

# SQL 実習支援システムのための学習項目生成法

A Generation Method of Learning Topics for SQL Exercise Support System

松本拓也<sup>1</sup>

Takuya Matsumoto

岡田信一郎<sup>2</sup>

Shin-ichirou Okada

茨城大学大学院 理工学研究科<sup>1</sup>

Graduate School of Science and Engineering, Ibaraki University

茨城大学 工学部<sup>2</sup>

Faculty of Engineering, Ibaraki University

## 1 はじめに

SQL の学習には単純な問題を反復して解くことが有效であると考えられる。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負担となる。そこで著者は、基本的な SQL の問題を自動的に出題し採点を行う「SQL 実習支援システム」[1] の開発を行っている。

このシステムの出題では、学習者の学習状況に応じた学習項目を決定し問題を作成する。この学習項目は、出題方針である教授戦略と学習記録をもとに決定される。現状では教授戦略は人手で作成しているが、これを自動的に生成したいと考えている。そこで、本研究では教授戦略の自動生成のために、学習項目を自動生成する方法について検討した。

## 2 人手による教授戦略

現状の人手による教授戦略について説明する。まず、学習項目を作成する。学習項目は、「SELECT A FROM R」のように出題する SQL 文の雛形とし、構文は基本的なものとする。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、システムに利用しているデータベースマネジメントシステム「HSQLDB」[2] でサポートする範囲を対象とする。

次に、学習项目的出題順を規定する。教授戦略は、学習項目を出題にふさわしい順序に並べた木構造で表現し、親要素の学習項目を理解してからでなければ子要素の学習項目は出題できないものとする。ある要素の学習項目を理解しており、その子要素が複数ある場合、その中からランダムに項目を選択し出題することを可能とする。

しかし、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、この指導順が適切と考えて作成されているが、それは作成者の主観によるものである。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となっていないと考えられる。

## 3 自動学習項目生成

人手の場合の問題を解決するために、自動で学習項目を生成する手法について検討した。

方針としては、資格試験の問題や SQL の入門書などの書籍 [3] から SQL 文を収集し、収集したデータから相関の強い予約語の組み合わせを抽出して学習項目と

する。これを実現するために、アプリオリアルゴリズムを用いた予約語間の相関ルール抽出を行う。抽出されたルールは関連の強い予約語の組を表しているため、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。抽出された相関ルールは「相関ルールツリー」として記録していく。この処理の流れを図 1 に示す。

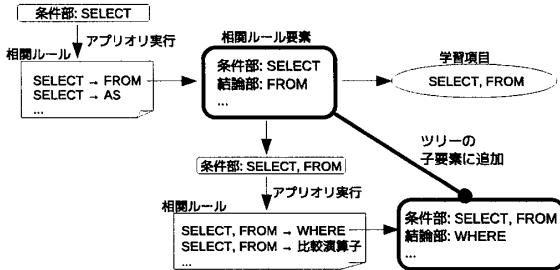


図 1 処理の流れ

学習項目自動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 2 のように表す。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なるためである。
- (2) (1) のデータについて、アプリオリアルゴリズムで相関ルールを抽出する。最初はコアとなる予約語を条件部とし、相関ルールを抽出する。コアとなる予約語は、「SELECT」、「CREATE TABLE」、「DROP TABLE」、「DELETE」、「UPDATE」である。
- (3) 抽出された相関ルールから要素を作成し、子要素として追加する。この要素は、相関ルールの条件部、結論部、サポート、確信度などの情報を持つ。作成した要素の「条件部+結論部」を新たな条件部として、相関ルール抽出を行う。
- (4) (3) を繰り返し、不適切な要素の削除（次節参照）を行い、相関ルールツリーを完成させる。この木構造は、親要素を理解していないければ子要素を出題できないという関係を表すものとなる。このツリーの各要素から、学習項目を生成する。

