し、出題を繰り返すシステムである。また、採点も自動で行うようになっている。本システムを用いることで、指導者の問題作成・採点の負担を軽減し、反復学習により学習者の理解を深めることができる。システムは Web アプリケーションと Java アプレットにより構成され、Web ブラウザ上で SQL を実行し結果を確認できるようになっている。

図 1 に本システムの動作画面を示す。画面右側が実習画面となっている。その上部に問題が表示され、下部に学習者が解答となる SQL 文を入力する。入力した SQL 文を実行して結果を確認することもできるようになっている。画面左側では学習の進捗状況が確認できるようになっている。

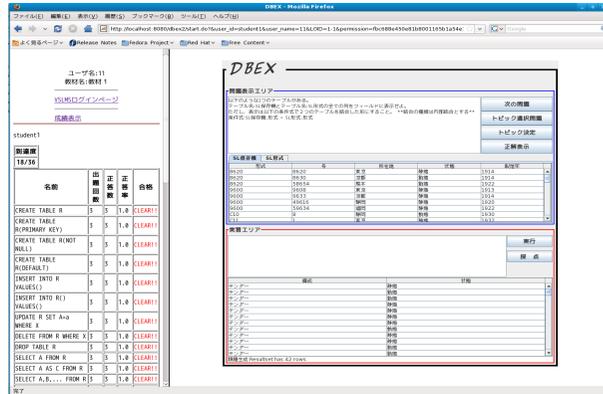


図 1 システムの動作画面

本システムの構成を図 2 に示す。また、本システムの処理の流れを以下に示す。

- (1) 学習項目決定機能が学習履歴データベースから学習者の理解状況を取得し、その状況

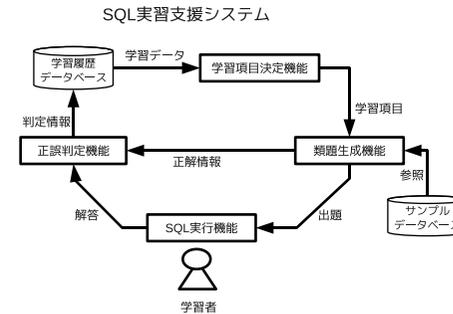


図 2 システムの構成

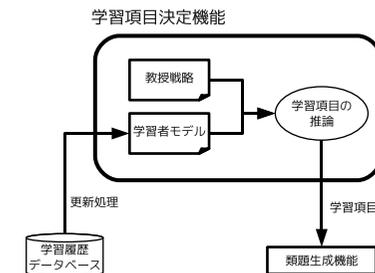


図 3 学習項目決定機能

に見合った学習項目を決定し類題生成機能へ送る。

- (2) 類題生成機能が、受け取った学習項目とサンプルデータベースをもとに問題を作成し、学習者へ出題する。
- (3) 学習者が出題された問題に解答する。SQL 実行機能で解答の SQL 文を実行し結果を確認することができる。
- (4) 正誤判定機能が、学習者の解答と類題生成機能から受け取った正解情報を比較し、正誤判定を行う。判定結果は学習履歴データベースへ送る。自動で判定できない場合は、指導者に判定を依頼する。
- (5) (1) から (4) の処理を繰り返し、出題する。

3.2 学習項目決定機能

先に述べたように、この機能では学習状況に見合った学習項目を決定する。

学習においては未習得の部分を繰り返し学習することが効果的である。しかし、初めて SQL を学習する者にとっては、学習の順序も重要である。本機能では、教授戦略と学習者モデルという 2 つの情報から学習項目を決定し、「未習得分野の反復」「適切な指導順」を両立した出題を行う。

教授戦略は、学習項目の出題順を規定するルールを列挙したものである。学習者モデルは、学習者の理解状況を学習項目ごとに記録したものである。学習項目決定機能の処理の流れを図 3 に示す。

