

# ベイズ推定を用いた適応型テストに関する研究

西條 緑

## 概要

本研究では、主に教育現場などで利用されるベイズを用いた適応型テストシステムにおける汎用性向上のための検討を行った。学習構造をネットワーク構造により表現し、問題選択および学習者の理解度推定に関する確率分布の設定法について理解状態と問題の関係をもとに問題の重要度を評価する新たな手法を導入、テスト終了条件を問題数に加え閾値の組み合わせによって設定する、としたところ学習者の理解度の推定精度の向上が確認できた。また、問題構造に異なる複雑性を持たせた場合にも同様の結果を得た。

## 1.はじめに

学校などの教育現場において、学習者である生徒に対して一人ひとりの学習の習熟度に合わせた指導をすることが理想とされている。しかし、教師による生徒に対する理解度の推定の多くは教師の経験則からされており、生徒一人ひとりに合わせた指導を行うことは現実的ではない。そこで、適応型テストによる学習者の理解状態推定法が注目されている。適応型テストは、学習者の理解度に基づいて個別に出題される問題が選択されるため、学習者一人一人に最適な学習を提供することができるとされている。学習者はシステムから出題された問題に対して回答を行う。システムはその回答をもとに学習者の理解度を推定し、次に適切な問題を出題する。このような出題と回答のサイクルが繰り返される。このプロセスにおいて、学習者の理解度をいかに正確に推定し、次の出題問題を適切に選択するかが、適応型テストの効果を大きく左右する重要な要素となる。これまで適応型テストには様々な手法の理解度推定法や問題選択アルゴリズムが提案・検討されてきた。秋山の[4]ではコンピュータ適応型テストは、初期条件設定、項目選択、出題、採点、能力推定、終了判定の6ステップで構成されているとしている。分寺の[3]では初期値の決定法は最終的な特性値の推定に大きな影響を与えないことを示した。加藤ら[1]の提案では、学習領域の持つ問題構造をネットワークとして表現した。学習者の理解度はテストの回答からベイズ的に推定している。また、出題問題の選択には、学習者の回答後の事後分布のエントロピーを最小化するように、相互情報量を評価関数とするアプローチを行っている。これらの計算法をもとに、学習者への出題を自動的選択し、「出題→(学習者からの)回答→理解度の推定」を繰り返すことによって学習者の理解度を診断するシステムの構築法が提案されている。大森ら[2]は、学習者が回答するうえで起こりうる間違い方(バグルール)をツリー構想で表現し、バグルールを持つ確率をベイズ的に推定した。[2]の手法では、学習構造に合わせて間違い方を全て検討する必要があり手動操作の負担が大きい。これらの先行

研究をもとにより汎用的な適応型テストの実装を目指す。様々な問題構造に対応するため、学習領域をネットワーク構造とした際に応用可能な[1]の手法を応用する。[1]ではモデルとして使用されている学習領域の問題構造が1種類のみであり技術の汎用性の検証がなされていない。また、システム内で用いられている確率分布の設定法について、先行研究では明記が見られないため、その決定法についても検討を行いたい。確率分布が自明でない構造に対しても機械的に設定可能な手法を提案し、学習領域の変化による、診断結果を比較しシステムの汎用性を検証する。また、推定をより精度よく行うための条件を実験により検証し提案を行う。

## 2. テスト構造

学習者の理解状態をもとにした適切な出題と、回答をもとにした理解状態のベイズ的な推定を繰り返す適応型テストを実装する。具体的には、「問題選択→出題→理解状態を推定し更新」といったサイクルを繰り返すことによって、学習者の理解状況の同定を目指した。テストが行われる空間を以下のように設定した。類似の解法を持つ問題項目の集合を  $T = \{t_1 \dots t_q\}$ 、問題に合わせて適当に設定した学習者の理解状態の集合を  $S = \{s_1 \dots s_n\}$  とし、学習者の理解状態  $s$  を  $s = \{x | x \text{ は学習者が理解している問題クラス}\}$  として定義する。学習者がどの理解状態にあるかを確率分布  $P(s|t)$  と設定した。出題から得られる結果は反応カテゴリーの集合として  $R = \{r_1, r_2, \dots\}$  とする。理解状態  $S$  と問題  $T$  の関係は、扱う学習領域によって変化し

「理解状態  $s_i$  のとき問題  $t_j$  を理解している」というように定義する。例えば、図1 学習構造 A のような問題の構造を持つ学習領域を扱う場合を考える。 $t_1$  から  $t_5$  の5つの難易度の問題が存在するこの学習領域では、 $t_1, t_3$  が前提知識を必要としない問題であり最も難易度が低いものといえる。反対に、 $t_5$  を習得するためにはこの問題領域における  $t_5$  以外の全ての問題  $t_1, t_2, t_3, t_4$  が前提知識として必要となる。これらの5問に対して、前提知識等つながりを考慮すると表1のような  $s_0$  から  $s_9$  の10個の状態が設定される。 $s_0$  は全ての問題を習得している、この学習領域を完全理解している状態である。問題が5問の場合、全ての組み合わせをとると32通りとなる。しかし、今システムでは問題の構造を図1のようなツリー構造をとることによって依存関係を明確にし状態数を削減している。例えば、問題  $t_2$  を理解している状態があるとする。問題  $t_2$  は前提知識として問題  $t_1$  を必要とするため、今回設定した状態  $s_4$  に収束されると考える。このような方式によって状態数の削減を行い、状態数32から10へと収束させ、起こりうる全ての問題の組み合わせそれぞれに理解状態が設定されている。秋山の[4]によって提案された6つのステップの具体的な設定法を定義していく。

