DEIM Forum 2017 B8-4

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†　　手塚　太郎†

† 筑波大学 情報学群 知識情報・図書館学類　〒305–8550 茨城県つくば市春日 1–2
E-mail: †s1513119@u.tsukuba.ac.jp, ††tezuka@slis.tsukuba.ac.jp

あらまし　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．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．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．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ことを試みる．

キーワード　教育データマイニング，スパース符号化，辞書学習

1. は じ め に

1.1 背　　景

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 [1]．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．

1.2 概　　要

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ことを試みる． 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まっ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まったスキル行列の比較を行なった．

1.3 本論文の構成

本論文は以下の通りである第 2 章では先行研究での手法を説明し本論文との関連性について説明する．3 章では本研究で用いる手法，KSVD と評価方法について説明する．4 章では実際に学生が解いた試験結果を用いて実験を行い，結果と考察を示す．5 章ではこれらのまとめを行う．

2. Q-matrix と試験結果から Q-matrix の抽出

2.1 Q-matrix の導入

教育機関や一般社会では，人それぞれの能力を客観的に評価する必要がある．教育機関においては，各種学校での入学試験として，受験者に対して教育を受けるために必要な知識や技能を有しているか確認するための試験を行う．一般社会においても，各種資格などに対して，知識や技能を有しているかを確認するために試験を行うことが多い．この様に評価したい知識や技能をスキルと定義すると，試験は学習者のスキルを測定するために用いられていると考えられる．試験には形態として口頭試問や実技試験，筆記試験など様々な方法が存在する．本論文において試験とは筆記試験を指し示すものとする．

一般的にスキルとは抽象的な概念であり，直接的にスキルの有無を測ることは難しい．そこで，一般的には試験に設問を与えることで試験を実施している．したがって，設問に正答するために必要なスキルが潜在的に設定されていると考えることができる．本来測定したかった学習者のスキルの有無は，これらの試験の正答数から判断されていると考えられる．

しかし，解答には当てずっぽうやケアレスミスなどが存在し，本来正解していたはずなのに不正解となるケースや不正解であ

ることが正しいのに正解しているケースが存在する．これらの事象から考えられることは，学習者の表層的な得点のみではスキルの状態を正しく観測できないということである．そこで，潜在的なスキルの状態を把握し，学習者の実力を測るために教育心理学の分野では Q-matrix という概念が用いられている [2]．Q-matrix とは，各設問に正解するためにどのスキルが必要とされているかを表すために，行に設問番号，列にはスキル番号を配した関係行列として表現するものである．この Q-matrix は有識者によって定義される．Q-matrix が用意された試験では，試験の結果から学習者のスキル状態を評価することが可能である [3]．図 1 に Q-matrix の外形を示す．

		Skils	
		s1	s2
Items	i1	1	0
	i2	0	1
	i3	1	1

図 1　Q-matrix の外形

例えば，この Q-matrix では，items が設問を示し，Skils がスキルの有無を示している．i1 では s1 が 1，s2 が 0 となっている．これは設問番号 1 では，この設問に正答するためにはスキル番号 1 が必要であり，スキル番号 2 は不要ということを示している．同様に考えると設問 2 ではスキル番号 2 が必要，設問 3 ではスキル番号 1 と 2 の両方が正答するにあたって必要ということを示している．

この Q-matrix の問題点として，試験問題を作成するだけでも大変な作業であることに加えて，その試験問題から対応するスキルを列挙し Q-matrix を作成するという作業は多大な労力を伴うため教育心理学の分野でも敬遠されてきた．そこで，Educational Data Mining(EDM) [1] の分野では少ない労力で Q-matrix を生成するために，Intelligent tutoring systems(ITS) や Leaning management systems(LMS) によって大量に蓄積されている試験結果から Q-matrix を自動生成しようとする試みがなされている [4]?[8]．これらは試験結果を学習者と設問のスコアを配した行列と解釈することで行列因子分解の問題として Q-matrix を求めている．試験結果は正答していれば 1 を，誤答であれば 0 を与える．これを設問とスキルの関係行列である Q-matrix と学習者とスキルの関係行列として因子分解する．