4. 教授戦略

4.1 教授戦略の概要

本研究で扱う教授戦略の概要について説明する。

教授戦略を図示したものが、図4の木構造である。それぞれの枠が学習項目を表しており、学習項目は「SELECT A FROM R」のようにSQL文の雛形とする。また、扱う文法事項は本学工学部情報工学科の講義「データベース論」で学習する範囲かつ、SQL実行機能で利用しているデータベースマネジメントシステム「HSQLDB」⁶⁾でサポートする範囲を対象とする。

上位の学習項目を理解してからでなければ下位の学習項目は出題できないものとして、出題順を規定する。ある学習項目を理解していた場合、その下位の学習項目が次の出題候補となる。例えば図4において太枠の項目を理解している場合、その次には実線の枠の項目が出題候補となる。点線の枠の項目は出題できない。この例のように出題候補が複数存在する場合、その中からランダムに選択して出題する。

4.2 人手による教授戦略の問題点

従来は、学習項目・教授戦略の作成を作成者が手作業で行っていたが、人手による場合の問題点として以下のことが考えられる。

まず、指導順が適切であるかという問題がある。教授戦略は作成者が適切な学習項目や指

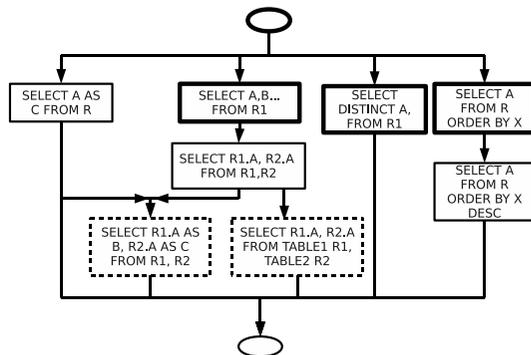


図4 教授戦略を図示した木構造

導順を考えて作成されているが、それはあくまでも作成者の考えであり、主観に偏ったものとなる可能性がある。指導順は基礎から応用という流れを踏まえつつ、あまり固定せずに可能な限りランダムに出題することが望ましい。また、学習項目と実際に利用されているSQL文との対応についての問題もある。人手で作成した学習項目はほとんどが単純なものであるが、実際に利用されるSQL文は複数の予約語を組み合わせて利用される場合が多い。単純な学習項目だけでは、実際の利用に即したものになっていないと考えられる。

5. 教授戦略自動生成

5.1 学習項目ツリーの生成

人手の場合の問題を解決するために、自動で教授戦略を生成する手法について検討した。まず、学習項目を作成し、教授戦略の基本となる「学習項目ツリー」の生成を行う。方針としては、資格試験の問題やSQLの入門書などの書籍⁷⁾⁻¹³⁾からSQL文を収集し、収集したデータから予約語間の相関ルールを抽出する。抽出されたルールは関連の強い予約語の組を表すこととなる。ルールの「条件部+結論部」という予約語の組を学習項目とすることで、実際の利用に即した予約語の組み合わせを抽出できると考えた。問題生成の際には、学習項目に存在する全ての予約語を利用したSQL文を解答させる問題を生成する。

学習項目ツリー生成のアルゴリズムを以下に示す。

- (1) 収集したSQL文を予約語の並びとして図5のように表す。このとき、予約語を上位概念でまとめることができる場合には上位概念を用いる。例えば「=」や「<」、「>」といった予約語は、上位概念「比較演算子」として扱う。これは、上位概念でまとめることにより意味のある相関ルールを抽出しやすくするためである。
- (2) (1)のデータから、予約語間の相関ルールを抽出する。相関ルール抽出にはアプリアルゴリズムを用いる。最初はあらかじめ定めたコアとなる予約語を条件部として与えて、相関ルールを抽出する。コアとなる予約語は、「SELECT FROM」「CREATE

