Vol.2018-MPS-118 No.37
Vol.2018-BIO-54 No.37
2018/6/15

テンソル分解を用いた教師なし学習による変数選択法のマルチビューデータ解析への応用

田口　善弘[1,a)]

概要: テンソル分解は昔からある手法であるが、三相以上のテンソルでは観測点が増えてしまい、フルにテンソルが埋まったデータはなかなか取得できず、テンソル分解を使うのは難しかった。そこで行列の積からテンソルを作ってテンソル分解を適用する方法を提案し、これをマルチビューデータセットにおける変数選択に用いることを提案する。

Y-h. Taguchi[1,a)]

1. はじめに

テンソルは三相以上のデータ（例えば、人 vs 購買履歴 vs 経歴）を記述する方法として非常に一般的な表現方法ではあるが、行列に比べるとデータサイエンスにあまり積極的に用いられてきたとは言えない。その理由は多岐に渡るが例えば

- 例えば、$N \times M \times K$ の三相のテンソルで表現されるデータの場合、テンソルを全ての要素を埋めるには膨大な観測データが必要である。
- ベクトルのバンドルで表現できる、従って、データ構造がイメージしやすい行列に比べると三相以上のテンソルはデータを視覚的にイメージしにくい。
- 特異値分解、ＱＲ分解、ＬＵ分解など目的別に多彩な操作が定義されている行列に比べると、テンソル分解はまだ数学的な整備が遅れている

があげられるだろう。

一方で、同じサンプルに対して異った特徴付けを行ったマルチビューデータセット（例えば、人に対して、経歴と購買機歴を別々に紐付けた2 つの行列）の解析の需要が高まっている。バイオインフォマティクスで言えば、マルチオミックスデータの解析などがそれに当たるであろう。しかし、行数と列数が完全に一致しているわけではない複数の行列を統合的に解析する方法はまだ発展途上であり、決定打は存在しない [1]。

本稿では、この「三相以上のテンソルを全て埋めるには膨大な観測が必要で困難である」という問題と「行数や列数が一致していない複数個の行列を解析する良い方法が存在しない」という問題を同時に解決する方法を提案することを目指す。[*1]

2. データと手法

2.1 テンソル分解

まず、テンソル分解について述べる。テンソル分解にはいろいろな種類があるが [3]、本稿ではタッカー分解と呼ばれる分解を用いる。簡単のため三相のテンソルを用いて説明するが、四相以上の場合への拡張は自明であろう。要素が $x_{ijk} \in \mathbb{R}^{N \times M \times K}$ であるような三相のテンソル $\mathcal{X}$ を考える。この時、タッカー分解は

$$x_{ijk} = \sum_{\ell_1=1}^{N} \sum_{\ell_2=1}^{M} \sum_{\ell_3=1}^{K} G(\ell_1, \ell_2, \ell_3) x_{\ell_1 i} x_{\ell_2 j} x_{\ell_3 k}$$

で定義される。$G \in \mathbb{R}^{N \times M \times K}$ はコアテンソル、$x_{\ell_1 i} \in \mathbb{R}^{N \times N}, x_{\ell_2 j} \in \mathbb{R}^{M \times M}, x_{\ell_3 k} \in \mathbb{R}^{K \times K}$ は特異値行列である。特異値行列は直交行列である。まず、これは明らかに過完備であり、ユニークな答えは存在しない。したがってタッカー分解の結果はタッカー分解を実行する具体的なアルゴリズムに依って変わってしまう。ここでは高次特異値分解 [3] と呼ばれるアルゴリズムを採用する。高次特異値分解で得られたタッカー分解では一般にコアテンソルは対角テンソルではない。従って、特異値ベクトル $\boldsymbol{x}_{\ell_1} \in \mathbb{R}^N, \boldsymbol{x}_{\ell_2} \in \mathbb{R}^M, \boldsymbol{x}_{\ell_3} \in \mathbb{R}^K$ は様々な組み合わせで複