式 (1) では 試験結果 R から 正しく Q-matrix と学習者の関係行列が抽出できた場合の行列を示している．試験結果を設問 n，学習者 m の行列 $R \in \{1,0\}^{n \times m}$ として考え，正答を 1，誤答を 0 として表現する．また，有識者が定義する設問とスキルの関係である Q-matrix を 設問 n，スキル k の行列 $Q \in \{1,0\}^{n \times k}$，スキルと学習者の関係はスキル k，学習者 m の行列 $S \in \{1,0\}^{k \times m}$ として表現する．Q, S では，それぞれスキルを有していた場合に 1，そうでなければ 0 が与えられている．S は学習者がどのスキルを有しているかを表していることになる．

$$R = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, S = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \tag{1}$$

式 (1) のように，Q, S が与えられたならば，理想的な試験結果の R が与えられると仮定して，逆に試験結果の R から Q, S を抽出することを考える．R と Q, S の関係は，

$$R \fallingdotseq Q \times S \tag{2}$$

と表すことができる．ここで記号 $\times$ を一般的な行列演算の積と定義する．

2.2　NMF と BMF による試験結果への行列因子分解の適用

先行研究 [6] では Non-negative Matrix Factorization (NMF) によって人工的に生成した試験結果から Q-matrix を抽出可能であると定量的に示している．さらに，先行研究 [7], [8] では，試験結果の関係行列は 1,0 の論理値行列であることに着目した上で Boolean Matrix Factorization (BMF) を用いて行列因子分解を行っている．特に先行研究 [7] では，NMF と BMF による Q-matrix の抽出結果を比較する計算機実験を行った結果，BMF は初期値依存することなく NMF よりも安定して Q-matrix を抽出できることが示されている．さらに先行研究 [8] では BMF に対して Minimum Description Length (MDL) を適用する [9] ことで，機械的に最適なスキル数を決定する方法について論じている．この論文で行なっている実データを用いた実験では，人間が決定した Q-matrix よりもスキル数が少なく現れており，少ない場合でも解釈可能なことから人間の設定した Q-matrix が間違っている可能性，機械により求められた Q-matrix が間違っている可能性，どちらも適切であるという可能性，どちらも不適切であるという可能性の計 4 つについて考察を行なっている．

2.3　Non-negative Matrix Factorization(NMF)

NMF とは，行列を非負値の 2 つの行列に分解する手法である．与えられた行列を $W \in \mathbb{R}^{\geqq 0, n \times m}$，NMF により分解される行列を $U \in \mathbb{R}^{\geqq 0, n \times k}$，$V \in \mathbb{R}^{\geqq 0, k \times m}$ とおくと，式 (3) のように表すことができる．

$$W \approx U \times V^{\mathrm{T}} \tag{3}$$

また，NMF の概念図を図 2 に示す．

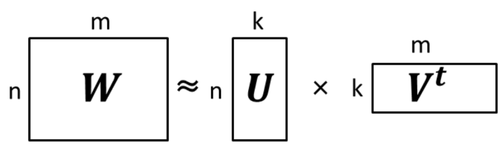

図 2　NMF の概念図

2.4 Boolean Matrix Factorization(BMF)

NMF は非負値の行列に分解するが，行列が 0,1 の論理値の場合は論理値に対応した手法を用いることが良いと考えられる．BMF は論理値演算を用いて行列因子分解を行う手法である．NMF が非負値の実数値に分解するのに対して，BMF では要素が論理値である行列に分解する．図 3 に示すように，$n \times m$ 行列 $R \in 0,1$ を，$n \times k$ 行列 $Q \in 0,1$ と $k \times m$ 行列 $S \in 0,1$ に分解する．なお，$\circ$ は行列の論理積和演算子を表す．

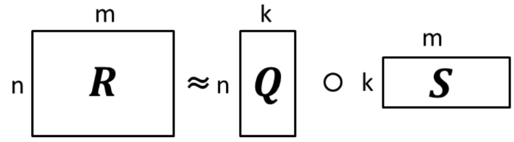

図 3　BMF の概念図

2.5 問　題　点

2.5.1 Q-matrix の問題点

Q-matrix は設問とスキルの関係行列として定義されている．この行列は論理値行列であり，スキルを有しているか否かという観点での判定はできるが，設問に対してあるスキルがどの程度影響しているかを測る術がない．これは言い換えると設問を解くために必要なスキルの優先順位が分からないという問題になる．設問に対してスキルの重要度合いを測るためには，各スキルが実数値で表されている必要がある．また，設問と学習者が解いた関係行列についても，本来は設問に対して点数が振り分けられており，作問者は本来設問それぞれに対して重要度を点数などの形で表現している．既存の方法では設問の重み付けがなされていない．