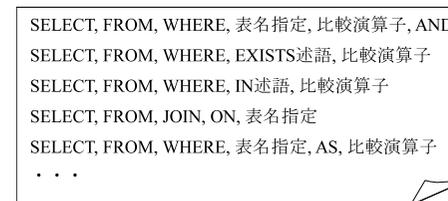


図5 収集したSQL文の予約語データ

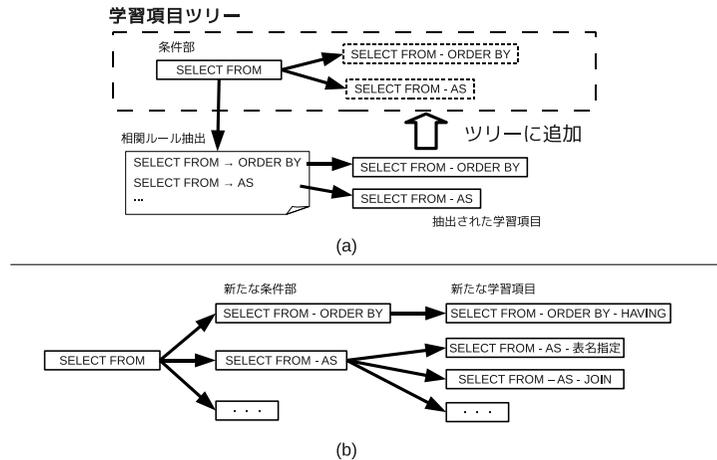


図 6 学習項目ツリー成長の処理の流れ

TABLE」「DROP TABLE」「DELETE FROM」「UPDATE SET」である。以降は「SELECT FROM」から始めた場合について、図 6 を用いて説明する。

- (3) 抽出された相関ルールから学習項目を作成する。図 6 (a) では、「SELECT FROM → ORDER BY」「SELECT FROM → AS」といったルールを抽出し、「SELECT FROM - ORDER BY」「SELECT FROM - AS」といった学習項目を生成している。抽出された学習項目「SELECT FROM - ORDER BY」などを、条件部であった「SELECT FROM」の子要素として追加する。こうして得られた木構造が学習項目ツリーとなる。
- (4) (3) で子要素として追加した「SELECT FROM - ORDER BY」などを新たな条件部として相関ルール抽出・学習項目生成を行う。抽出された「SELECT FROM - ORDER BY - HAVING」などの新たな学習項目を「SELECT FROM - ORDER BY」などそれぞれの要素の子要素に追加していき、図 6 (b) のようにツリーを成長させていく。
- (5) アプリリアルゴリズムの最小サポート・最小確信度を満たすルールが抽出されつづける間、学習項目生成とツリーの成長を繰り返す。

5.2 不適切な学習項目の削除

5.1 で生成した学習項目ツリーには、不適切な学習項目が含まれる可能性がある。そうし

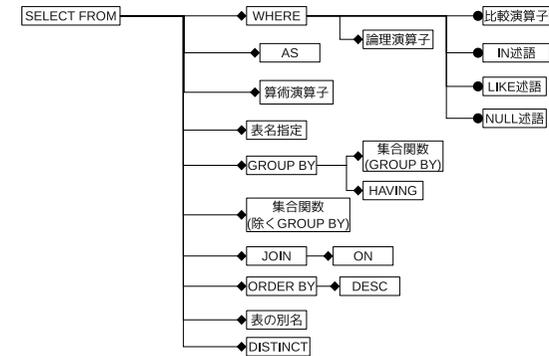


図 7 依存関係ツリー

た項目を削除して、ツリーを完成させる。

5.2.1 依存関係上不適切な項目の削除

予約語には、ある予約語との共起が必須であるといった予約語間の依存関係が存在する。学習項目作成の際には、この依存関係を考慮する必要がある。そこで、図 7 のように予約語の依存関係を表す「依存関係ツリー」を作成した。このツリー構造は、親要素の予約語が存在すれば子要素の予約語が存在できるという関係を表す。例えば、「論理演算子」が存在するためには「WHERE」が、「DESC」が存在するためには「ORDER BY」が必要である。相関ルールを抽出した際にこの依存関係と照合し、不適切なものは学習項目ツリーに加えない。

