

学習支援を目的とした知識ベースの自動生成に関する研究

小形 淳[†] 長谷川 理[†] 砂原 悟[†] 小松川 浩[†]

千歳科学技術大学[†]

1. はじめに

我々は初等中等教育課程の理数系を中心とした体系的な知識定着教材を保有する e ラーニングシステム(CIST-Solomon)の実証開発を行っており,平成 19 年度末現在では,数学・物理・化学等を中心に教材数は約 1 万 5 千を超えている.特に数学は,小学校高学年から高校 3 年に至る学習指導要領に準拠した体系的な教材の整備を行ってきた.また 60 を超える教育機関(2 万人を超える利用者)での運用に至っている[1].

本研究では,こうした実証フィールドを活用し,e ラーニング教材群の類似度と利用状況に応じた知識ベースの自動生成を試みる.特に,生成する知識の粒度を e ラーニングで学習するコース(1 回の授業)程度とすることで,自律的な学習支援への可能性も模索する.

2. 知識ベースの作成

e ラーニングで自学自習を行う際に必要とする学習内容を検索する場合,キーワード検索を行うのが一般的である.しかし,入力するキーワードの内容によって提示されるコンテンツの数が不均一となる場合があり,必ずしも学習者が必要とする内容になっているとは限らない.そこで,e ラーニングが保有するコンテンツや学習者全体の学習履歴情報を基にした,知識の繋がりを示すアルゴリズムを作成した.

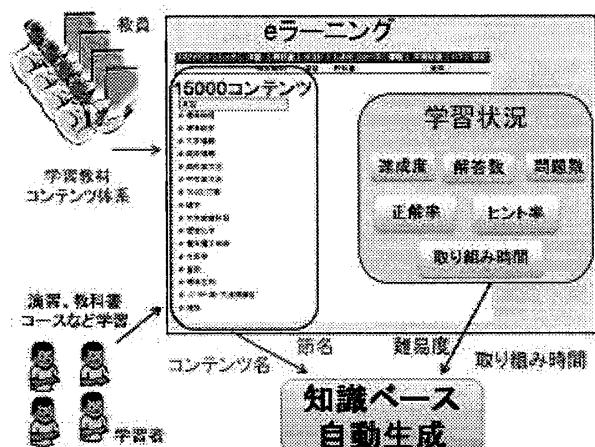


図 1: アルゴリズムモデル

作成したアルゴリズムのモデルを図 1 に示す.教員はコンテンツがどの分野なのかといったコンテン

ツ体系を登録する.学習者は演習問題や教科書,教員の提示したコースを取り組み,その取り組み時間や達成度などの学習状況を e ラーニングに蓄積していく.動的に変化し続けるこれらの情報を基に知識ベースを自動生成した.

2.1 構文解析

節名と教材名に含まれる「の」や「と」,「基礎,応用,例,計算」などの直接コンテンツの中身に関係しない情報の除去を行い,キーワードを抽出する.しかし,複数の教員が節や教材を作成しているため,同義語などの表記ゆれが起こる.そこで e ラーニングに実装されている検索機能で,同義語や含有語などを検索結果に反映するためのオントロジーを参照し,同義語の表記ゆれを修正している.例えば「sin,サイン,正弦」というキーワードは「sin」で統一される[2].

2.2 類似度の計算

節ごとに 2.1 で抽出されたキーワードに階層的クラスタリング手法(最短距離法)を用い,式 1 より類似度を算出する.図 2 に $n(i), n(j), n(k)$ の類似度を算出したマトリックスを示す.この場合 $n(i), n(j)$ の類似度が 0.33 で類似しているといえる.

$$\text{Max}(r_{ij}) = \{n(i) \cap n(j)\} / \{n(i) \cup n(j)\} \quad (1)$$

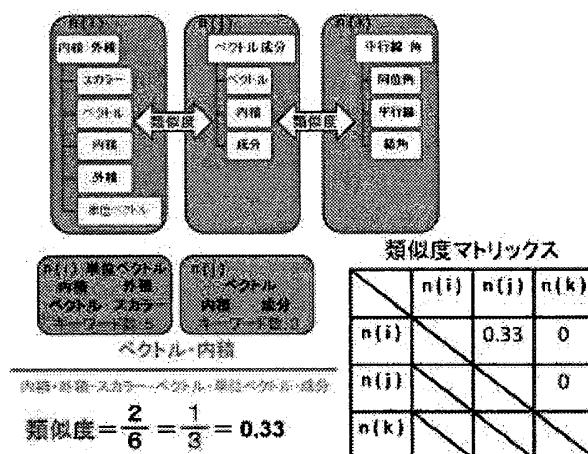


図 2: 類似度計算及び結合例

2.3 節の結合

類似した節を結合する.図 2 に例を示す.類似

していると判断された節 $n(i), n(j)$ を結合する。結合した集合は名前を「ベクトル,内積」とし「内積,外積,スカラー,ベクトル,単位ベクトル,成分」というキーワードのコンテンツを持つ。この集合を知識とする。

