

スパース符号化を用いた試験結果からスキル構造の抽出

菊池 祥平

近年, Educational Data Mining(EDM)と呼ばれる教育分野におけるデータ解析に注目が集まっている. この分野では生徒の試験結果や学習の記録を解析し, 有益な情報を獲得することを目的としている. その中に, 学生の試験結果を用いて学生が持つスキル, 設問そのものを解くために必要となるスキルを抽出する方法が研究されている.

試験は学生のスキルを測るために用いられる. 設問に対して正解ならば 1, 不正解であれば 0 とすることで学生と設問から形成される試験結果を行列として表現することができる. 本来, 設問には設問を解くために必要なスキルが決定されている. 設問とスキルの関係を 0,1 の行列で表したものを Q-matrix という. 各設問におけるスキルの関係は有識者によって定義されており, 多大な時間と費用を要する. そのため, EDM の分野では大量の試験データから Q-matrix を自動生成しようとする試みが行われている. これまでの研究では Q-matrix を如何に正確に得るかという観点で研究がなされている. しかし, これまでの研究ではスキルの表現性に着目しておらず, 0,1 で表現されたデータにしか適用されていない. 本来ならば試験の設問によってスキルの重要性が変化するはずである. また, 既存の手法では Q-matrix は表現できるものの, 一つの設問が無数のスキルの組み合わせで表されることがあるため意味解釈が困難である.

本研究では「スキルは論理値ではなく負値などの実数値を含む値で表現されるはず」という仮説のもと, スキル構造を論理値から実数空間に拡張し, 抽出されるスキル構造をスパース符号化という手法を用いて重要な特徴のみで表現するというを試みる. 0 となる成分が多いことをスパース性と呼ぶ. スパース符号化とは, ある行列 A とベクトル y が与えられた時にもっともスパース性を課したベクトル x を求めることである. このスパース符号化を試験結果に応用した, 行列因子分解と呼ばれる行列の分解を行うことで実数空間への拡張とスパース性を考慮した Q-matrix を生成する実験を行った. その結果, 試験結果の行列と行列因子分解によって求めた行列との誤差を測ることでどのスキル数で分解した行列が適切なスキル数となるのかを機械的に判定をすることができた. また, 実際の行列の誤差の他にも生成された行列との誤差やスキル行列の類似性を測ることで適切なスキル数, が設定できるかを追実験し, それぞれの手法で求めたスキル行列の比較を行なった.

今後はスキル数とスパース性を情報理論などのモデルで評価できないかという観点での追実験を行うことで, 行列因子分解でのスキル行列最適化を行う.

(指導教員 手塚太郎)