[1] 中央大学理工学部物理学科
東京都, 112-8551, 日本
[a)] tag@granular.com

[*1] 本研究は原著論文として出版済みである [2]

数回掛けあわされた上で和が取られる。

2.2 行列からのテンソル作成

三相のテンソルの要素は全部で $N \times M \times K$ 個あり、全てを埋めるような計測をこなすのは難しい。一方、行列形式のデータはたくさん存在する。そこで列を共有する2つの行列 $x_{i_1 j}, x_{i_2 j}$ または行を共有する2つの行列 x_{ij_1}, x_{ij_2} からテンソルを生成することを考える (以下、行が形質、列がサンプルに対応しているとする)。

$$x_{i_1 i_2 j} = x_{i_1 j} x_{i_2 j} \tag{1}$$

または (これを以下 Case I とする)

$$x_{ij_1 j_2} = x_{ij_1} x_{ij_2} \tag{2}$$

(これを以下は Case II とする) とする。この様にすることで実際には観測されていない $N \times M \times K$ 個の要素を全て埋めることができる。

更に、共通している列または行について和をとって次元を落としたものを考えておこう。これらを

$$\tilde{x}_{i_1 i_2} = \sum_j x_{i_1 i_2 j} \tag{3}$$

または

$$\tilde{x}_{j_1 j_2} = \sum_i x_{ij_1 j_2} \tag{4}$$

を定義する。以下では和を取らないもの (式 (1)(2)) を type I、和を取ったもの (式 (3)(4)) を type II と呼称することにする。

2.3 type II の場合の特異値行列

type II の場合、列、または、行の和が取られてしまっているので $\tilde{x}_{i_1 i_2}$ または $\tilde{x}_{j_1 j_2}$ をテンソル分解しても $x_{\ell j}$ や $x_{\ell i}$ を計算することはできない。このような欠点があるにも関わらず type II のテンソルを導入する理由は単純にテンソル分解に必要な計算時間やメモリーが $\frac{1}{M}$ や $\frac{1}{N}$ になるためである。

失われた特異値行列は以下の式で計算するものとする。

$$x^s_{\ell_s j} = \sum_{i_s} x_{\ell_s i_s} x_{i_s j}, s = 1, 2$$

または

$$x^s_{\ell_s i} = \sum_{j_s} x_{\ell_s j_s} x_{i j_s}, s = 1, 2$$

とする。特異値ベクトルが2種類出来てしまうという問題があるが、計算量的には少ない要求で計算できるメリットがある。

この計算の意味は特異値行列は直交行列であり、射影であるとみなすことができるからである。元の行列の射影で失われた特異値行列を計算する。

3. 結果

テンソル (行列) の積で高次のテンソルを作成し、得られたテンソルをテンソル分解することで解析を行うという方法は一種のマルチビューデータ解析であるとみなすことができる。なぜならばテンソルを構成するために用いた複数個のテンソルは、行または列を共有しているという以外には表面的には関係がないからである。

この様な複数個の、互いに (表面上は) 無関係なテンソルを統合的に解析することの動機は、例えば，サンプル間に形質に依らない関係がある (例えば、二群に分割できる、とか) があるかどうかを調べる、あるいは、逆に同じサンプル依存性を持っている、異った形質が存在する (例えば、購買履歴と経歴のが同じような二群で差があるのであれば、その経歴は購買履歴と関係あると、みなすことができる) かどうかを調べる、などである。要するに、個々のテンソルを解析しただけでは得られない結果をマルチビューデータ解析で得ようということである。

一般的なマルチビュー解析 [1] では個々のテンソルを統合するときになんらかの重み付けを必要とする。例えば、行を共有している 10×100 の行列と 10×10000 の行列を同じ重みで解析してしまっては結果はほとんど 10×10000 の行列を個別に解析した場合と同じになってしまうであろうことは想像に難くない。しかし、それは逆に言えば結果はどのように重み付けをするかに依っているということであり、そうなると結果がそもそもデータに由来するのか、あるいは、重み付けに由来するのかの区別が曖昧になってしまう。