したがって，重み付けを行うためには設問と学習者が解いた関係行列に対しても論理値ではなく負値を含む実数空間に拡張した際に，正しく重み付けを行なった上での行列因子分解が行われれば有用であると考えられる．

2.5.2 NMF,BMF の問題点

NMF，BMF はその名前の通り，行列が非負値であるという制約のもと分解を行う手法，行列が論理値であるという制約のもと分解を行う手法である．負値を含む実数空間への拡張を行なう場合，NMF や BMF ではこの行列を正しく因数分解することができない．

2.5.3 意味解釈の問題点

既存の手法のもと，行列因子分解を行うことで Q-matrix と学習者とスキルの関係行列を表すことが可能である．しかし，実データに対してこの手法を行うと，表される Q-matrix の各設問は，無数のスキルの結合で表現されることがある．行列因子分解により求まった Q-matrix のスキルはラベル付けがされておらず，人間がそれぞれラベル付け作業を行う必要がある．その場合，無数のスキルの結合で表された設問はどのスキルを有していれば正答できるのかを解釈するのが困難になる．

したがって，意味解釈を行うためには用いるスキルの中でも重要な要素のみで表現されている方が適切であると考える．そのために，意味解釈を容易にするためには設問に用いられるスキル数の限度を動的に決定する必要がある．

3. 研 究 手 法

本研究では 2.5 で述べたように Q-matrix 本来の問題点と NMF/BMF による行列因子分解での問題点を挙げた．これらの問題に対処するためにスパース符号化を用いた辞書学習の手法を導入する．

3.1 スパース符号化

人間の脳は情報のスパース化が行われていると考えられている．情報のスパース化とは，ある概念を表現する時により疎化された表現でその概念を表現するということである．視覚を例にすると，人間は画像を見た時に画像を基底画像の線形和として処理することによって認識している．スパース化が可能なものは画像や音声情報，信号などのベクトルとして表現されるものが挙げられる．分解したい観測データを n 次元のベクトル y とすると，$y = Dx$ で表現される．D は $n \times k$ 次元行列で辞書と呼ばれ，観測データを表現するための k 個の基底ベクトルから成る．x は $n \times k$ の係数ベクトルである．観測データが与えられた時，適切な辞書 D を求めた上でスパースなベクトル x を求めるのがスパース符号化と呼ばれる．辞書というのは，観測データを表現するために，基底ベクトルからなる行列を辞書と呼んでいる．この辞書から任意の基底ベクトルを選択することであらゆる行列やベクトルを表現することは可能である．辞書を冗長にすることで行列，ベクトルをほとんどの要素が 0 となるベクトル，つまりスパースなベクトルで表現することが可能である．

このスパース符号化には，観測データが論理値や非負値である必要がなく，任意のベクトルの組み合わせで表現されたデータを与えられる．つまり観測データの制約が無く，与えられるデータは負値を含む実数で表現されたベクトルの組み合わせで良い．従って Q-matrix が任意の要素を含んでいても良いことになる．

また，NMF/BMF では非負値，論理値という制約があったが，スパース符号化を行うことで行列因子分解中の片方の行列を辞書とすることでスパースな行列を求めることができる．これは教育データに当てはめると，設問と学習者のスコア行列をスパース符号化することで，従来使われていなかった学習者とスキルの関係行列を辞書とし Q-matrix はスパースな行列として表現するという命題になる．この時，従来の行列因子分解と異なる点として，学習者とスキルの関係行列は任意の要素を含む辞書になるという点である．辞書自体はスパースではないため，スパースな行列として生成されるのは Q-matrix のみとなる．また，Q-matrix においても 0,1 の論理値行列では無く，負値を含む実数のスパースな行列として生成される．したがって，従来人間が考案していた Q-matrix と比較した際に差が生まれることになる．そこで本研究では辞書を学習させることで求められた Q-matrix と従来の Q-matrix の比較も行うこととする．

3.2 スパース符号化のための辞書学習

3.2.1 K-means

観測データからスパースなベクトルを求めるためには適切な辞書を学習する必要がある．そこで，本研究ではスパース符号化のための辞書学習の手法として K-SVD という手法を導入する [10]．K-SVD とは K-means の一般化でもある．K-means とはクラスタリング手法の 1 つであり，初めはランダムにデータをクラスタに割り当てる．そのクラスタ毎の重心を求め，各データについて，一番重心が近いクラスタへとデータを再度割り当てる．そしてまたクラスタ毎に重心を求め，一番近い重心のクラスタへとデータを再割り当てするということを繰り返しクラスタリングを行う．クラスタの数を K 個とし，重心の計算には算術平均が用いられることから K-means と名付けられている．

