

RK-002

講義資料からの意味ネットワーク作成による学習事項の形式化 Formalization of Learning Items with Semantic Network from Lecture Notes

大野 公裕[†]
Masahiro Ono

原田 史子[‡]
Fumiko Harada

島川 博光[‡]
Hiromitsu Shimakawa

1. はじめに

大学において、教員が多数の学習者に対して指導しなければならない科目がある。このような科目で、教員がすべての学習者に対しマン・ツー・マンで指導することは困難である。よって、このような科目では Teaching Assistant(以降 TA と表記する) が配備される。TA は教員とともに学習者を指導しなければならない。

教員が多数の学習者に対し指導を行わなければならない科目としてプログラミング演習がある。プログラミング演習において、TA は個々の学習者に対し、学習者が習得すべきと教員が考える内容を指導し、その内容を学習者が理解しているかどうかを確認する。しかし教員が教えようとする内容は、通常複数の知識や概念とそれらの間の関係からなる。これら知識や概念とその関係を、教員は TA に正確に把握させなければならないが、複数の項目とそれらの関係を正確に伝達することは容易ではない。習得させるべき内容を TA が正確に把握していなければ、学習者がその内容を習得できているかを確認するさいに、確認漏れが起こる。また、複数の TA が設けられている場合、これら内容を学習者が習得できているかを判断する基準は TA によって異なる。これにより、TA ごとに異なった判断基準で指導が行われるため、すべての学習者に対し均質な指導を行うことができない。

教員の考える学習者が習得すべき内容を TA が正確に把握するためには、TA と教員の間でその内容を共有していなければならない。そのためには、教員がその内容を形式的に表現することが必要である。本論文では、TA が教員の考える学習者が習得すべき内容を把握できるようにするため、その内容を形式化する手法を提案する。本研究では、知識や概念を形式的に表現することのできるモデルである意味ネットワークに着目した。本手法では、教員の考える学習者が習得すべき内容である知識や概念を講義資料から抽出し、それらを意味ネットワークとして形式的に表現する。

2. TA の学習事項の把握

2.1 プログラミング演習における TA の役割

大学において、教員が多数の学習者に対して指導するプログラミング演習では、一般的に指導者が学習者に対しマン・ツー・マンで指導するので、教員を補助するため多くの TA が配備される。教員や TA は、個々の学習者に対し指導すべき内容を指導し、その指導すべき内容を学習者が習得しているか確認する。演習課題を解く過程で理解につまずいた学習者は、理解につまずいた内容に関する質問を TA に投げかける。TA は質問を投げかけた学習者に対し、教員の指導すべき内容のうち、どの内

容でつまづいたかを問答を交わすことで特定し、その内容について指導を行う。

2.2 学習事項の把握

TA は、学習者に指導するさい、教員が学習者に指導すべき内容を正確に把握していなければならない。本論文では、教員が学習者に習得させるべき内容を、教員の考える学習事項と呼ぶことにする。TA は、学習者と問答を交わすことで、学習者が学習事項を確実に理解できているか確認する。このさい、TA が学習事項を把握していなければ、学習者が理解しなければならない学習事項の確認漏れが起こる。また、複数の TA が設けられている場合、学習者が学習事項を理解しているかどうかを判断する基準は、TA によって異なる。この判断基準の設定が、TA 個人の裁量に任されると、学習者に対して均質な指導が行われない。このため、TA は教員の考える学習事項を正確に把握することが必要である。

2.3 学習事項の形式化

TA が学習事項を正確に把握するためには、TA と教員の間で学習事項を共有しなければならない。学習事項を誤解なく共有するためには、教員の考える学習事項が形式的に表現されなければならない。学習事項は、講義で強調されるキーワードと、そのキーワードが示す概念とみなせる。例えば「整列」というキーワードに対して、「整列」というキーワードそのものに加え「要素 $S_1 \dots S_n$ が与えられたときに、ある規則にしたがって要素を並べ替えること」という定義や、整列が「基数ソーティング」と「比較に基づくソーティング」の2種類をもつといった概念が、学習事項として挙げられる。

知識を形式的に表現する手法として、意味ネットワーク [1][2][3] が存在する。意味ネットワークとは、ある知識や概念をひとつの「ノード」として表現し、ノード同士の関係を「リンク」と呼ばれる線で結ぶことにより、多様な知識のつながりをネットワーク構造で形式化するモデルである。このような意味ネットワークにおいて、知識を講義で強調されるキーワード、概念をそのキーワードが示す概念とすると、学習事項を形式的に表現できると考えられる。