この点、複数のテンソルの掛け算から新しいテンソルを作るという場合には重みをどうするかという問題から独立である。なぜなら式 (1),(2),(3),(4) では、2つの行列がすべて同じ寄与で新しいテンソルが作られているからである。この様にこの論文で提案するアプローチではマルチビュー解析につきものの重みの設定が不要であるという利点がある。

3.1 人工データ

提案手法がどのような役に立つかを人工データでデモンストレーションする。1000×50 の行列を2つ用意する。行が形質の種類で列がサンプルであるとする。つまり、５０個のサンプルが有り、１０００種類の計測が行われたとしよう。2つの行列はサンプル (列) を共有しており、一方で、2つの行列で行われた観測は全く別のものであるとする (つまり、簡単のため１０００種類としたが、この値は2つの行列で異なっていてもいいとする)。従って、これはサンプルを共有している Case I のマルチビューデータ解析である。１０００種類の形質のうち、それぞれ５０個だけが「意味がある」データであとはノイズであるとする。こ

Vol.2018-MPS-118 No.37
Vol.2018-BIO-54 No.37
2018/6/15

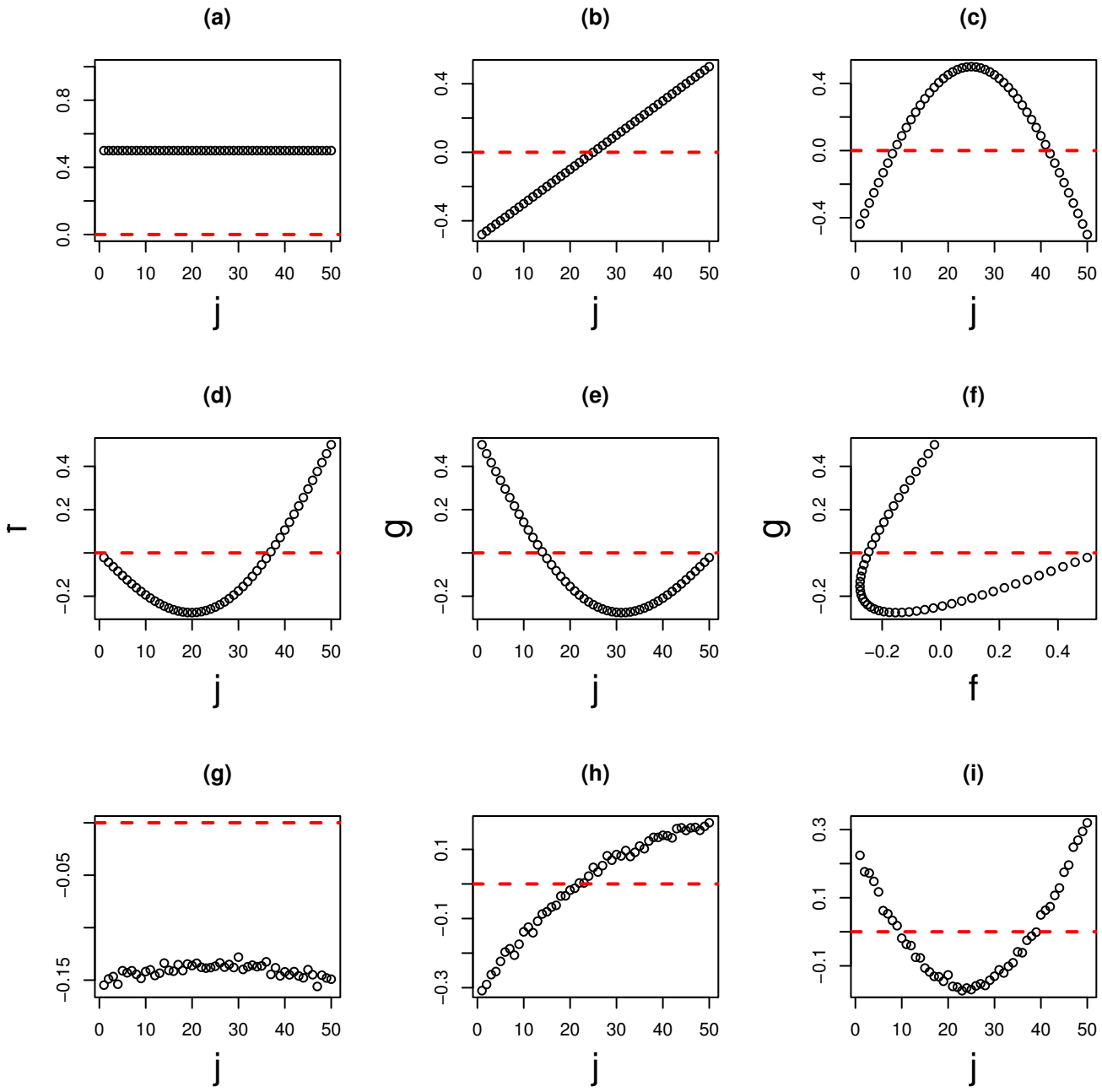

図 1　人工データの説明。(a)〜(c):サンプル依存性を共有する５０個の形質の基底となる３つの直交基底。(d)(e):(a)〜(c) から作られたサンプル依存性。２つの行列のぞれぞれに対応する。(f):(d) と (e) の相関プロット。(g)〜(i):type I テンソルのタッカー分解で得られた $x_{\ell_3,j}, 1 \leq \ell_3 \leq 3$

