

E-33

仮想教室の学習コミュニティにおける発想支援エージェントの検討 Information Support System in the Virtual Class Room System for e-Learning Community

山田一郎 有安香子 住吉英樹 柴田正啓 八木伸行
 Ichiro Yamada Kyoko Ariyasu Hideki Sumiyoshi Masahiro Shibata Nobuyuki Yagi

1. まえがき

我々は、放送局が所有するコンテンツを教育に活用するマルチメディア教育支援システムの研究を進めている。今回、ネットワーク上に仮想的な教室を作成し、話し合いを通じた授業環境を実現した[1]。この仮想教室での話し合いには、複数の生徒以外に、多種の情報を持つエージェントが参加し、会話が途切れた際に、その内容に関連する新たな話題の提示を行う。本稿では、この話題提示によって発想を支援するエージェントの実現のために、放送用ニュース原稿に出現する重要な用語と説明文、さらに説明文の意味的な役割を自動抽出し、会話内容に関連した情報として提示する手法を提案する。

2. 仮想教室の会話に関連するニュース話題抽出

発想支援エージェントは、仮想教室で生徒が何について会話しているか推定し、その内容に関連する情報を質問形式で話題として自発的に提示する。これらの処理の情報源に、NHKの放送用読み原稿として使われるニュース原稿を利用した。このニュース原稿から社会の話題を抽出して、仮想教室における会話内容との比較を行い、最も関連が深いニュース原稿中の話題を利用する。

以前、我々はニュース原稿から一定期間に発生した話題を抽出する手法を提案した[2]。表1に2002年5月のニュース原稿から抽出した話題の一部を示す。

- ・ 亡命を求めた北朝鮮住民を拘束した問題
- ・ 鈴木宗男衆議院議員逮捕の秘書事件
- ・ サッカー・ワールドカップの国内の開幕戦
- ・ 大相撲の夏場所の休場
- ・ 福岡県看護師仲間保険金殺人事件
- ・ イスラエル軍が包囲の問題
- ・ 個人情報保護法案など重要法案

表1. 2002年5月の話題 (一部)

この手法により抽出された各話題は、話題性により重み付けされた単語をベクトルの要素とするベクトル空間に配置される(以後、話題ベクトル)。そこで、仮想教室における会話に出現する名詞を監視し、直近の5つの名詞をベクトルの要素とする単位ベクトルと、ニュース原稿から抽出された話題ベクトルとの内積が最大のものを、その瞬間における仮想教室の会話に関係が深いと判断する。この際、長期に継続する話題は、トラッキング処理[3]により同じ話題と判断している。

3. ニュース原稿を利用した用語集作成

前章で抽出した仮想教室の会話に関係が深い話題を構成するニュース原稿から、重要な用語と説明文を抽出し、説

明文の意味的な役割を推定した結果を、会話の活性化を図る発想支援の情報として利用する。

重要な用語がニュースで扱われる場合は、視聴者が容易に理解できるよう、用語の説明を伴うことが多い。この説明部分を利用して用語集を作成する。

まず、ニュース原稿集合から、対象とする用語を抽出する。ここでは、強調を意図する鍵括弧で囲まれた名詞句に限定して抽出した。鍵括弧で囲まれていても一般的な語には定義を抽出する必要が無いため、ここでは辞書に登録されている語は処理対象から除いている。

次に、用語を説明する部分を抽出する。ニュース原稿において用語を説明する場合、連体修飾節などの従属節が多用される[4]。そこで、本手法では、用語を修飾する連体修飾節を処理対象とした。

処理対象として抽出された用語と連体修飾節との意味的關係には、様々な種類が考えられる。本手法では、語の意味記述を体系付けた Qualia structure[5]を参考にして分類する。Qualia structure では意味記述の要素として、以下の4つの役割を定義している。

- A) Formal role : 他の語と区別する情報
- B) Constitutive role : 語が持つ内部的な性質、構成要素
- C) Telic role : 語が持つ典型的な機能・目的
- D) Agentive role : 語の起源、語が引き起こす事象

この意味記述の役割を、用語を修飾する連体修飾節の役割に当てはめる。下記の例では、下線部の連体修飾節が、鍵括弧で囲まれた用語における Qualia structure(A~D)のそれぞれの役割となる。

- A) ビル街や舗装された道路に囲まれた都会の気温が上昇する「ヒートアイランド現象」が、～
- B) この他、細長い花卉を水牛の角のように張る「スカーホセバラム」の仲間など～
- C) 森林の整備や緑化の推進を目的とする「緑の募金」を、～
- D) ハタミ大統領が提唱する「文明間の対話」を、～

ニュース原稿では5W1Hの情報を的確に伝えることが重要な役割とされている。そこで、5W1Hの関係や特徴となるConstitutive roleについては、さらに細分化し、「Time(時間)」、「Location(場所)」、「Instrument(道具)」、「Contain(包含)」、「Is_a_member_of(メンバー)」、「Is_a_part_of(部分)」そして、それ以外のConstitutive roleの要素を「Constitutive」に分類した。

連体修飾節の役割を推定するために、最大エントロピー法による学習[6]を利用した。学習における素性は、連体修飾節中の、用語を直接修飾する動詞の標準表記・時制・態・完了形の有無、動詞の格、その格が含まれる節の自立語表記、自立語の属性、用語の属性とした。ここで、自立語と用語の属性は IREX 固有表現抽出タスクにより付加された8つの固有表現(組織名、人名、地名、固有物名、日