3. 意味ネットワークによる学習事項の形式化

3.1 講義資料からの学習事項の形式化

本論文では、学習事項を TA が正確に把握できるようにするため、学習事項を形式的に表現する手法を提案する。本手法では、学習事項は意味ネットワークとして形式的に表現される。講義では、Power Point などで作成されたスライド形式のファイルが講義資料として用いられ、この講義資料には学習事項が含まれている。そこで本研究では、教員が作成した講義資料から、学習事項の意味ネットワークを作成することで学習事項を形式化することを目的とする。

[†]立命館大学大学院 理工学研究科

[‡]立命館大学 情報理工学部

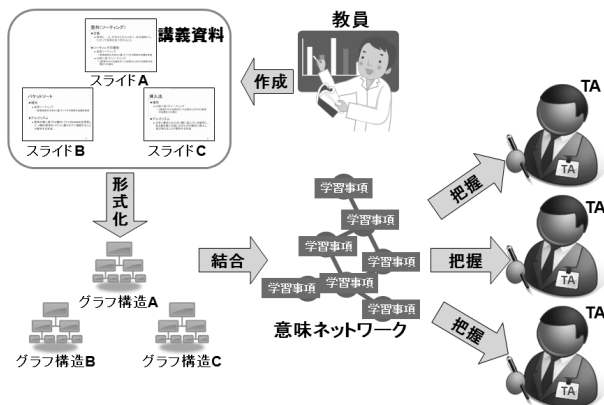


図 1: 意味ネットワークによる TA の学習事項の把握

講義資料は、複数のスライドから構成され、それぞれのスライドに学習事項が含まれている。さらにひとつのスライド内では、複数の学習事項が列挙される。また、ひとつの学習事項と、より詳しい内容を表す学習事項が階層的に表現される。提案手法では、このようなスライドの表記方法に基づき、学習事項を表す意味ネットワークを作成する。提案手法の概要を図1に示す。

まず、講義資料のそれぞれのスライドにおいて、学習事項を抽出し、その学習事項をスライド内の構造にしたがってグラフ構造で形式化する。これをスライド別グラフ構造と呼ぶ。そして、スライド別グラフ構造を講義資料の全スライド間で統合することにより、講義資料における学習事項の意味ネットワークを作成する。本手法により、学習事項を形式的に表現することで、TAは学習者に指導すべきことを正確に把握することができる。また、全員のTAが学習事項を把握できるため、すべての学習者に対する均質な指導が実現できる。

3.2 スライドからの木構造成成

講義資料の各スライドから学習事項を抽出し、スライドの構造に従って木構造として表現する。一般的にスライドの列挙文には階層関係があると考えられる。あるひとつの列挙文Aを、より詳細に列挙文Bを用いて説明する場合は、列挙文Aの次の行に字下げされた新たな列挙文Bが加えられる。列挙文Bは、列挙文Aを親の階層とする階層関係をもっている。

本手法では、このような各スライドにおける列挙文の階層関係を利用し、スライドから木構造を抽出する。木構造は、スライドのスライドタイトルと各列挙文を、それぞれノードとする。ある列挙文Aが列挙文Bの親階層であるとき、Bに対応するノードの親ノードをAに対応するノードとする。すなわち、AのノードからBのノードへのエッジを生成し、Aを親ノードと定める。

ソーティングを例にして、スライドから木構造を作成する過程を図2に示す。まず、スライドタイトルである「整列（ソーティング）」が木構造のルートノードとなる。このルートノードに、列挙文から作成されたノードが子ノードとしてつながられる。列挙文から作成されたノードは、列挙文の階層関係に基づいて作成される。たとえば、スライドにおいて「基数ソーティング」の親の列挙文が「ソーティングの種別」であるため、「基数ソーティ

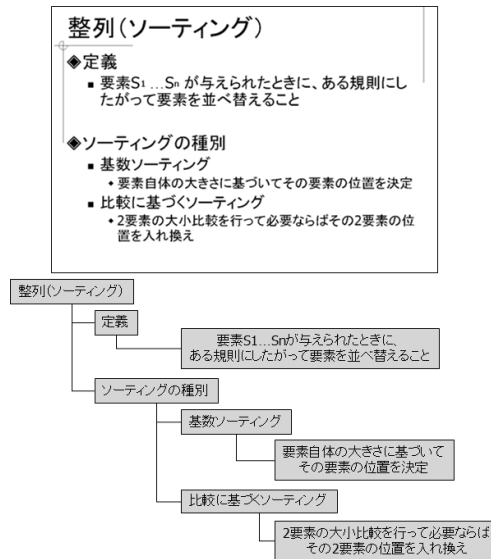


図 2: スライドからの木構造成成の例

ング」の親ノードは「ソーティングの種別」となる。また、「定義」や「ソーティングの種別」といった親の階層を持たない列挙文は、スライドタイトルの「整列（ソーティング）」を親ノードとする。このようにして、スライドから図2下部のような木構造が作成される。