こで「意味がある」とは具体的には、図 **1**(d) や (e) の様な形状をしめす。２つの行列の１０００個の形質のうち、５０個だけはそれぞれ図 1(d)(e) に示すようなサンプル依存性を共有している。しかし、１０００個のうち、何番目が図 1(d) や (e) で表現されるサンプル依存性をもった形質であるかは全く対応していないとする（つまり、場合に依っては５０個ずつである必要さえなく、１０００個中２５個と７５個の形質がそれぞれの行列で共通のサンプル依存性を持っているとしても全く構わない）。実はこの図 1(d) と (e) で表現されるサンプル依存性は図 1(a),(b),(c) で表現されるような、３つの基底の線型結合で作られている。しかし、(d) と (e) は相関を見た場合には (f) にあるように無相関にしかみえなくなっている。人間が目で見れば図 1(d) と (e) は左右を反転させれば同じものだとすぐにわかるが、実際には横のサンプルの並びはこの順番である必要は全くない。横方向の順序がデタラメの場合、なんらかの秩序を見いだせるかというとそれは困難であろう。

この2つの行列 $x_{i_1 j} \in \mathbb{R}^{1000\times 50}$ と $x_{i_2 j} \in \mathbb{R}^{1000\times 50}$ から

Vol.2018-MPS-118 No.37
Vol.2018-BIO-54 No.37
2018/6/15

式 (1) から $x_{i_1 i_2 j} \in \mathbb{R}^{1000\times1000\times50}$ を作成し、タッカー分解を計算する。目的は2 つの行列で (d) 及び (e) の様なサンプル依存性を共有している５０個の形質を見つけることである。図 **2** は $x_{\ell_1,i_1}, x_{\ell_2,i_2}, \ell_1, \ell_2 = 1,2$ の散布図であ

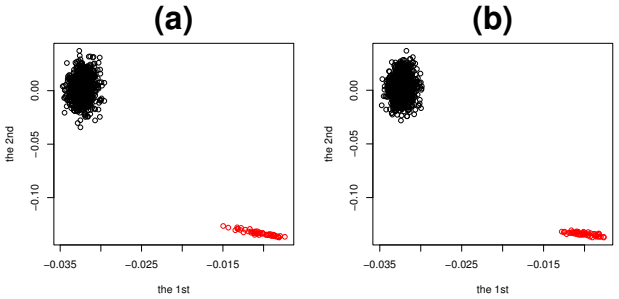

図 **2**　人工データのタッカー分解の結果。赤は図 1(d)(e) にあるようなサンプル依存性を共有した５０個の形質。(a) $x_{\ell_1,i_1}, \ell_1 = 1,2$, (b) $x_{\ell_2,i_2}, \ell_2 = 1,2$