5.2.2 重複する項目の統合

予約語の並び順が異なるだけで、同一の学習項目となるものが複数出現する場合がある。例えば、ツリーを成長させていくと図 8 左部のように「SELECT FROM - AS - 表名指定」と「SELECT FROM - 表名指定 - AS」という 2 つの項目が抽出される。しかし、この 2 つの項目に存在する予約語は並びが異なるだけで同じ組み合わせであるため、この 2 つは同一項目である。このように重複する項目が存在する場合には、図 8 右部のように一方を削除しリンクをもう一方へ繋ぎ直すことで、項目を一つに統合する。

5.2.3 問題生成に利用できない項目の削除

学習項目の中には、学習項目ツリーの成長には必要だが実際の問題作成には利用できない項目が存在する場合がある。図 9 において、「SELECT FROM - WHERE」は WHERE 句の条件記述に必要な比較演算子などが無いため、問題作成には利用できない。図 7 にお

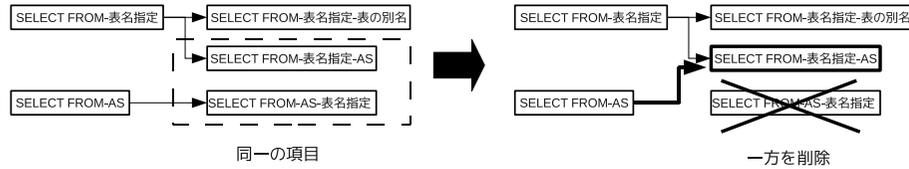


図 8 重複する項目の統合

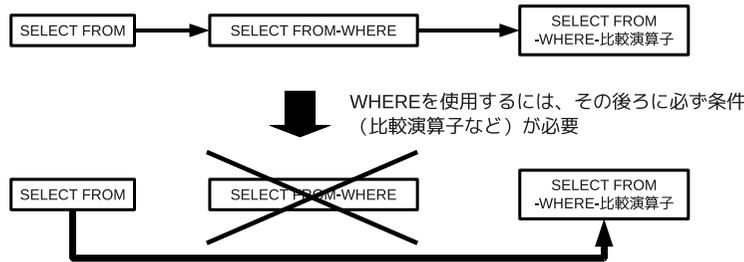


図 9 問題作成に利用できない項目の削除

丸矢印は「親要素と子要素の共起が必須である」ということを表しており、「WHERE」は「比較演算子」などとの共起が必須となっている。しかし、ツリーの成長過程で、依存関係上不適切であるとして「SELECT FROM - WHERE」を削除してしまうと、「SELECT FROM - WHERE - 比較演算子」といった本来抽出されるはずの項目を抽出できなくなる。そのため、このような項目についてはツリーの成長過程では残しておき、最後に図9のように問題作成に利用できない項目として削除する。

5.3 教授戦略の生成

得られた学習項目ツリーから実際に教授戦略を作成する。教授戦略は XML 形式のルールとして記述する。¹⁴⁾ 図 10 に教授戦略の例を示す。この例は、「SELECT FROM GROUP BY」についてのルールである。上部の枠の記述が「『SELECT FROM GROUP BY』を理解していない」という条件であり、下部の枠の記述が「『SELECT FROM GROUP BY』を学習する前に理解しておかなければならない項目『SELECT FROM』を理解している」という条件である。

それぞれの項目を理解したかどうかは、出題回数と正答率によって判断する。教授戦略生成の際に最低出題回数と合格正答率を指定し、「出題回数が最低出題回数以上」かつ「正答