3.3 スライド別グラフ構造の作成

一般的に教員は、スライドの列挙文中で重要だと位置づける文字列に対し、文字の色を変更する、下線を引くなど強調して表示する。この強調表示された文字列は、学習事項の中で重要なものである。本研究では、強調表示された1文字列をひとつの学習事項とみなして意味ネットワークを作成する。そこで、3.2節の手順にしたがってスライドより作成された木構造のノード縮約を行い、強調表示された各学習事項をノードとするスライド別グラフ構造を作成する。まず、3.2節の手順にしたがって作成された木構造のノード集合の中で、強調表示された文字列を含む列挙文に対応するノードとスライドタイトルに対応するノードを選定する。選定されないノードは、そのノードに接続されたエッジとともに削除する。ただし、選定されていないノードの子孫ノードに、選定されたノードが含まれる場合、その選定されていないノードは削除しない。これは強調表示の文字列を含む列挙文の親列挙文が強調表示されていない場合でも、子の列挙文とスライドタイトルの間の意味的、あるいは概念的なつながりを保つためである。この手順により、各スライドに含まれる学習事項を的確に木構造のノードとして抽出することができる。また、ある選定されたノードに対応する列挙文に、強調表示された文字列が2つ以上含まれる場合がある。このとき、当該ノードを各強調箇所に対応する複数のノードに分割する。分割により生成された各ノードは、分割前のノードの親ノードと子ノードを、自身の親ノードおよび子ノードとして持つ。このとき木構造はグラフ構造に変化する。

赤字表記を強調表示として用いている場合の、ノード

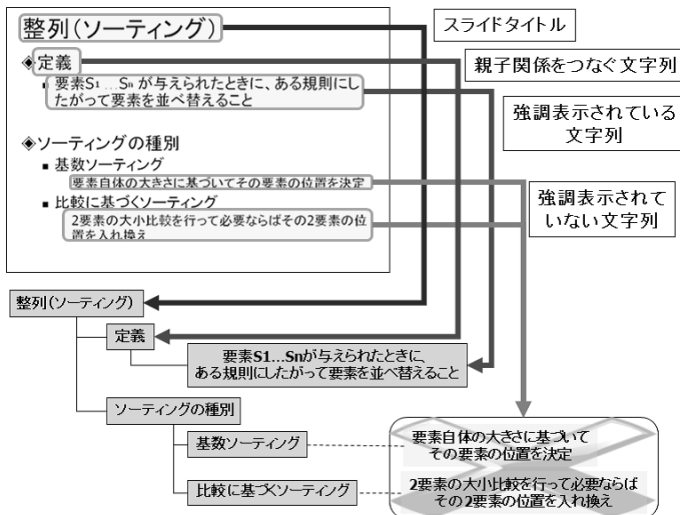


図 3: 木構造からスライド別グラフ構造の作成

を選定する例を図3に示す。図3の上部に示すスライドから、スライドタイトルと各列挙文に対応する、合計8つのノードが抽出される。図3において、「要素 $S_1 \dots S_n$ が与えられたときに、ある規則にしたがって要素を並べ替えること」、「基数ソートリング」、「比較に基づくソートリング」の3つの文字列が赤字で強調表示されている。このそれぞれの文字列に対応するノードと、スライドタイトルに対応するノードを選定する。ここで、ルートノードと「要素 $S_1 \dots S_n$ が与えられたときに、ある規則にしたがって要素を並べ替えること」ノードの中間ノードである「定義」ノードは、強調表示されていないため、選定されない。しかし、「要素 $S_1 \dots S_n$ が与えられたときに、ある規則にしたがって要素を並べ替えること」ノードとルートノード「整列（ソートリング）」との概念的つながりを保つために「定義」ノードは削除せずに残される。「ソートリングの種類」ノードも同様である。その他の強調表示を持たないノード「要素自体の大きさに基づいてその要素の位置を決定」、「2要素の大小比較を行って必要ならばその2要素の位置を入れ換え」は削除される。これにより、強調表示された文字列に対応するすべてのノードは、そのスライドタイトルであるルートノードを祖先に持つ構造になる。図3の例では、図3の下部に示すようなグラフ構造が抽出され、これがスライド別グラフ構造となる。子ノードは親ノードの内容に関連する事柄を表現でき、そのスライドにおける学習事項を的確に抽出することが可能になる。