K-means は非ゼロ要素を 1 つに限定し，かつその値を 1 つに限定したスパース符号化とも取れる．N 個のデータをベクトル集合 $Y = \{y_{:1}\}_{i=1}^{N}(N \gg K)$ として考え，各クラスタの重心を表すベクトルを横に並べて行列 $C = [c_{:1}, c_{:2}, \cdots, c_{:K}]$ と表す．ここでデータの数 N はクラスタの数 K より十分多くする必要がある．

各データ $y_{:i}$は C の中にある列 $c_{:j}$，つまりどこかのクラスタに割り当てられる．これを係数ベクトル $x_{:i}$を用いて $y_{:i} = Cx_{:i}$ と表すと，$x_{:i}$は割り当てられている C の列 $c_{:j}$に対応した j 成分 x_{ji}のみが 1 で，残りは全て 0 になる．$y_{:i}$に対応する $c_{:j}$はL2 ノルムを用いて

$$\forall k \neq j. ||y_{:i} - Cx_{ji}||_2^2 \leqq ||y_{:i} - Cx_{ki}||_2^2 \qquad (4)$$

を満たすものが選ばれる．これによって $x_{:i}$を更新する．次に $X = \{x_{:i}\}$ を固定して X を更新する．$y_{:i}$における二乗誤差を $e_i^2 = ||y_{:i} - Cx_{:i}||_2^2$と定義し，$Y$ 全体での二乗誤差を

$$\sum_{i=1}^{N} e_i^2 = ||Y - CX||_F^2 \qquad (5)$$

と定義する．$||A||_F$ はフロベニウスノルムを表し，$||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$ である．二乗誤差を最小化するように C を更新する．

以上のように K-means は 1) クラスタ C に基づいて X を決定する段階と， 2) X を固定して C を書き換える段階の 2 つの工程を交互に繰り返すことでデータをクラスタリングしている．まとめると，

$$\min_{C,X}\{||Y - CX||_F^2\} \text{s.t.} \forall i ||x_{:i}||_0 \leqq 1 \qquad (6)$$

とかける．$||x_{:i}||_0$ は 0 ノルムと呼ばれ，非ゼロ成分の数を表す．$x_{:i}$ の非ゼロ成分が 1 つという条件で E を最小化するような C と X を求めることを表している．

3.2.2 K-SVD

K-SVD では，式 (6) において値が 0 または 1 であった非ゼロ成分数を閾値以下の任意の値にできる．したがって K-means の一般化と言われている．K-SVD は式 (6) と対応させて考えると．

$$\min_{D,X}\{||Y - DX||_F^2\} \text{s.t.} \forall i. ||x_{:i}||_0 \leqq T_0 \qquad (7)$$

とかける．ここで D は辞書を表す $n \times k$ 行列であり，T_0は係数ベクトル $x_{:i}$における非ゼロ成分の上限を表す．D の各列は基底を表し，一般にアトムと呼ばれる．データ $y_{:i}$は基底の線形結合で表されるが，使われる基底の数は T_0によって制限される．

K-SVD も K-means と同様に 2 つの工程で構成されている．1) 辞書 D を固定して係数行列 X を求めるスパース符号化の段階と，2) X に基づいてより良い D へと書き換える辞書の更新の段階である．2) の段階で，D とともに関係する X の値 の変更も許すというのが他の K-means を一般化した手法に見られない K-SVD の特徴であり，それゆえ列ごとに独立して働く K-means の一般化に近いと言われている．

K-SVD のアルゴリズムを 2 つの段階に分けて考える．まず 1 つ目のスパース符号化の段階だが，行列のままでは扱いにくいので， $||Y - DX||_F^2 = \sum_{i=1}^{N} ||y_{:i} - Dx_{:i}||_2^2$ と置き換える．すると式 (7) は，

$$\min_{x_{:i}}\{||y_{:i} - Dx_{:i}||_2^2\} \text{s.t.} ||x_{:i}||_0 \leqq T_0 (i = 1, 2, \cdots, N) \qquad (8)$$

