一般口演|医療データ解析
一般口演11
機械学習
2021年11月20日(土) 09:10 ~ 11:10 G会場 (2号館3階232+233)

[3-G-1-05] 臓器ごとの遺伝子発現データに対する継続学習技術の検討

*鈴木 麻由美¹(1. 日立製作所)

*Mayumi Suzuki¹(1. 日立製作所)

キーワード: Life-long Learning, Machine Learning, Genome Informatics

【目的】 医療費の削減と患者 QoL向上を目指し,がんの早期発見および信頼性の高い治療効果予測を実現するためには,日進月歩の医療技術に追従し,かつ過去知識も兼ね備えた予測モデル構築が重要である。そこで,逐次的に複数のデータを学習した際に,一度学習したデータを再学習することなく,学習したすべてのデータに対して高精度な予測を実現する Run Once Memory Aware Synapses (RO-MAS)継続学習手法を提案する。

【方法】 RO-MAS手法では、当該データの学習により生成された予測モデルに対する出力の勾配を正則化係数に 利用することで、当該データの知識を保存する。実験データは、GTExおよび TCGAにおける RNA-Seqデータに 対し、mskccによるパイプライン処理を行った4種の臓器に関するデータセットとした[1]。継続学習手法を用い ない従来手法と RO-MAS手法を用いて4種のデータを逐次的に学習させ、がん細胞由来サンプルの予測モデルをそ れぞれ生成した。なお、機械学習モデルは deep neural networkとした。

【結果】 RO-MAS手法における予測精度は、4種のデータ全体で平均して88.33%となった。従来手法と比較して、14.26%高かった。

【考察】 RO-MAS手法において,4種のデータにおける各予測精度のうち最大で22.28%の差があった。最後に学習したデータと過去のデータにおいて発現している遺伝子の種類が大きく異なる場合に,正則化項による過去知識の保持が不十分になっていることが原因と考えられる。

【結論】 RO-MAS手法により,逐次的に複数のデータを学習した際に,一度学習したデータを再学習することな く全データで高精度な予測が可能であることを確認できた。

[1] Qingguo Wang et al.: bioRxiv 110734, 2017.

臓器ごとの遺伝子発現データに対する継続学習技術の検討

鈴木麻由美^{*1}、吉田英恵^{*1} *1株式会社 日立製作所 研究開発グループ

Examination of Life-long Learning Technology for Gene Expression Data in Each Organ

Mayumi Suzuki^{*1}, Hanae Yoshida^{*1}

*1 Hitachi, Ltd. Research and Development Group

Early cancer detection and optimal treatment selection by treatment effect prediction is expected to reduce medical costs and improve patients' quality of life. In order to achieve treatment effect prediction which can also follow the fast-evolving cancer diagnosis, there is a need for lifelong learning technology that can learn and update it with new data while ensuring the prediction performance. A catastrophic forgetting is generally caused in machine learning due to the loss of the previously trained task information when a new task is trained. In this study, we proposed a lifelong learning technique, Run Once Memory Aware Synapses (RO-MAS), which stores the knowledge of previous task to the regularization term. This method is possible to train efficiently because of using no previous task data for the new prediction models. We applied the proposed method and evaluated the prediction accuracy with sequential task training using by MNIST dataset and gene expression dataset. Through the experiment, we achieved high prediction accuracy without forgetting for all tasks by using the proposed method. In gene expression dataset results, there was a maximum of 22.5% difference in prediction accuracy and reduce catastrophic forgetting using by proposed method.

Keywords: Life-long Learning, Machine Learning, Genome Informatics

1. 背景および目的

医療費の削減と患者 QoL 向上へ向け、がんの早期発見お よび治療効果予測による個別最適な治療法選択支援が期待 されている。実現には、日進月歩の医療技術の情報に追従し、 かつ過去知識も兼ね備えた予測モデル構築が重要である。 一般的に機械学習では、新しいタスクを学習すると以前に学 習したタスクの情報を失う、破滅的忘却(catastrophic forgetting)が発生する¹⁾。逐次的に異なるタスクが与えられる 場合、各タスクのデータをすべて記憶しておき、学習時に再 度利用することが考えられるが、タスク数に比例したメモリおよ びロード時間が必要となる。