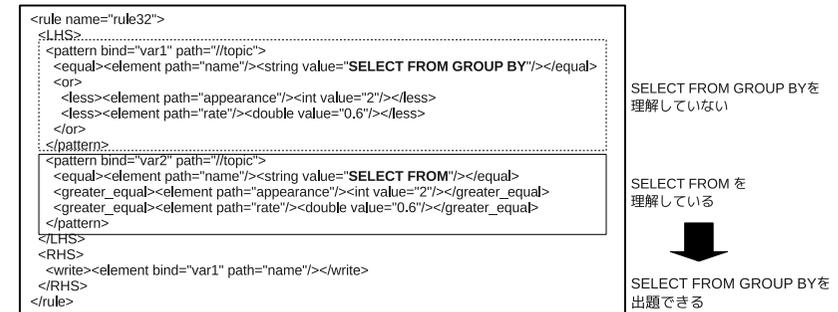


図 10 教授戦略ルール

率が合格正答率以上」であればその項目を理解したとみなす。現在は最低出題回数を 2 回、合格正答率を 0.6 としている。

6. 評価

6.1 生成された教授戦略

7)–13) から収集した 397 例の SQL 文を用いて、教授戦略の自動生成を行った。関連ルール抽出の際のアプリオリの最小サポートは、「SELECT FROM」から始める場合は 2.0%、それ以外では 0.5%とした。また、最小確信度は 5.0%とした。すべてのコアとなる予約語について教授戦略の生成を行った結果、生成された学習項目は 53 項目となった。形式が異なるために単純比較はできないが、これらは人手による場合の約 90%をカバーしている。また、「SELECT FROM-WHERE-比較演算子-表名指定-AS」のように、人手による場合には無かった、多くの予約語を組み合わせた新たな学習項目も生成された。本稿では、コアとなる予約語「SELECT FROM」を指定して生成された学習項目ツリーを図 11 に示す。

6.2 システムでの実用評価

平成 22 年度の本学工学部情報工学科の「データベース論」の講義において、生成した教授戦略を適用した本システムを運用し、履歴データを収集した。システムを利用した学習者は 78 名であった。また、システムを利用した学習者に対してアンケートを行い、56 名の学習者から回答を得た。履歴データとアンケート結果から教授戦略の評価を行った。

6.2.1 学習項目数の妥当性

学習項目数の妥当性について検証を行った。今回の運用では、全 53 項目のうち 35 項目を理解することをノルマとした。これは、53 項目全てのクリアを課すことは学習者の負担