という N 個の独立した問題として考えることができる．T_0は任意の数に設定できるが，$x_{:i}$の次元数と比べて小さいので，$x_{:i}$は多くの成分が 0 となりスパースなベクトルとなる．よって，ここではスパース符号化を用いて式 (8) を満たす $x_{:i}$を求める．辞書 D は変化させず，一定とした上で係数ベクトル $x_{:i}$の非ゼロ成分が T_0以下となるまでスパース符号化を行う．次に辞書 D を更新する．こちらも 1 段階目と同様にして列ごとに動かしたいので，D の k 列目のみを取り出せるように工夫する．その際，辞書 D の各列に対応した X の行も同時に考える．

$$\begin{aligned} &\|Y - DX\|_F \qquad (9) \\ &= \|Y - \sum_{j=1}^{k} d_{:j} x_{j:}\|_F \\ &= \left\| \left(Y - \sum_{j \neq k} d_{:j} x_{j:} \right) - d_{:k} x_{k:} \right\|_F \\ &= \|E_k - d_{:k} x_{k:}\|_F \end{aligned}$$

式 (10) では D の k 列目とそれに対応する X の k 行目のみを取り出した残りが E_kとなっている．これは k 列目が取り除かれた時の N 個のデータの誤差を表している．2 段階目では特異値分解 (SVD) を用いて E_kとの差を最小化する $d_{:k}$及び $x_{k:}$ を求める．

しかし，SVD にはスパース性に関する制約がない．式 (10) を最小化するために，$x_{k:}$の 0 である成分にも非ゼロの値が入る．このままだと 1 段階目で制約をかけて生成したスパース性が失われてしまう．そこで SVD でも x のスパース性を保持する

ために，$x_{k:}$の非ゼロ要素のみを変化させるようにする．これは，$d_{:k}$が使われている $y_{:i}$を選ぶことに等しい．$x_{k:}$の非ゼロ成分の位置を指定するために，$\omega_k = 1 \leqq i \leqq N, x_{ki} \neq 0$ を定義する．さらに，ω_k成分が 1，それ以外が 0 となる $N \times |\omega_k|$ 行列としてΩ_kを定義する．$\tilde{x}_{k:} = x_{k:}\Omega_k$ とすると，$\tilde{x}_{k:}$は $x_{k:}$から非ゼロ成分を抜き出したベクトルとなり，列の数が N から $|\omega_k|$ へと変化する．同じように Ω_kを Y にかけたものを$\tilde{Y}_k$とすると，$\tilde{Y}_k$は $d_{:k}$が使用されている列のみを取り出した行列を表し，列の数も$\tilde{x}_{k:}$と同じく N から $|\omega_k|$ へと減らされる．同様のことが $\tilde{E}_k = E_k\Omega_k$でも起こる．$\tilde{E}_k$は $d_{:k}$を使用している $y_{:k}$に対応した誤差のみを E_kから抜き出したものを表す．これらを使って，式 (10) の右辺の代わりに以下の式を最小化する．

$$\|E_k\Omega_k - d_{:k}x_{k:}\Omega_k\|_F = \|E_k - d_{:k}x_{k:}\|_F \qquad (10)$$

$\tilde{E}_k$ に対して SVD を行うと，$\tilde{E}_k = U\Lambda V^T$と分解できる．$\Lambda$ は対角行列であり，対角成分はそれぞれ対応する U と V の列の重みを表していると言える．式 (10) を最小化することが目的なので，$d_{:k}$と$\tilde{x}_{k:}$は最も大きい特異値に対応する特異ベクトル，つまりΛの対角成分が最大のものに対応する U と V の列ベクトルから作れば良い．最大特異値をλ_{11}で表す．そして$\tilde{d}_{:k} = u_{:1}$，$\tilde{x}_{k:} = \lambda_{11}v_{:1}^T$と更新すれば，式 (10) を最小化することができる．ただし，$\tilde{d}_{:k}$は毎回正規化しなければならない．

以上の工程を繰り返して K-SVD はデータを辞書とスパースな行列に分解することができる．少ない情報量でデータを表現できるので画像の圧縮などにも用いられている．

4．研究手法の評価実験

4.1 実験方法

ここでは実際に実データを用いて辞書学習を用いたスパース符号化による Q-matrix の抽出を試みる．実験データは統計言語 R の CDM パッケージにある "fraction.subtraction.data" および"fraction.subtraction.qmatrix" を用いた [11]．このデータは分数の問題 20 問を 536 人が解いた結果である．この試験における Q-matrix は有識者によってスキル数 8 と定義されている．したがって，定義されている Q-matrix と推定した Q-matrix の比較を行う．図 4 に辞書学習によるスコア行列の行列因子分解結果を示す．