3.4 講義資料に対応する意味ネットワークの作成

各スライドにおける学習事項を表すスライド別グラフ構造を結合することで、講義資料全体として学習事項が形式的に表現される。よって、各スライドから作成されたスライド別グラフ構造を結合することで、講義資料において学習事項が形式的に表現された意味ネットワークを作成する。

講義資料において、複数のスライドで同じ文字列が強調表示されている場合がある。このとき、それらのスライドで、学習事項が重複していることを表している。本手法では、複数のスライドから作成したスライド別グラフ構造を、スライド間で重複した学習事項である、強調表示された文字列のノードで結合する。そのため、複数

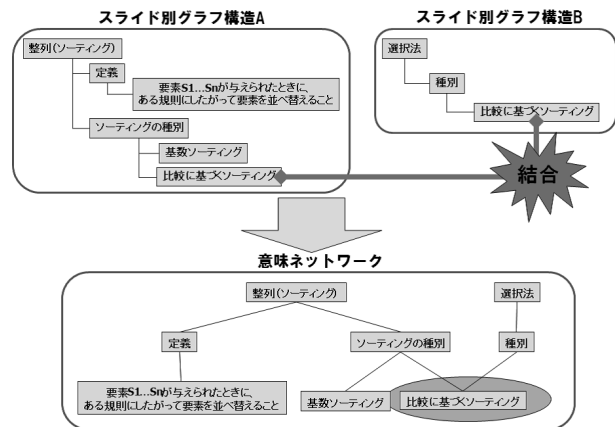


図 4: スライド別グラフ構造の結合

のスライドのグラフ構造で重複するノードは、講義資料の意味ネットワークでは、同一のノードとして扱われる。これにより、重複するノードを持つスライド別グラフ構造同士は、その重複するノードで結合され、ひとつの意味ネットワークとなる。

スライド別グラフ構造を結合し、意味ネットワークを作成する例を図4に示す。図4は異なる2つのスライドから、それぞれスライド別グラフ構造AとBが作成されたことを示す。グラフ構造Aとグラフ構造Bには「比較に基づくソートリング」という同じ文字列に対応するノードが存在している。そこで、両グラフ構造の「比較に基づくソートリング」ノードを同一のノードとみなして、両グラフ構造を結合し、ひとつの意味ネットワークを作成する。意味ネットワークは、各スライド別グラフ構造のすべてのノードをノード集合としてもつ。2つ以上のスライド別グラフ構造のノードが同一のノードとして結合されたとき、結合後のノードの親ノードと子ノードは、それぞれ結合前の各ノードの親ノードの和集合および子ノードの和集合となる。これにより図4の下部のような意味ネットワークが生成される。このようにして、各スライドから作成したグラフ構造を結合し、講義資料全体としてひとつの意味ネットワークを作成する。

4. 評価実験

4.1 実験環境

提案手法により講義資料から学習事項を形式化することができるかを検証するための評価実験を行った。本実験では、立命館大学の情報理工学部1回生を対象とした講義「データ構造とアルゴリズム」で担当教員Aが作成した「ソートアルゴリズム」に関する講義資料から意味ネットワークを作成し、学習事項を形式化することができているかを検証する。講義資料は上記講義の教科書のソートアルゴリズム記述部[4]に基づき作成されている。この記述部は、クイックソート、バケットソート、ヒープソート、マージソート、選択法、挿入法から構成される。

講義資料は全65枚のスライドから構成されているが、その中で「演習課題」や「講義連絡」などの提案手法の対象とはならないスライド24枚は除外した。この講義資料から意味ネットワークを作成し、学習事項を形式化