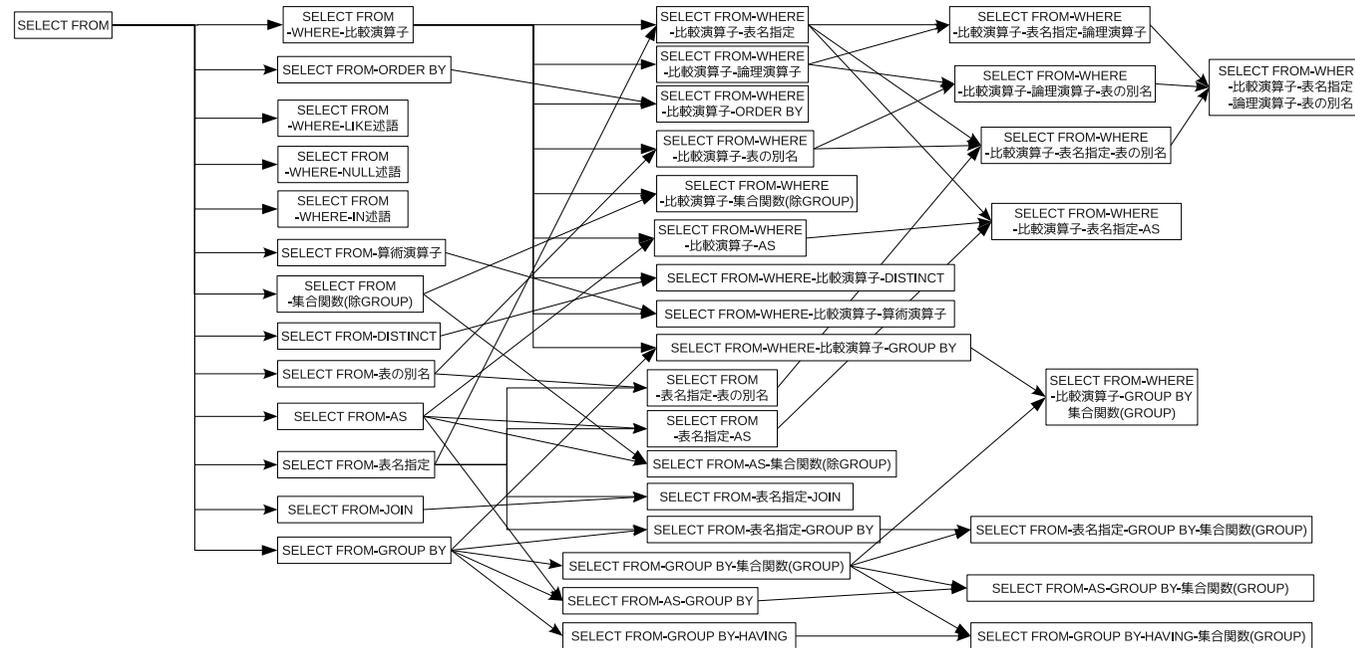


図 11 生成された学習項目ツリー

が大きいと考えたためであるが、74 名がノルマである 35 項目以上をクリアしており、全体の半数以上となる 42 名が全ての項目をクリアしていた。また、図 12 に示すように、「問題数は適切でしたか?」という質問に対して、「ちょうど良かった」と回答した学習者が 7 割以上であった。こうしたことから、学習項目数は妥当であったと言える。

6.2.2 学習項目の正答率への影響

学習項目の正答率への影響について検証を行った。履歴データから学習項目ごとの正答率を調べた。正答率が 0.8 未満の項目を表 1 に示す。正答率が低い項目には比較的単純な項目も含まれていた。一方で、「SELECT FROM-WHERE-比較演算子-表名指定-論理演算子-表の別名」のように予約語の数が多い項目は一見すると複雑そうに見えるが、学習の後半で出題される項目であるため正答率はそれほど低くはなかった。アンケート結果では、図 13 「不正解や指導者判定となった解答があった場合、特に大きな原因は何であったと思

いますか?」という質問に対して、「自分のケアレスミス (39 名)」「自動採点が厳密過ぎる (12 名)」などとなり、「問題自体が複雑」と回答した学習者はいなかった。図 14 「このシステムの悪いところはどこですか?」という質問では、「問題文のわかりやすさ」と回答した学習者が多かった。こうしたことから、正答率が低くなっている原因は学習項目に存在する予約語の数による影響よりも、ケアレスミスや採点機能、問題文がわかりにくいといったことによる影響が大きいと考えられる。

6.2.3 学習項目・出題順の妥当性

学習項目とその出題順の妥当性について検証を行った。図 15 「このシステムで学習を進める中で、同じような問題が多く出題された (同じようなトピックが多かった) と感じましたか?」という質問に対して、「同じような問題が多かったが、反復学習には良かった」という意見が多数であった。このことから、類似した学習項目が存在し類似問題も多かった

が、学習効果を高めるといって学習者からは肯定的に受け止められていると言える。また、図16「序盤に難しい問題が出題される、終盤に簡単な問題が出題されるなど、問題の出題順に不自然さがありましたか?」という質問に対して、ほとんどの学習者が「いいえ」と回答していた。このことから、ほとんどの学習者が出題順に不自然さを感じておらず、基本から応用という順序での出題を行うことができたと言える。