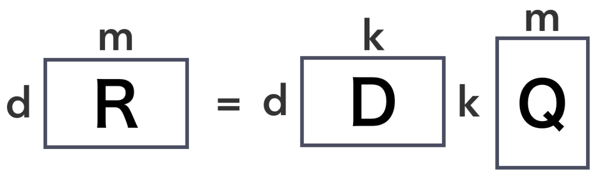

図 4 辞書学習によるスコア行列の分解

この図は生徒数 d，問題数 m が与えられた時に，$d \times k$ の辞書行列 D と $k \times m$ のスパースな Q-matrix に分解を行うということを示している．

この実験において辞書学習で設定するパラメータは 2 つである．1 つはスキル数のサイズを表す k のサイズである．k のサイズが大きくなる場合はスキル数が多いと言える．2 つ目は m の中での非ゼロ成分数 n である．もしこの非ゼロ成分数 n が小さい場合を考える．例えば $n = 1$ のとき，スキル数 k がどれだけ大きい場合でも設問 m の中で実数として現れるのは 1 つのみである．この n は設問中のスキルの複合度合いを示していると言い換えられ，$n = 1$ ならば 1 問につき 1 つのスキルで表現がなされ，$n = 2$ では 1 問につき 2 つのスキルの複合問題として表現することが可能である．したがって，n が小さい場合は少ない要素で表現を行うため制約が厳しく，n が大きい場合は複数の要素で表現が行えるため制約が緩いと解釈することができる．

4.2 制約付き辞書学習

スキル数の増加と非ゼロ成分数の値は従属関係にある．スキル数は増加すればするほど，行列内で表現できる要素が増えるため誤差値は下がると考えられる．また，非ゼロ成分においても，スキル数の増加に伴って値を増やすことができる．非ゼロ成分が多くなると行列としての制約が緩くなるため収束しやすくなり，誤差値が下がると考えられる．このように，スキル数と非ゼロ成分数は従属しているため，どちらがどの程度収束に影響を及ぼすのか判断することが難しいと言える．

そこで，ここではさらに制約付き辞書学習という手法を導入する．図 4 による辞書学習では，1) R が与えられた時，辞書学習によって生成された辞書 D は，任意の Q を生成するためのテンプレートである必要がある．また，2) 一般的な Q-matrix の場合は，生徒と設問のスコア行列 R から Q-matrix を生成する際に，生徒の人数が変化しても Q-matrix は変化してはならない．1) は一般的な辞書学習についての考え方であり， 2) は Q-matrix 本来の役割を考えた際に必要な事項である．そこで，得られた Q-matrix と辞書それぞれについて，最適さを測る手法を考案する．

図 5 に制約付き辞書学習の方法を示す．

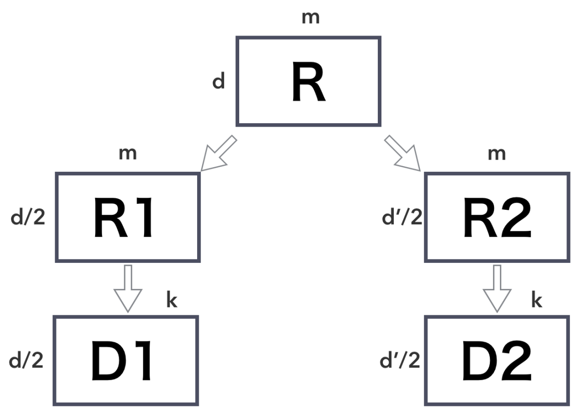

図 5 スコア行列 R の分割

図 5 では，まず与えられた R を 2 分割する．そして，分割された R と R' のそれぞれについて KSVD による辞書学習を行い，辞書 D と Q-matrix を求める．次に，それぞれの 辞書 D と Q-matrix を比較し，その誤差が最小となるような スキル数 k を探索する．辞書 D と Q-matrix の最適化にはフロベニウ

