

## HMM に基づく理解状況推定による学習支援の効率化

興治 崇<sup>†</sup> 新井 浩志<sup>†</sup>千葉工業大学 工学研究科<sup>†</sup>

## 1. はじめに

本報告では、学習支援システムにおける学習者とシステムの対話を時系列として捉え、確率過程に基づいて理解状況を表現する方法を提案する。確率過程は隠れマルコフモデル (Hidden Markov Model, HMM) を用いて表現する。あらかじめ理解状況が分かっている学習者から作成した HMM と学習者の動向から、Viterbi アルゴリズムを用いて理解状況を推定する。この推定を基に学習者が次に学習すると良い知識を提示する事で、効率的に学習支援を行う事が出来る。

学習支援システムの一形態として、学習者が主体となって自らの発見を促すシステムが提案されている。しかし、システムが学習者の理解状況を把握しづらく、学習が効率的に行われる保証がない。この問題点を解決する為には、学習者の理解状況に応じて膨大な教育リソースの中から学習者が学習すべき内容を抽出して提示する仕組みが必要である。

理解状況を推定しながら学習支援を行うシステムに関しては、いくつかの研究が報告されている。津森らの研究<sup>[1]</sup>では、学習領域の語彙に対して個別に理解度を設定し、多肢選択問題の解答状況により学習者の理解度を判定している。この研究では学習者の理解状況に対応した知識提供を行っており、正誤が一意に定まる問題や、一般的な解説が確立されている問題に対しては有効である。しかし、正解が複数存在する問題や理解への道筋が明確ではない問題への適用は困難である。本提案手法では、学習者との対話の系列をリアルタイムに観測し学習モデルと対比する事で、正解が一意に定まらない問題に対しても理解状況を推定する事が出来る。

## 2. 提案手法

本節では、着目する学習者の反応、反応からの理解状況の推定、推定結果に基づく知識の提示の3つの段階について提案手法を概説する。

本研究では、特定の学習分野に必要な知識の集合をトピックと呼ぶ単位に切り分け、学習者がトピックを読み進む順序を反応として捉える。

これは、教師と学習者の会話のやり取りの代替とみなす事が出来る。1つのトピックは、質問文と回答文の対からなり、固有の ID を持つ。特定領域に関する知識は全トピックの集合である。トピックでよく使われる単語をキーワードと呼ぶ。図 1 における Q は質問文、A は回答文を表し、それぞれ少なくとも 1 つ以上のキーワードを含む。

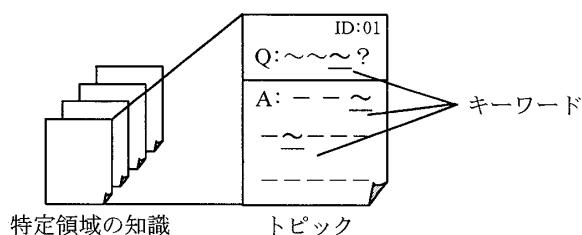


図 1 トピックとキーワードの関係

次に、上記の手法で取得した反応、すなわち学習者がトピックを読み進んだ順序から理解状況を推定する。その分野の概要だけ知っているのか、詳細まで理解しているのかなどの習熟度は、学習者が参照する知識の順序に表れると仮定する。例えば、学習者が図 2 に示す順序でトピックを選択した場合を考える。

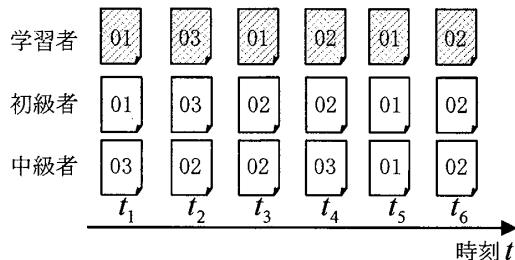


図 2 理解状況の推定モデル

学習者が選択したトピックと、初級者および中級者の選択トピックを比較すると、時刻  $t_5$  および  $t_6$  ではどちらも学習者と同じトピックを選択しているが、時刻  $t_1$  から  $t_4$  までに選択したトピックは初級者に近い。よって、この学習者は初級者に近いと推定出来る。この様に学習者の選択順序と、過去の学習者の学習過程から理解状況を推定する事が出来る。

最後に、推定に基づいてトピックを提示する。学習者は学習したいキーワードについてシステムに問い合わせをする。システムはこれに対し、

A learning support system using HMM based user model.

<sup>†</sup> Takashi Okiharu, Hiroshi Arai

Graduate School of Engineering, Chiba Institute of Technology.