付表現、時間表現、金額表現、割合表現)を利用し、抽出アルゴリズムは内元らの手法を利用した[7]。表2に素性の例を示す。

千葉県柏市を ホームタウンと する「柏レイソル」～			
表記	千葉県柏市	ホームタウン	する
属性	地名	その他	組織名
格助	を格	と格	
動詞の時制・態・完了形の有無 現在形・能動態・無			

表2. 付加する素性例

2002年6月のニュース原稿に出現した670組の連体修飾節と用語を対象として、その意味的な役割を推定する実験を行った。学習データは、実際の2002年5月以前のニュース原稿から連体修飾節と用語を抜き出し、その3712組にQualia structureの役割を手作業で与えて作成した。学習データに少数しか出現しない素性はノイズとなる可能性があるため、出現頻度が8回以上観測された素性のみ利用している。処理の結果、各役割に所属する確率値が出力される。この値が一定値(実験では0.5)以上のものを、用語を修飾する連体修飾節の役割とした。評価結果を表3に示す。

役割	適合率	再現率
Formal	0.853	0.858
Telic	0.891	0.661
Agentive	0.792	0.811
Constitutive	0.683	0.545
Time	1.000	0.286
Location	0.900	0.900
Instrument	1.000	0.571
Is_a_member_of	0.684	0.650
Is_a_part_of	1.000	0.500
Contain	0.625	0.500
Total	0.814	0.760

表3. 用語と連体修飾節の意味的役割判定結果

この処理を、毎日更新されるニュース原稿を対象にして行い、結果をデータベースに登録することにより、常に最新の用語を含む用語集システムを構築した。この用語集は、エージェントの重要な知識源となる。

4. 発想支援エージェント

エージェントは、仮想教室に参加する生徒の会話を監視し、一定時間発話が無い場合に、会話が行き詰っていると判断して、その内容に関連する質問を行い会話の活性化を図る。この際、まず、生徒の会話に関連が深いニュース話題を抽出する。次に、その話題を構成するニュース原稿から抽出された用語を、提供する質問の候補とする。この質問の候補とされた用語の中で、生徒の会話に最も関連が深い用語を質問対象として抽出し、その用語を利用した質問文を生成する。

生徒の会話に最も関連が深い用語を抽出する処理では、各用語に、Formalと判定された連体修飾節に含まれる単語をベクトルの要素に持つ用語ベクトルを定義する。用語ベクトルの要素の重みは、その用語が抽出されたニュース原稿が属する話題の話題ベクトルの要素の重みと同じ値とし

た。Formalの役割の連体修飾節は、用語を他の用語と区別する説明を与えるため、用語を特徴付ける指標となる。そのため、用語ベクトルは、Formalの役割の連体修飾節のみを利用して定義した。

この用語ベクトルと、仮想教室における会話に出現する直近の5つの名詞をベクトルの要素とする単位ベクトルとの内積が最大のものを会話に関連した用語であると判断し、質問対象とした。すべての用語ベクトルとの内積が0の場合は、用語ベクトルのノルムが最大のものを質問対象用語とした。ここで、会話に既出の用語は、処理対象から除いている。

質問は、あらかじめ決めておいたテンプレートを利用し、質問対象用語について問いかける。例えば、「“ヒートアイランド現象”って何か知ってる?」といった質問文を生成した。生徒から「知らないから教えて」と言われた場合、Formalと判定された連体修飾節を利用して、「ビル街や舗装された道路に囲まれた都会の気温が上昇する現象のことだよ」と返答できる。この対話処理は、オートマトンによる単語のパターンマッチングを基本とする住吉らの手法を利用した[8]。続けて、その用語について質問があった場合、他の役割と判定された連体修飾節を利用して、「どういった特徴?(Constitutive)」、「何のため?(Telic)」、「いつ?(Time)」、「どこ?(Location)」などの質問の回答も可能となる。

5. まとめ

ニュース原稿を解析して、重要な用語と、その役割が推定された説明文を自動抽出する手法を提案した。この用語集を利用して、インターネット上の仮想教室における学習で、エージェントは、会話内容に関連した情報を、適宜、提供することができる。

今後、新しい教育放送サービスの実現に向けたシステムとして、評価、改良を進める予定である。

謝辞

本研究を進める上で、実験システムのプログラミングに協力して頂きました小杉広和氏((株)漢字情報サービス)に感謝いたします。

参考文献

- [1]住吉ほか: “学習コミュニティの対話を支援する仮想教室のシステム化”, FIT2002 情報科学技術フォーラム, Sep.(2002)
- [2]I. Yamada et al.: “Topic Event Detection using Japanese News Articles”, 5th NLPRS1999, pp375-380, Nov. (1999)
- [3]山田ほか: “ニュース原稿を利用した話題トラッキングの検討”, 情処学会第61回全国大会, 5U-6, pp193-194, Oct.(2000)
- [4]山田ほか: “ニュース原稿を利用した用語集作成の検討”, 言語処理学会第8回年次大会, pp507-510, Mar.(2002)
- [5]Pustejovsky, J.: “The generative Lexicon. Cambridge”, MA: MIT Press. (1995)
- [6]北: “確率的言語モデル”, 東京大学出版社
- [7]内元ほか: “最大エントロピーモデルと書き換え規則に基づく固有表現抽出”, 自然言語処理 Vol. 7, No. 2, pp63-90, Apr. (2000)
- [8]住吉ほか: “エージェントを利用したマルチメディア教育支援システム”, 電子情報通信学会 教育工学研究会, ET2001-35, Sep, (2001)