る。図 1(d),(e) のサンプル依存性を共有している５０個の形質は赤で表記されているが、見事に他と分離されていることが解る。つまり、提案手法は教師なし学習でランダムなサンプル依存性しか持たない形質から他とサンプル依存性を共有している形質を分離する能力を持っていることが解るだろう。つまりテンソル分解を用いた教師なし学習は CaseI/typeI テンソル(式(1)) に用いることで変数選択をすることに利用可能であるということである。

また、図 1(g),(h),(i) は $x_{\ell_3,j}, 1 \le \ell_3 \le 3$ である。見事に図 1(a),(b),(c) を再現している。このことからマルチビュー解析を行うことで、別々に解析することでは決して得られない元の直交基底の再現に成功していることが解る。

3.2　マルチオミックスデータ解析

次に現実のデータの例として mRNA と miRNA のマルチビュービュー解析を行う。これらはサンプルを共有する Case I の2 つの行列である。これらをそれぞれ $x^{\mathrm{mRNA}}_{i_1 j}, x^{\mathrm{miRNA}}_{i_2,j}$ と表記する。式 (1) で三相のテンソル $x_{i_1 i_2 j}$ を作成した後タッカー分解

$$\begin{aligned} x_{i_1 i_2 j} &= x^{\mathrm{mRNA}}_{i_1 j} x^{\mathrm{miRNA}}_{i_2,j} \\ &= \sum_{\ell_1,\ell_2,\ell_3} G(\ell_1,\ell_2,\ell_3) x^{\mathrm{mRNA}}_{\ell_1 i_1} x^{\mathrm{miRNA}}_{\ell_2,i_2} x_{\ell_3 j} \end{aligned}$$

を行う。まずは、$x_{\ell_3 j}$ を観察し、健常者と患者の間で差がある ℓ_3 をみつける必要がある。図 **3** は $x_{\ell_3 j}$ であるが、$1 \le \ell_s \le 5$ では健常者と7 種類の乳がんサブタイプの間になんからの有意な差があるようになっており、タッカー分解は教師なし学習的に、サンプルのクラス間差を反映したベクトルを抽出できている。

次にこのようなサンプル差を反映した mRNA と miRNA を抽出したい。そのためには $G(\ell_1,\ell_2,\ell_3), 1 \le \ell_s \le 5$ のう

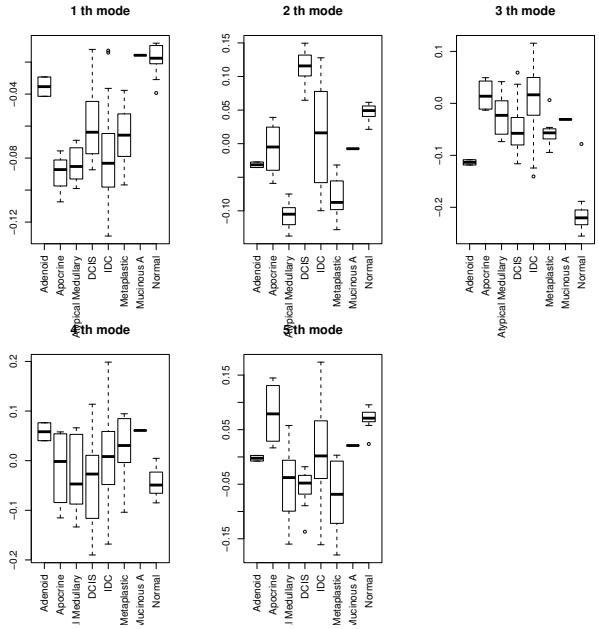

図 **3**　マルチオミックスデータの type I テンソルにタッカー分解を適用した結果得られた $x_{\ell_3 j}, 1 \le \ell_3 \le 5$。