## 4 不適切な学習項目の削除

### 4.1 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関係などの予約語間の依存関係が存在する。この依存

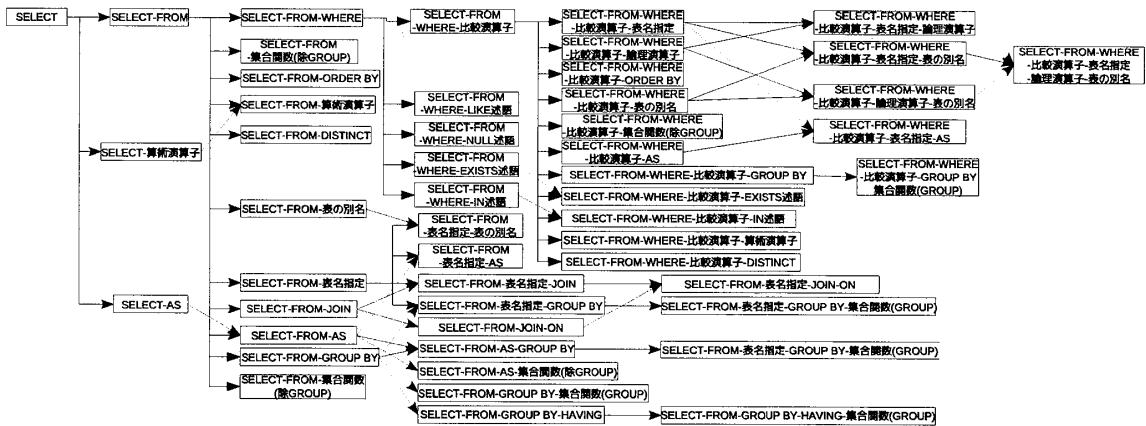


図 3 抽出された学習項目

```

SELECT, FROM, WHERE, 表名指定, 比較演算子, AND
SELECT, FROM, WHERE, EXISTS述語, 比較演算子
SELECT, FROM, WHERE, IN述語, 比較演算子
SELECT, FROM, JOIN, ON, 表名指定
SELECT, FROM, WHERE, 表名指定, AS, 比較演算子
...

```

図 2 収集した SQL 文の予約語データ

関係を考慮するために、図 4 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（角矢印）、親要素が存在する場合にいずれかの子要素との共起が必須である関係（丸矢印）を表す。

例えば、「FROM」が存在する場合「WHERE」は存在が可能である。また、「WHERE」が存在する際には「比較演算子」「IN述語」「EXISTS述語」「LIKE述語」「NULL述語」のいずれかと共に起する必要がある。

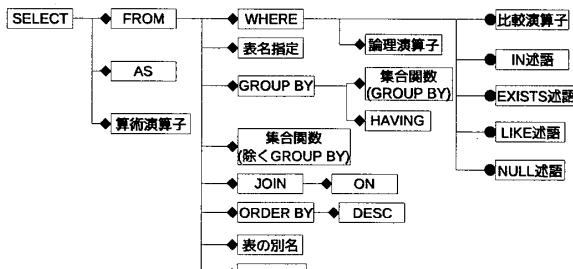


図 4 依存関係ツリー

相関ルールを抽出した際にこの依存関係と照合し、不適切なものは相関ルールツリーに加えない。

#### 4.2 重複する項目の削除

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。ある要素の子要素が既に存在する要素と重複した場合、その子要素を削除し既に存在している要素を新たに子要素とする。このようにして、重複した要素は複数の親要素をもつ一つの要素

とする。

#### 5 検証

収集した 397 例の SQL 文を用いて、提案手法の検証を行った。アプリオリアルゴリズムの実行の際の制約は最小サポート 2%、最小確信度 5%とした。今回は、コアとなる予約語「SELECT」を指定して検証した結果を図 3 に示す。生成された学習項目は 50 個であった。その中で、人手によるものに対応するものが 17 個あり、これらは人手による場合の約 82% をカバーしている。また、「SELECT-FROM-WHERE-比較演算子-表名指定-AS」のような人手では作成していない新たな学習項目も生成された。しかし、図中「SELECT-FROM-表名指定」と「SELECT-FROM-WHERE-比較演算子-表名指定」の間など、リンクが存在すべきところにリンクが無い箇所があった。また、実際の問題としては出題されず子要素へのリンクのためだけに存在する要素もあり、これらの扱いについて検討する必要がある。

#### 6 結論

抽出された学習項目は人手のものと多く一致している。ただし、それらは収集した SQL 文から得られた結果であり、より一般性の高いものであるといえる。このことから、提案手法は妥当であると考えられる。さらに、提案手法により生成された新たな学習項目は実際の SQL 利用に即していると考えられるので、より実用的な問題の出題に利用できると考えている。今回目標としたのは学習項目の生成であるが、これは教授戦略の基幹部分を生成しているといえる。今後は検証で述べた問題点を解決した上で、サポートや上位概念でまとめた予約語の出現確率などを考慮した教授戦略の生成について検討する。

#### 参考文献

- [1] 岡田, 青山, 菊池, 松本: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 2009 年総合大会講演論文集,D-15-19,2009
- [2] <http://hsqldb.org/>
- [3] 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