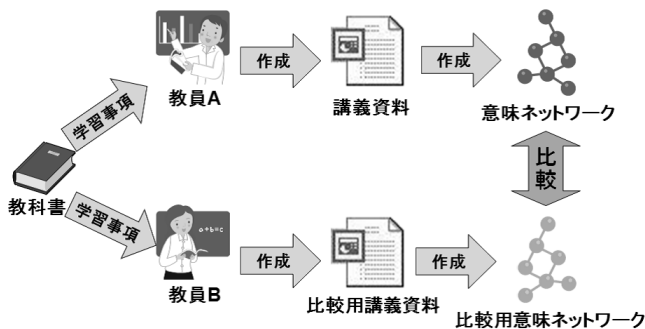


図5: 実験の流れ

しているかを検証する。

4.2 実験の流れ

図5に実験の流れを示す。本実験では、前述した講義資料から作成した意味ネットワークとは別に、比較するための意味ネットワークを用意する。比較するための意味ネットワークは、教員Aとは別の教員Bが同じ教科書[4]の記述部から作成した比較用講義資料から、提案手法を用いて作成する。教員Aと教員Bは、どちらも同じ教科書の学習事項をもとに講義資料を作成している。そのため、どちらの講義資料にも、同じ学習事項が含まれていると考えられる。それぞれの講義資料から作成された意味ネットワークがもつ学習事項が一致すれば、提案手法により学習事項を正しく形式化できているといえる。本実験では、それぞれの講義資料から提案手法を用いて作成された意味ネットワークの一致性を算出することで、学習事項を形式化することができるかを検証する。

4.3 実験結果

教員Aの講義資料から作成した意味ネットワークと、教員Bの比較用講義資料から作成した意味ネットワークの各ノードを比較し、クイックソート、バケットソート、ヒープソート、マージソート、選択法、挿入法の各項目について一致率を求めた結果を表1に示す。

表1: 意味ネットワークのノード一致率

ソートアルゴリズム	一致率
クイックソート	47%
バケットソート	10%
ヒープソート	25%
マージソート	50%
選択法	13%
挿入法	22%
平均	28%

本実験結果より、それぞれの講義資料から作成した意味ネットワークには十分な一致性はみられなかった。これには大きく2つの要因が考えられる。

1つ目は、講義資料は作成する教員によって独自の工夫が凝らされるためであると考えられる。教員Aが作成した実際の講義で用いられた講義資料には、教員A独自の工夫により作成されたスライドが含まれる。一方、教員Bが作成した講義資料は、教科書から抜粋した学習事項をそのまま用いて作成されている。そのため、教員Bの意

味ネットワークには教員Aの独自の工夫によるノードが含まれず、これが一致率の低下につながった要因の一つと考えられる。

2つ目に、提案手法では図式から学習事項を抽出することができないことが考えられる。教員Aが作成した講義資料には、図式を含むスライドが多く存在し、提案手法の対象となるスライド41枚中19枚のスライドに図式が含まれていた。アルゴリズムを説明する場合、図式やアニメーションなどを用いて視覚的に理解させることが有効であるとされている[5]。そのため、ソートアルゴリズムに関する学習事項は、講義資料における図式の中にも含まれている可能性が高い。しかし、提案手法ではスライドタイトルと列挙文中から学習事項を抽出しているため、図式に含まれる学習事項を抽出することはできない。そのため、図式を多く含む教員Aの講義資料から作成された意味ネットワークには、学習事項の抽出漏れがあると考えられる。よって、この抽出できなかった学習事項が一致率の低下につながった要因の一つと考えられる。

これら2つ要因を考慮し、教員独自の工夫により作成されたスライドを識別し、スライド別グラフ構造を作成するスライドから除外する。また、講義資料中の図式も含めて学習事項を抽出することで、より正確な学習事項の抽出が可能になると考えられる。

5. おわりに

本論文では、教員の考える学習事項をTAが正確に把握できるようにするため、スライド形式の講義資料から学習事項を抽出し、その学習事項を意味ネットワークとして形式的に表現する手法を提案した。本手法の有効性を検証するための実験を行った結果、平均28%の一致率が算出され、提案手法によって講義資料から抽出された学習事項に高い一致性は見られなかった。その要因として、教員による独自の工夫が凝らされたスライドが講義資料に含まれていたことと、本手法ではスライド内の図式から学習事項を抽出することができないことの2つが考えられる。

今後の課題として、教員による独自の工夫が凝らされたスライドを識別すること、またスライド内の図式から学習事項を抽出することを考えている。

参考文献

- [1] A.M.Collins, E.F.Loftus: A Spreading-Activation Theory of Semantic Processing, *Psychological Review*, Vol.82, No.6, pp.407-428, 1975.
- [2] A.M.Collins, M.R.Quillian: Retrieval Time from Semantic Memory, *Journal of Verbal Learning and Verbal Behavior*, Vol.8, pp.240-247, 1969.
- [3] 佐藤 基治, 坂元 滋信: 活性化拡散モデルに関する実験的研究, 福岡大学人文論叢, Vol.38, No.3, pp.729-762, 2006.
- [4] 渡邊 敏正: データ構造と基本アルゴリズム, 共立出版, pp.127-150, 2000.
- [5] 阿部 哲也: アルゴリズム教育のための教材の開発, 日本教育工学雑誌, Vol.27, pp.45-48, 2004.