ち、G の絶対値が大きい ℓ_1, ℓ_2 の組み合わせを見つける必要がある。このような $x^{\mathrm{mRNA}}_{\ell_1 i_1}, x^{\mathrm{miRNA}}_{\ell_2,i_2}$ が図 3 にあるような疾患・健常者依存性を反映していると考えられる。 表 **1**

表 **1**　マルチオミックスデータの Case I/type I テンソルに対するタッカー分解で得られた、絶対値が大きい順に 10 位までの $G(\ell_1,\ell_2,\ell_3), 1 \le \ell_1,\ell_2,\ell_3 \le 10$

multi-omics

ℓ_1	ℓ_2	ℓ_3	$G(\ell_1,\ell_2,\ell_3)$
1	1	1	1.67×10^5
2	1	2	-1.03×10^5
4	1	4	7.48×10^4
3	1	3	-6.64×10^4
5	1	5	6.23×10^4
3	2	3	3.00×10^4
1	2	3	-2.87×10^4
3	1	5	-2.33×10^4
2	2	3	-2.02×10^4
1	2	2	-1.48×10^4

は上位1 ０位の G である。１０位までであるが $1 \le \ell_3 \le 5$ しか出てこないので、これらが主に寄与が大きいことが確認できる。次に mRNA に関係する ℓ_1 の方は、$1 \le \ell_1 \le 5$ がでてきているのでこれら全てが患者と健常者の差を反映している特異値ベクトルであると期待できることが解る。一方、miRNA に関係する ℓ_2 の方は、$1 \le \ell_2 \le 2$ しか出てこないので、この2 つの特異値ベクトルが主に患者と健常者の差を反映している特異値ベクトルであると期待できることが解る。

実際に図 3 に見るような患者と健常者の差を特に強く反映している mRNA や miRNA はどれだろうか？人工デー

表 2 mRNA に関して選ばれた 427 プローブの MSigDB による評価。上位１０位まで。上の行: type I, 下の行: type II テンソル。"BREAST_CANCER/_DUCTAL_CARCINOMA" が強調表示。 K:各遺伝子セットの総遺伝子数, k: そのうち、４２７プローブに属するもの