2.4 クラスタリングの制御

知識の結合を行う際に, 取り組み時間を評価関数に利用している。類似度のみの結合を行った場合, キーワードによって知識の粒度にばらつきが起きる。例えば三角関数という知識は, 膨大な数の節と結合してしまい, 学習者に提示するには困難な知識となってしまう。そこで, コンテンツの取り組み時間から節ごとの平均取り組み時間を算出し, 一つの知識が 150 分以下となるようにルールを追加した。また, 教育機関ごとの取り組み時間を反映した知識の作成が可能となった。

これら 2.1~2.4 で作成した知識群を知識ベースと呼ぶ。知識ベースは知識習得のため, 基礎知識から専門知識までの知識の繋がりを重視し, e ラーニングからコンテンツの学年情報を取得し, 中学, 高校, 大学の教育機関ごとに作成した。

3. オントロジーの作成

作成した知識ベースに対し, owl [3] を用いてオントロジーの適用を図った。オントロジーの作成には jena フレームワーク [4] を用いた。シソーラスは単に語彙の差を吸収するだけでなく, is-a, has-a の関係が明確に示せるため, 学習手順のひとつの指標になると考えられる。図 4 に作成したオントロジーを可視化した図を示す。

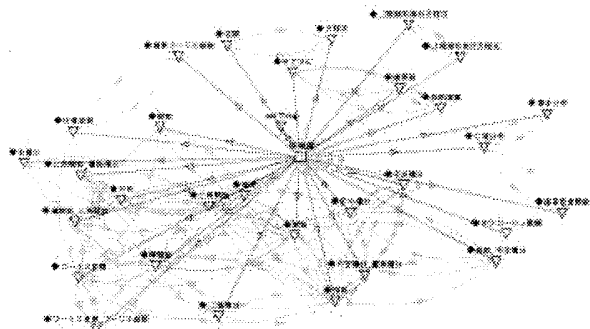


図 4 : 作成した大学の知識を可視化した図

4. e ラーニングへの適用

e ラーニングへの適用にあたり, 学習者側のシステムに新たにコース自動生成機能を追加した。これは, 学習者が習得したい知識を知識の一覧から選択すると, システムが知識習得に必要なコンテンツを選択しコース形式で提示するシステムである。図 5 に例を示す。図 5 は知識一覧から「フーリエ級数」を選択した画面である。選択された知識に関連する

コースは「基礎」と「標準」の二つの難易度から選択することが可能である。難易度は e ラーニングの取り組み履歴にある正解率などから標準偏差を算出し, 基礎と標準に分類している [5]。

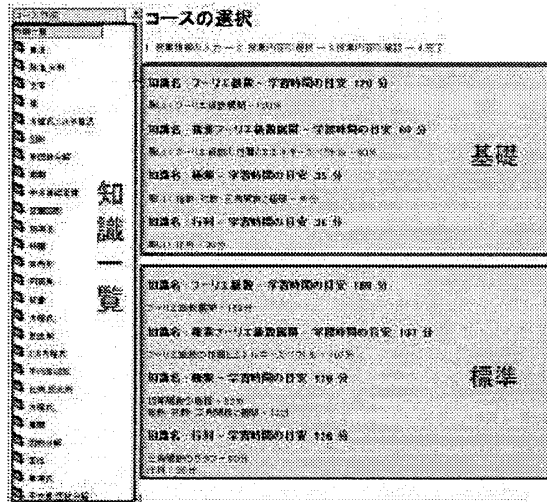


図 5 : コース自動生成画面

5. おわりに

e ラーニングの膨大なコンテンツと長年の利用に伴い蓄積され, 日常に変化していく様々な学習情報を基に動的に知識ベースの作成を行い, 学習者に知識習得を目標とするコースを提示し検証を行った。

今後の課題として知識ベース生成エンジンのブラッシュアップを行いノイズの除去や関連知識の少ないコンテンツへの知識補充をどのように行うか工夫する必要がある。また, 授業と連携した環境での実証評価を行い大規模な評価を行う必要がある。

参考文献

- [1] 小松川浩 : 理工系の知識共有に向けた e-Learning の実証研究, メディア教育研究第 1 巻第 2 号, 11-22 (2005)
- [2] 砂原悟, 友田成則, 今井順一, 小松川浩 : コンテンツ検索のためのオントロジーの構築と e-Learning システムへの実装, 情報処理学会第 69 回全国大会論文集, 445-446 (2007)
- [3] W3C:OWL Web Ontology Language Reference, <http://www.w3.org/TR/owl-ref/>
- [4] Jena - A Semantic Web Framework for Java <http://jena.sourceforge.net/>
- [5] 庄司慶市, 上野春毅, 小松川浩 : 個別指導エージェントによる適応型 e ラーニングシステムの開発と評価, 教育システム情報学会第 31 回全国大会講演論文集, 317-318 (2007)