そのキーワードを質問文に含む全てのトピックをピックアップする。そして、過去の学習者が次に選択したトピックを、理解状況の推定結果に基づいて確率的に求め、これを優先度とする。この優先度が高い順に一定の個数を学習者に提示する。学習者は、トピックの選択とキーワードの選択を交互に繰り返す事により学習を行う。この一連の流れを図3に示す。

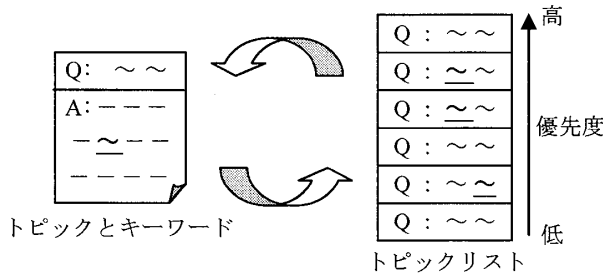


図3 トピックの選択

### 3. HMMの適用

HMMは次の5つのパラメータによって定義される。このモデルと、観測記号列  $O = \{o_1, \dots, o_T\}$  を用いる事で最尤な状態遷移系列の推定が出来る。

(1)  $S = \{s_1, \dots, s_N\}$ : 状態の有限集合

(2)  $R = \{o_1, \dots, o_M\}$ : 出力シンボルの有限集合

(3)  $T = \{a_{ij}\}$ : 状態遷移確率分布

$a_{ij}$  は状態  $s_i$  から状態  $s_j$  への遷移確率

(4)  $E = \{b_i(o_j)\}$ : シンボル出力確率分布

$b_i(o_j)$  は状態  $s_i$  でシンボル  $o_j$  を出力する確率

(5)  $\pi = \{\pi_i\}$ : 初期状態確率分布

$\pi_i$  は状態  $s_i$  が初期状態である確率

本研究では、状態は学習者の理解状況を表し、出力シンボルは学習者が選択する時系列のトピックを表すものとする。状態遷移確率分布  $T$  はある状態から次の状態へ遷移する確率であるため、学習者がある理解状況から、他の理解状況へ推移する確率分布である。シンボル出力確率分布  $E$  は、ある理解状況において学習者がどのようなトピックを選択するかを表す確率分布である。初期状態確率分布  $\pi$  は、最初に判定される状態の確率分布である。図4に示す例では、このモデルに対してトピック 02 の次にトピック 02 が観測された場合、その状態遷移は  $s_3$  から  $s_2$  と一意に定まる。

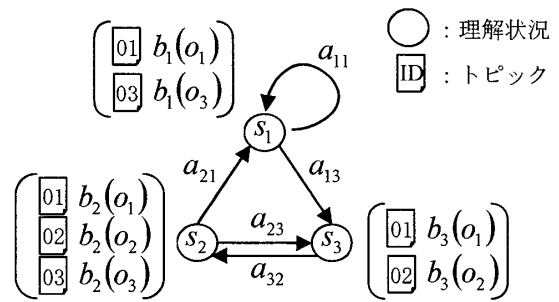


図4 HMMの適用例

推定は選択されたトピックの ID の順序を観測記号とし、あらかじめ用意されたモデルに対して Viterbi アルゴリズムで最尤経路を計算する事で行なう。最尤経路から、学習者の現時刻  $t$  における内部状態  $s_t$  を推定出来る。現時刻における状態  $s_t$  が推定されると、次時刻  $t+1$  における各出力シンボル  $o_j (j=1 \dots M)$  に対するシンボル確率  $P(o_j)$  は次式で求める事が出来る。

$$P(o_j) = \sum_{i=1}^N a_{s_i, t} \times b_i(o_j)$$

全ての出力シンボルに対して算出された値がトピックリストに付加する優先度である。

### 4. まとめと今後の課題

本稿では、HMMに基づく理解状況の推定と、推定結果から学習支援を効率化する手法について提案した。本提案に基づいて教育システムを実装し、就職活動に関する FAQ を教材として評価を進めている。図5の左から順に起動時のウィンドウ、トピックリスト、トピックである。

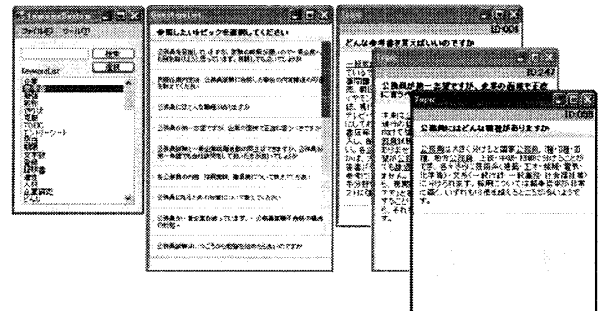


図5 提案手法の実装

本手法の有効性を評価する為には、あらかじめ理解状況の分かっている学習者から HMM を生成し、HMM に基づく推定が正しいか評価する必要がある。

### 参考文献

[1] 津森, 海尻: “理解状況に適応した多肢選択式問題の自動生成に関する構想”, 教育システム情報学会研究報告, Vol. 21, No. 4, 2006.