スノルムを用いる．最適化したい行列を X, X' とすると，これらのフロベニウスノルムは式 (11) によって求められる．

$$\min_{X,X'} ||X - X'||_F \tag{11}$$

4.3 評価方法

この実験では定量的評価を行う．辞書学習による学習の反復回数は 500 と固定し，制約条件となる n の値を元の Q-matrix での値を参考に 3 とした．

4.4 実験結果

4.4.1 素朴な実験

実験結果を図 6 に示す．

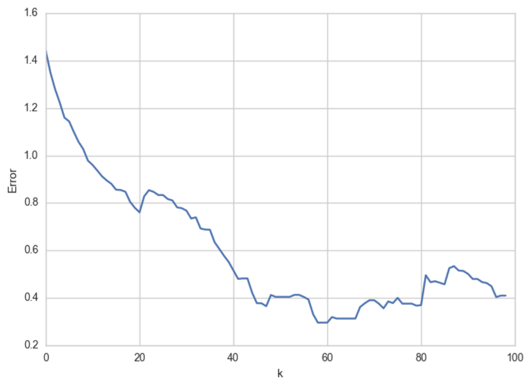

図 6　非ゼロ成分数 3 とした時のスキル増加によるスコア行列の誤差値の関係

この図では，非ゼロ成分数を固定した際に，スキル増加によってスコア行列の誤差値にどのような変化が見られるかを示したものである．この図では，スキル数が 60 前後の段階で誤差値が最も小さくなっており，スキル数を 100 程度まで調べた段階ではスキル数 60 前後が最適なスキルであることがわかる．

4.4.2 制約付き辞書学習による実験

次に，制約付き辞書学習による実験結果を図 7 に示す．

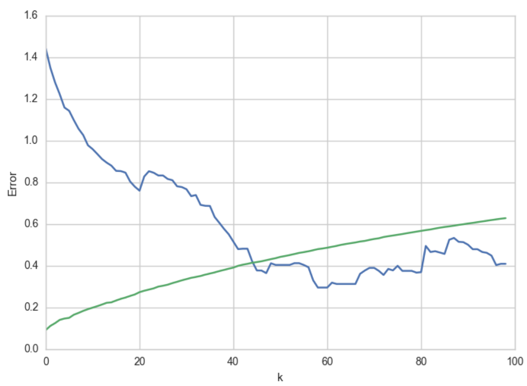

図 7　非ゼロ成分数 3 とした時のスコア行列の誤差値と辞書制約を施した辞書の誤差値

この図では，青線が図 6 において求められたものである．緑線は辞書間のフロベニウスノルムを求めたものである．この図によって比較を行うと，スキル数 40 - 60 の間で誤差値が逆転している．スコア行列の誤差に対して辞書の誤差は緩やかに増加し続けていることがわかる．各誤差の重み付けによって最適なスキル数を判定するのが難しいが，緑線の特徴から 40 - 60 以下のスキル数が最適であると考えることができる．

4.4.3 Q-matrix を制約した場合の実験

Q-matrix を制約した場合の実験結果を図 8 に示す．

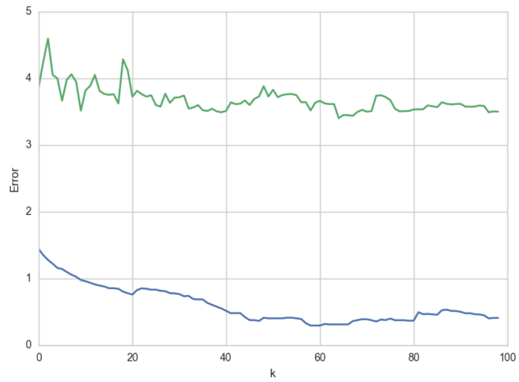

図 8　非ゼロ成分数 3 とした時のスコア行列の誤差値と Q-matrix を制約した場合の誤差値

この図では，青線が図 6 において求められたものである．緑線は Q-matrix のフロベニウスノルムを求めたものである．この図では，Q-matrix の誤差値がスコア行列の誤差値と比較して大きくなっていることがわかる．また，Q-matrix の誤差値はスキルの増加に伴って減少するという傾向は見られず，ほぼ横ばいであることがわかる．

4.5 考　　察

非ゼロ成分数を固定した場合にスキルの増加を調べた．その結果スコア行列はスキルの増加に伴って減少する傾向があることがわかった．また，非ゼロ成分数を固定した場合に最適なスキル数 k を見つけるという実験を行なった際に，制約付き辞書学習を行なったが，辞書については徐々に増加，Q-matrix においては最適な結果が見られなかった．これらについて考察を行う．