Gene Set Name	(K)	Description	(k)	k/K	p-value	FDR q-value
SMID_**BREAST_CANCER**	564	Genes down-regulated in the luminal B	100	0.1773	4.00E-105	1.36E-101
_LUMINAL_B_DN		subtype of breast cancer.	88	0.1560	2.34E-090	7.94E-087
SMID_**BREAST_CANCER**	701	Genes down-regulated in basal	91	0.1298	2.06E-082	3.50E-079
_BASAL_DN		subtype of breast cancer samples.	86	0.1227	6.42E-079	1.09E-075
DOANE_**BREAST**	112	Genes up-regulated in breast cancer	44	0.3929	5.78E-063	6.56E-060
_**CANCER**_ESR1_UP		samples positive for ESR1 compared	38	0.3393	6.29E-053	4.28E-050
		to the ESR1 negative tumors.				
SMID_**BREAST_CANCER**	315	Genes down-regulated in bone	—	—	—	—
_RELAPSE_IN_BONE_DN		relapse of breast cancer.	51	0.1619	1.17E-052	6.63E-050
JAEGER_METASTASIS_DN	258	Genes down-regulated in metastases	—	—	—	—
		from malignant tumors. melanoma	43	0.1667	5.25E-045	2.55E-042
		compared to the primar				
WALLACE_PROSTATE	299	Genes up-regulated in prostate cancer	55	0.1839	5.86E-058	4.98E-055
_CANCER_RACE_UP		samples from African-American	—	—	—	—
		patients compared to those from the				
		European-American patients.				
SMID_**BREAST_CANCER**	476	Genes up-regulated in the normal-like	61	0.1282	2.40E-054	1.63E-051
_NORMAL_LIKE_UP		subtype of breast cancer.	—	—	—	—
FARMER_**BREAST**	330	Genes which best discriminated between	54	0.1636	5.31E-054	3.01E-051
_**CANCER**_BASAL		two groups of breast cancer	54	0.1636	5.03E-056	4.27E-053
_VS_LULMINAL		according to the status of ESR1 and AR: basal				
		(ESR1- AR-) and luminal (ESR1+ AR+).				
POOLA_INVASIVE_	288	Genes up-regulated in atypical ductal	51	0.1771	7.55E-053	3.67E-050
BREAST_CANCER_UP		hyperplastic tissues from patients	—	—	—	—
		with (ADHC) breast cancer				
		vs those without the cancer (ADH).				
MCLACHLAN_DENTAL	254	Genes up-regulated in pulpal tissue	47	0.1850	1.12E-049	4.77E-047
_CARIES_UP		extracted from carious teeth.	—	—	—	—
SMID_**BREAST_CANCER**	648	Genes up-regulated in basal subtype of	63	0.0972	1.38E-048	5.23E-046
_BASAL_UP		breast cancer samples.	78	0.1204	1.18E-070	1.34E-067
SMID_**BREAST_CANCER**	172	Genes up-regulated in the luminal B	38	0.2209	1.95E-043	6.63E-041
_LUMINAL_B_UP		subtype of breast cancer.	37	0.2151	3.29E-043	1.40E-040
DELYS_THYROID_UP	443	Genes up-regulated in papillary	—	—	—	—
_CANCER		thyroid carcinoma (PTC)	48	0.1084	5.47E-041	2.07E-038
		compared to normal tissue.				
TURASHVILI_**BREAST**	198	Genes down-regulated in ductal	—	—	—	—
_**DUCTAL_CARCINOMA**		carcinoma vs normal ductal breast cells.	36	0.1818	3.16E-039	1.08E-036
_VS_DUCTAL_NORMAL_DN						

Vol.2018-MPS-118 No.37
Vol.2018-BIO-54 No.37
2018/6/15

タの例 (図 2) で見たように、この様な mRNA/miRNA は他の mRNA/miRNA から大きく外れた位置にあることが期待される。そこでここでは、$x^{\mathrm{mRNA}}_{\ell_1 i_1}, x^{\mathrm{miRNA}}_{\ell_2, i_2}$ が多重ガウス分布に従っているという帰無仮設に基づいて、各 mRNA/miRNA に対して帰無仮説に対する対立仮説(多重ガウス分布には従わない) の棄却確率を付与し、この値が十分小さいものを外れ値として選択することを考える。

これを実行するため、カイ二乗分布を採用してP 値を計算した。具体的には

$$P_{i_1} = P_{\chi^2}\left[> \sum_{1 \le \ell_1 \le 5} \left(\frac{x^{\mathrm{mRNA}}_{\ell_1 i_1}}{\sigma_{\ell_1}} \right)^2 \right]$$

$$P_{i_2} = P_{\chi^2}\left[> \sum_{1 \le \ell_2 \le 2} \left(\frac{x^{\mathrm{miRNA}}_{\ell_2 i_2}}{\sigma_{\ell_2}} \right)^2 \right]$$

を用いる。ここで $P_{\chi^2}[> x]$ は引数が x より大きいカイ二乗分布の累積確率、$\sigma_{\ell_s}, s = 1, 2$ は標準偏差である。$P_{i_s}, s = 1, 2$ はベンジャミーニ＝ホッホベルク法 [4] で個別に多重比較補正した上で、補正P 値が 0.01 以下の mRNA,miRNA を選択した。