近年、破滅的忘却を起こすことなく、学習を行う継続学習 (lifelong learning, continual learning, sequential learning また は incremental learning とも呼ばれる)の研究が活発化してい る。継続学習の重要な要件として、sample efficient、memory efficient、time efficient の3 つがある $^{2)}$ 。Sample efficient は、 新しいタスクの学習時に、過去のタスクのデータを参照しない ことを指す。Memory efficient は、過去のタスクのデータを保 存するメモリとモデルのネットワーク構造のサイズがタスクの数 に比例して増加しないことを指す。Time efficient は、新しいタ スクの学習をリアルタイムで処理可能であることを指す。これら の要件は、ヘルスケアデータが抱える問題であるデータの永 続的な増加によるストレージメモリのひっ迫や患者の個人情 報保護、医療技術の進歩による新しい薬剤や治療法の出現 への対応可能性を示唆している。そこで、逐次的に複数のタ スクを学習する際に、一度学習したタスクのデータを再学習 することなく、学習したすべてのデータに対して高精度な予測 を実現する Run Once Memory Aware Synapses (RO-MAS)継 続学習手法を提案する。

2. 方法

2.1 継続学習手法の特徴

継続学習手法は、過去タスクのデータの知識の保持方法と、 新タスクを学習する際にどのような学習プロセスで過去の知 識を使用するかに基づき、大きく3種類に分けられる3。まず、 1 種類目である replay-based methods は、raw または圧縮等に より保存しておいた過去タスクのデータを、新タスクの学習に おいて再度使用(replay)する手法である。続いて、 regularization-based methods は、新タスク学習時の損失関数 に過去タスクの知識を持った正則化項を追加する手法である。 最後に、parameter isolation-based methods は、各タスク専用 のモデルパラメータやモデル構造を使用する手法である。本 研究では、1度学習したタスクを再学習しないこと、新タ スクの学習時に保持したい過去タスクを選択できること、 各タスクで共通のモデル構造を使用可能であることを重 視し、regularization-based methods である RO-MAS を提案し た。RO-MAS では、当該タスクの学習により生成された予測 モデルに対する出力の勾配を正則化項に利用することで、当 該タスクの知識を保持する。

2.2 Run Once Memory Aware Synapses

機械学習において、タスク T_m に関する予測モデルを生成 する際、予測関数に学習データを入力した際の予測値と実 際の値や識別クラス番号である正解ラベルとの誤差が最小と なるようにモデルパラメータ θ_m を計算する損失関数 $L(\theta_m)$ を 設定する。損失関数 $L(\theta_m)$ は、誤差関数 $E(\theta_m)$ 、正則化項 $R(\theta_m)$ とモデルパラメータ θ_m を用いて、式(1)で与えられる。

$$L(\theta_m) = E(\theta_m) + R(\theta_m) \tag{1}$$

一般的に、正則化項 $R(\theta_m)$ は、L1 ノルムやL2 ノルムなどが 使用される。L2 ノルムを使用した際の正則化項 $R(\theta_m)$ は、正 則化係数 λ とモデルパラメータ θ_m を用いて、式(2)で与えられ る。

$$R(\theta_m) = \lambda \theta_m^2 \tag{2}$$

一方、Memory Aware Synapses (MAS) 手法では、学習した データの知識を保持するために、正則化項 $R(\theta_m)$ は式(3)で 与えられる⁴。

$$R(\theta_m) = \lambda \sum_{ij} \Omega^{ij} \left(\theta_m^{ij} - \theta_{m-1}^{ij} \right)^2 \tag{3}$$

ここで、 θ_{m-1}^{ij} は学習開始時に予測モデルの初期値としたモ デルパラメータ、 Ω^{ij} は保持するタスクの知識を表す合成知識 係数、ijはモデルパラメータのかかるニューロンの位置である。 タスク T_m の学習時に、過去に学習したタスク T_1 およびタスク T_2 も予測可能とする場合、合成知識係数 Ω^{ij} はタスク T_1 およ びタスク T_2 の知識係数から生成する。RO-MAS における、タ スク T_1 の知識係数 $I^{ij}(x_l; \theta_l)$ は式(4)で、合成知識係数 Ω^{ij} は 式(5)で与えられる。

$$I^{ij}(x_m;\theta_m) = \frac{\partial \|h_m(x_m;\theta_m)\|}{\partial \theta^{ij}}$$
(4)

$$\Omega^{ij} = \frac{1}{P} \sum_{p=1}^{P} \sum_{l}^{U} \left\| l^{ij}(x_l; \theta_l) \right\|$$
(5)