スコア行列の減少傾向が見られたのは，辞書 D の要素がスキル行列の増加と伴い相対的に表現性能が上がったためだと考えられる．辞書 Q はスキル数 k のサイズによって変化する．例えば，k が小さいということは，わずかな辞書の要素で Q-matrix を求めることと等しい．この際，辞書 D は Q-matrix を決定付けるための重要なアトムで構成される．逆に，k のサイズが大きい場合はより柔軟な表現で Q-matrix を求めることが可能になる．このことから，辞書のサイズが大きくなることによって元の行列を再現する能力が向上することが考えられる．再現性が上がるということは，基のスコア行列により近似した行列ができるということであり，結果としてスコア行列の誤差が減少傾向にあるのは自然といえる．

次に辞書 D について考える．先ほど k のサイズが上がること

で辞書が柔軟な表現をすることが可能になると説明した．これはQ-matrix を決定付けるアトムの他にも，Q-matrix に影響の少ないアトムが用いられることになるとも考えられる．Q-matrix に影響の少ないアトムが辞書に増えれば増えるほど，辞書そのものについて最小化を行なった際に一致度は下がることが考えられる．以上の理由から，辞書 D がスキル数の増加に伴って誤差値が上昇したと考える．

Q-matrix の一致度がスコア行列に比べて高いのは生成された辞書の影響によるものだと考えられる．辞書そのものはテンプレートなので要素こそ大きく変化しないものの，それを用いて生成する Q-matrix は大きく変化する．他 2 つの行列と比較した際にも誤差値は大きく，今後最適化すべき点はこの Q-matrix である．

5. まとめ

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ことを試みた．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まっ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まったスキル行列の比較を行なった．

課題として誤差値という観点のみでは一概に最適なスキル数であると断定することができないという問題がある．その理由として，誤差値はスキル数の大きさと非ゼロ成分数のどちらもが従属関係にあるためである．これはスキル数を固定していても非ゼロ成分数を増やした場合収束しやすくなるという点，非ゼロ成分数を固定してもスキル数の増加によって収束しやすくなるという側面を併せ持っているためである．

今回の研究では制約付き辞書学習という方法を提案し辞書行列，Q-matrix のどちらに対しても誤差が最小となるような k を求めるための追実験を行なった．その結果，辞書 Q においてはスキル数の増加により徐々に誤差値が増加していくという傾向が見られたものの，Q-matrix においては優位な傾向は見られなかった．今後の実験ではこれらの制約に加えて情報モデルの評価という観点から検証を行う必要がある．

文献

[1] Cristobal Romero, Sebastian Ventura, Mykola Pechenizkiy, Ryan S.J.d. Baker, “Handbook of Educational Data Mining (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)”, CRC Press, 2010.

[2] K.K. Tatsuoka, ”Rule space: An approach for dealing with misconceptions based on item response theory” Journal of Educational Measurement, Vol. 20, pp. 345-354, 1983.

[3] K.K. Tatsuoka, ”Cognitive Assessment: An Introduction to the Rule Space Method”, Routledge Academic, 2009.

[4] Barnes, Tiffany. ”The Q-matrix method: Mining student response data for knowledge.” American Association for Artificial Intelligence 2005 Educational Data Mining Workshop. 2005.

[5] Winters, Titus, et al. ”Topic extraction from item-level grades.” American Association for Artificial Intelligence 2005 Workshop on Educational Datamining, Pittsburgh, PA. Vol. 1. No. 2. 2005.

[6] Desmarais, Michel C. ”Mapping question items to skills with non-negative matrix factorization.” ACM SIGKDD Explorations Newsletter 13.2 (2012): 30-36.

[7] 大枝真一, 天野恵理子, and 山西健司. ”行列因子分解を用いた時系列試験結果からの潜在スキル構造の抽出 (ポスターセッション, 第 16 回情報論的学習理論ワークショップ).” 電子情報通信学会技術研究報告. IBISML, 情報論的学習理論と機械学習 113.286 (2013): 123-130.

[8] 清野真理子, and 大枝真一. ”論理値行列因子分解を用いた試験結果からのスキル構造の抽出.” 情報処理学会第 77 回全国大会 2 (2015): 03.

[9] Miettinen, Pauli, and Jilles Vreeken. ”mdl4bmf: Minimum description length for Boolean matrix factorization.” ACM Transactions on Knowledge Discovery from Data (TKDD) 8.4 (2014): 18.

[10] Aharon, Michal, Michael Elad, and Alfred Bruckstein. ”K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation.” IEEE Transactions on signal processing 54.11 (2006): 4311-4322.

[11] Tatsuoka, Kikumi K., ed. Analysis of errors in fraction addition and subtraction problems. Computer-based Education Research Laboratory, University of Illinois, 1984.