選ばれた mRNA に対する 427 プローブと、7 miRNA について、生物学的な妥当性を検討した。mRNA は MSigDB[5] に、miRNA は DIANA-mirpath[6] にアップロードされた。表 **2** が結果である(ここでは触れないが type II の結果も合わせて示してある)。上位1 0 位のうち、9 位までが type I か typeII のテンソルのタッカー分解を用いた結果で、乳がん関連の遺伝子セットと有意に重なっていることが解るだろう。このことから提案手法は生物学的に妥当な遺伝子群を選択する能力を持っていることがわかる。表 **3** は miRNA についての DIANA-mirpath の結果である(こちらもここでは触れないが type I/II 両方の結果が掲載されている)。こちらもがん関係のＫＥＧＧパスウェイが上位にならんでいることがわかるだろう。

これらのことからテンソル分解を用いたマルチオミックスデータのマルチビューデータ解析は、人工データのみならず、現実の遺伝子/miRNA 発現プロファイルの統合解析においても力を発揮することが確認できた。

4. 終わりに

本稿では、複数種のテンソルの積から構成した高次のテンソルをタッカー分解することで教師なし学習的にサンプルのクラス依存性を自動的に腫瘍室する方法を提案し、それを実行データと現実のデータに応用した。原論文 [2] は長大な長さでその全体をここで説明することはできない。興味を持たれた方は是非、原論文をあたっていただきたい。

表 **3** miRNA に対する DIANA-mirpath の結果。上位1 0 位までの KEGG パスウェイが表記されている。gene: 7 miRNA の標的との重なり数, miRNA: 7 miRNA との重なり数。 ／ の左右がそれぞれ type I/type II。

KEGG pathway	FDR q-value	gene	miRNA
MicroRNAs in cancer	3.29E-88/4.68E-68	115/141	7/22
Proteoglycans in cancer	5.36E-12/9.48E-17	116/159	7/22
Cell cycle	1.26E-10/2.61E-12	80/104	7/22
Renal cell carcinoma	—/2.25E-011	—/61	—/22
Protein processing in endoplasmic reticulum	—/3.81E-10	—/134	—/22
Hepatitis B	6.57E-09/5.37E-09	79/107	7/22
Prion diseases	5.14E-08/—	16/—	7/—
Central carbon metabolism in cancer	2.76E-07/—	42/—	7/—
Hippo signaling pathway	3.27E-07/2.40E-07	78/109	7/22
Chronic myeloid leukemia	—/2.40E-07	—/62	—/22
Viral carcinogenesis	—/2.40E-07	—/158	—/22
Pancreatic cancer	—/1.84E-06	—/55	—/22
Lysine degradation	1.15E-06/—	27/—	6/—
FoxO signaling pathway	2.89E-06/—	79/—	7/—
Prostate cancer	4.52E-06/—	56/—	7/—

参考文献

[1] Li, Y., Wu, F.-X. and Ngom, A.: A review on machine learning principles for multi-view biological data integration, *Briefings in Bioinformatics*, p. bbw113 (online), DOI: 10.1093/bib/bbw113 (2016).

[2] Taguchi, Y.-H.: Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing, *PLOS ONE*, Vol. 12, No. 8, p. e0183933 (online), DOI: 10.1371/journal.pone.0183933 (2017).

[3] 石黒勝彦，林　浩平：関係データ学習 (機械学習プロフェッショナルシリーズ)，講談社 (2016).

[4] Benjamini, Y. and Yekutieli, D.: The Control of the False Discovery Rate in Multiple Testing under Dependency, *The Annals of Statistics*, Vol. 29, No. 4, pp. 1165–1188 (online), available from ⟨http://www.jstor.org/stable/2674075⟩ (2001).

[5] Liberzon, A., Birger, C., Thorvaldsdttir, H., Ghandi, M., Mesirov, J. and Tamayo, P.: The Molecular Signatures Database Hallmark Gene Set Collection, *Cell Systems*, Vol. 1, No. 6, pp. 417 – 425 (online), DOI: https://doi.org/10.1016/j.cels.2015.12.004 (2015).

[6] Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T. and Hatzigeorgiou, A. G.: DIANA-miRPath v3.0: deciphering microRNA function with experimental support, *Nucleic Acids Research*, Vol. 43, No. W1, pp. W460–W466 (online), DOI: 10.1093/nar/gkv403 (2015).