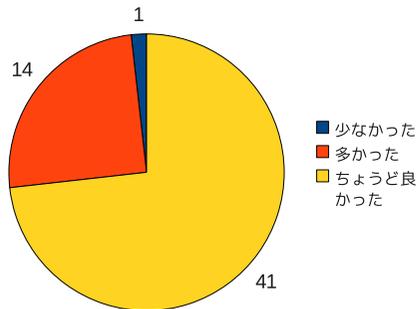


図12 「問題数は適切でしたか?」という質問への回答

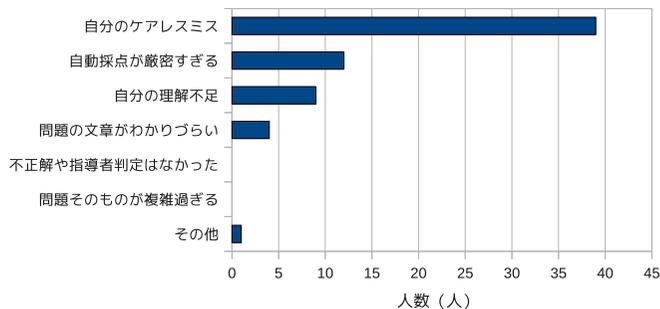


図13 「不正解や指導者判定の原因は何でしたか?」という質問への回答 (複数回答)

表1 正答率が低い項目 (正答率 0.8 未満)

学習項目	正答率
SELECT FROM GROUP BY HAVING	0.65
SELECT FROM GROUP BY HAVING 集合関数 (GROUP)	0.67
INSERT INTO INSERT の列名指定	0.77
SELECT FROM WHERE NULL 述語	0.78
SELECT FROM WHERE 比較演算子 論理演算子	0.78

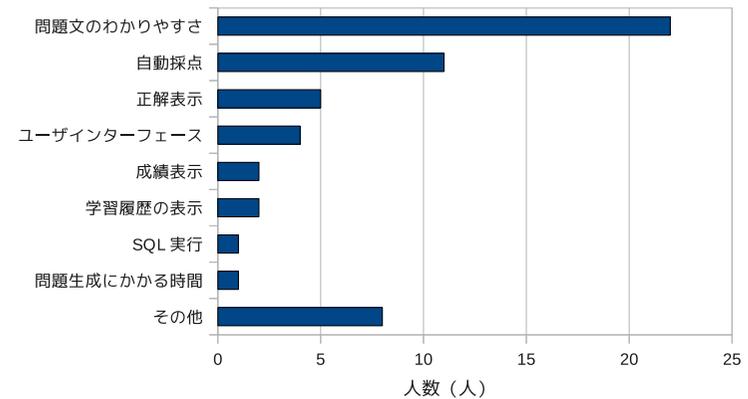


図14 「システムの悪いところは?」という質問への回答 (複数回答)

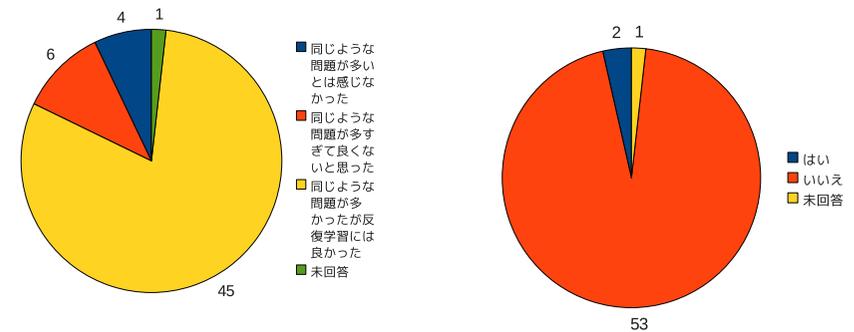


図15 「同じような問題 (トピック) が多く出題されま 図16 「出題順に不自然な点はありませんでしたか?」という質問への回答

6.2.4 問題点

良い評価がある一方で、「後半は1度経験したような問題だらけなので全項目やる気にならない」「あまりに長い命令文が多く、後半疲れてミスが多くなった」といった意見もあった。予約語が多い学習項目についての履歴を見ると、ケアレスミスと言える解答も散見された。後半になると、学習項目に含まれる予約語の数の増加に伴い命令文も長くなり、ケアレスミスの一因となっていることも考えられる。特に、「表名指定」というキーワードは、多くの学習項目に含まれていた。

また、全項目をクリアしなかった学習者の中に、一部の基本的な項目を解かずにノルマをクリアして学習を終えてしまった学習者がごくわずかであるが存在した。これは、ランダムな学習項目選択や学習の進捗具合によって、基本的な項目が後半に残ってしまう場合があったためであると考えられる。