6ステップのうち、出題、採点は実装時に自明的に設定できるものとする。

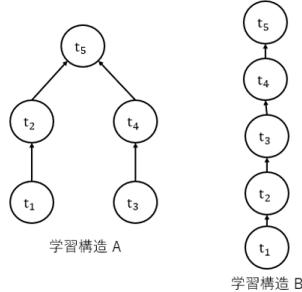


Figure 1とある学習領域における問題の構造図

### ● 項目選択

問題選択法は、[1]によって提案されている相互情報量を評価関数とする方式を用いた。学習者の理解状態の候補を絞り込むために問題を選択したい。したがって、事後分布のエントロピーを最小化する問題を選択する以下の式を用いる。

$$\max_y \sum_{r_k^y \in R^y} \sum_{s_i \in S} P(s_i) P(r_k^y | s_i, t_y) \log \frac{P(r_k^y | s_i, t_y)}{\sum_{s_i \in S} P(s_i) P(r_k^y | s_i, t_y)} \quad (1)$$

によって最大値をとるときの  $t_y$  が、相互情報量が最も大きくなる問題であり次の出題として選択される。

### ● 能力判定

学習者からの回答を得たのち、回答をもとに学習者の現在の理解状況を推定する必要がある。問題  $t_j$  を学習者に出題、その回答として  $r_k^j$ を得たとする。その時の事後確率  $P(s_i | r_k^j, t_j)$  は 1999 年の加藤ら[1]によって提案されている以下のベイズ推定を利用した式によって求める。

$$P(s_i | r_k^j, t_j) = \frac{P(r_k^j | s_i, t_j) P(t_j | s_i) P(s_i)}{\sum_{s_i \in S} P(r_k^j | s_i, t_j) P(t_j | s_i) P(s_i)} \quad (2)$$

以上の式によって推測される確率分布を事後確率とする。学習者から回答を得た際、 $S$  のすべての要素に対して計算した事後確率を、事前確率  $P(S)$  と置き換えることで学習者が持つと推定される確率分布  $P(S)$  を更新する。

## 4.まとめ

本研究では、先行研究で提案された適応型テストのアルゴリズムの実装を行った。先行研究では 1 種類の問題構造においてのみでしか、実験がされておらずその汎用性に疑問が残る結果であった。そのため、問題同士の関係として枝分かれの構造を持つ学習構造 A、枝分かれのない学習構造 B を用意した。その結果、学習構造 A では 89.2%，学習構造 B では 91%と高い確率で適切な診断を得ることができた。また、実装において必要とされる変数の決定方法を提案し、それらの方法に基づいた精度比較を行った。学習者がある理解状態  $s_i$ を持つとき、問題  $t_j$  が出題される確率である  $P(rt|si)$  の決定法について提案し比較実験を行った。具体的な重みづけの具体的な方法として、学習領域において以下の 3 点を適用させた。

1. 自身より上位に当たるノードの数
2. 状態  $s_i$  において理解している問題に -1, 理解していない問題に +1

3. 状態  $s_i$  の時理解している問題のなかで、最も難易度の高い問題 +1.5

4. 状態  $s_i$  の時理解している問題のなかで、最も難易度の高い問題の親ノード見当たる問題 +3

こうして得られたそれぞれの重みを状態ごとに確率化し適用することで、適切な診断の結果を得ることができた。また、上記の設定法では重みが 0 となる場合が発生する可能性が考えられる。その際には、重み全体に微量の値を付与することによって確率に 0 が発生することを防ぐ必要がある。先行研究によって提案されたアルゴリズムや、上記の確率分布決定法が、診断における精度向上に寄与することが確認され、理論的な予測が実際の結果と一致することが示された。今回の重みづけの方法では意図的な部分が大きいという懸念もある。しかし、上記に示した重み 1 における確率への反映方法を変更したところ、適切な推定結果を得ることができなかった。このことから、単に状態  $s_i$  と問題  $t_j$  の間に意図的な関係を設定し、 $P(rt|si)$  という確率を適用すればよいというわけではないことがわかった。実際には、状態と問題の関係性だけでなく、全体の確率分布における整合性や調整が重要であり、そのバランスを取ることが診断精度の向上につながるのではないかと考える。

## 文 献

- [1] 加藤浩・赤堀侃司、ベイズ推定による適応的問題 演習システムのための問題選択方式、電子情報通信学会論文誌, 82(1)(1999), 147-157.
- [2] 大森拓哉・繁糾算男、ツリーモデルを用いたバグ ルールの適応型ベイズ診断、日本教育工学会論文誌, 21(2)(2007), 211-218.
- [3] 分寺杏介、制限時間のある適応型テストにおける項目選択アルゴリズムの比較検討、日本テスト学会誌, 15(1)(2019), 1-20.
- [4] 秋山寛、潜在ランク理論に基づくコンピュータ適応型テストのアルゴリズム、日本テスト学会誌, 10(1)(2014), 81-94.