6.3 実用評価のまとめ

評価結果から、学習項目は適切であり、項目数や出題順にも問題は無いことがわかった。したがって、生成された教授戦略はおおむね妥当なものであると言える。

一方で、いくつかの問題点も見つかった。これらの問題点の改善には、学習項目の統合や絞り込みを行い項目の数を減らすことが有効であると考えられる。そうすることで、学習意欲の低下やケアレスミスの増加を抑えることができると考える。項目数を減らすかわりに全項目の学習を課すことで、未習得の項目を残さないことも可能となる。

7. まとめ

SQL 実習支援システムにおける学習者適応型の出題のための教授戦略を、自動生成する手法について検討した。従来の教授戦略は人手で作成したため、作成者の主観に偏る可能性があった。本研究では書籍などから SQL 文を収集し、そこから抽出した関連ルールをもとに学習項目を作成することで、実際に利用されている SQL 文をもとにした客観的な教授戦略の生成を行った。この教授戦略は人手の場合の大部分の項目をカバーしており、新たな学習項目も得ることができた。

さらに、生成された教授戦略を適用したシステムを運用して実用評価を行った。運用後に行ったアンケートでは、学習項目やその出題順などの教授戦略に関連する質問に対してはおおむね高評価が得られた。

一方で、「似たような問題が多すぎて全項目をやる気にならない」「命令文が長くミスが増えた」という意見もあった。類似問題が多い点や解答の命令文が長くなりすぎることなどが

ら、意欲の低下やケアレスミスの増加を招いていることが考えられる。さらに、学習の進行過程の違いにより、基本項目を学習せずにノルマクリアで学習を終えてしまった学習者も存在した。

こうした問題を改善するためには、学習項目を絞り込むことで意欲の低下を防ぎ、全ての項目の学習を課すことで未習得の項目を残さないことが有効であると考えられる。今後は、学習項目の修正や重要度を付加した推論なども検討し、より効果的な学習のための教授戦略の生成に取り組んでいく予定である。

参考文献

- 1) 岡田信一郎, 青山優太, 菊池孝典, 松本拓也: ユーザ適応型課題生成機能を備えた SQL 実習支援システムの試作, 電子情報通信学会総合大会, D-15-19, 2009
- 2) 松本拓也, 岡田信一郎: SQL 実習支援システムのための学習項目生成法, 情報処理学会全国大会, 4ZK-3, 2010
- 3) 鈴木智樹, 藤原祥隆, 岡田信一郎, 吉田秀樹: ユーザ適応型 e-Learning システム KUSEL の設計, 電子情報通信学会技術研究報告, AI, 人工知能と知識処理, vol.103, No.726, pp.33-38, 2004
- 4) 津森伸一, 海尻賢二: 理解状況に適応した選択問題生成方法の検討, 教育システム情報学会誌, vol.26, No.3, pp.240-251, 2009
- 5) 内田英喜, 森本純一, 史一華, 徐海燕: SQL 演習問題自動出題演習システムの開発, 火の国情報シンポジウム, 2009
- 6) <http://hsqldb.org/>
- 7) 山平耕作: 平成 21 年度データベーススペシャリスト合格教本, 技術評論社, 2008
- 8) 高橋麻奈: 情報処理技術社試験 やさしく学ぶデータベース技術, 翔泳社, 2007
- 9) 谷尻かおり: データベース基礎の基礎, 技術評論社, 2009
- 10) 上島紳一, 上田真由美: データベース-活用のための基礎知識-, 昭晃堂, 2009
- 11) 阿部武彦, 木村春彦: 初歩のデータベース論, 共立出版, 2007
- 12) 藤本壺: これだけは知っておきたいデータベースの常識, 技術評論社, 2009
- 13) 川越恭二: 楽しく学べるデータベース, 昭晃堂, 2007
- 14) 岡田信一郎, 小林寿啓: XML による自己書き換え型プロダクションシステム実行環境の試作, 電子情報通信学会総合大会, D-8-13, 2008