ここで、 $h_m(x_m; \theta_m)$ はモデルパラメータ θ_m から構成される タスク T_m の予測関数 h_m にタスク T_m の入力データ x_m を入力し た出力結果、Uは予測可能とする過去に学習したタスク群、P は総データポイント数である。式(5)が示すように、過去に学習 したタスク Tiのデータとタスク Tiを学習した際に生成したモデ ルのパラメータθ,を用いて知識係数I^{ij}(x₁; θ₁)を算出する。つ まり、タスク Ti のデータは予測モデルを決定し、知識係数 I^{ij}(x_l; θ_l)を算出した後には、新タスクの学習において不要と なる。RO-MAS における処理フローの説明図を図1に示す。 以下、簡易化のため、タスク Tmの知識係数は Imと表す。タス ク T_1 のデータ D_1 を deep neural network や decision tree など の機械学習手法により学習し、タスク Tiを高精度に予測可能 な予測モデル M1を生成する。同時に、予測モデル M1とデー タD1を用いて、知識係数L1を生成する。次に、タスクT2が到 来した際には、予測モデル M1 を初期値として、知識係数 I1 を用いて、タスクT2のデータD2をRO-MAS 手法により学習し、 タスク T1 およびタスク T2 を高精度に予測可能な予測モデル M2を生成する。同時に、予測モデル M2とデータD2を用いて、 知識係数 L2を生成する。

2.3 実験方法

2.3.1 逐次的なタスクの設定方針

RO-MAS では、各モデルパラメータの変動を制御すること で過去タスクの知識を保持するため、全タスクにおいて同じモ デル構造である必要がある。つまり、モデルの入力と出力の 次元数が同じである必要がある。また、例えば入力が医用画 像のタスクと血液検査値のタスクなど特徴量の意味が異なる タスクを逐次的に学習する場合、結果の解釈が難しくなる。さ らに、実際のヘルスケアデータにおいて想定される各タスク は、異なる病院や診療科のデータや同じ診療科で取得され た同種のデータであるが半年ごとに到来するデータなどであ ると想定される。そのため、本稿では、各タスクの条件を以下 とした。

1. 各タスクにおいて、入出力データの次元数が同じ

2. 各タスクにおいて、入出力データの単位が同じ

3. 各タスク間で、入力データの分布や傾向が異なる 上記条件に伴い、本稿における学習シナリオは、学習後に到 来する、または精度検証を行うデータが属するタスクが自明 である incremental task learning とし、 incremental task learning で典型的な multi-headed output layer を使用する ⁵。



図1 RO-MAS 手法における処理フローの説明図

2.3.2 MNIST データセット

実験データとして、手書き数字認識データセットである MNIST を用いた^の。MNIST データセットは、28×28 画素の0 から9までの数字からなる10クラスのデータである。このデー タをタスク T1とした。タスク T2から T4は、タスク T1のデータに 対して permutation 処理を行い、各タスクにおいて画素の並 びのみが異なるを生成した⁷⁾。Permutation 処理は, 画像内の ピクセルを一定の法則で並び替えることで、新しいタスクを生 成する手法である。各タスクともに、全 70,000 サンプルのうち 学習および検証データに 60,000 サンプル、テストデータに 10,000 サンプルを用いた。また、学習したタスクの増加に伴う 継続学習の効果を見るため、逐次的に到来するタスク数は 4 とした。4 つのタスクを逐次的に学習し、最終的な予測モデル を用いて各タスクのデータに対する予測精度により評価を行 った。なお、機械学習モデルは deep neural network とし、grid search により最適なモデルを決定した。比較のため、従来法と して、継続学習手法を用いず、直前のタスクにて生成した予 測モデルを初期値とする finetuning を行った。 ハイパーパラメ ータの探索範囲を表1に示す。

2.3.3 臓器ごとの遺伝子発現データセット

2つ目の実験データには、2.3.1項で述べた逐次的なタスクの設定方針を満たすヘルスケア関連データセットとして、臓器ごとの遺伝子発現データセットを用いた。GTEx および TCGAにおけるRNAシーケンスデータに対し、mskccを用いたパイプライン処理を行った遺伝子発現データセットから4種の臓器データを用いた。なお、各臓器データは、以下のサイトから2018年5月15日に commit されたバージョンを取得した

(https://github.com/mskcc/RNAseqDB/tree/master/data/norma lized)。TCGA のがん由来サンプルをがんクラス、TCGA の非 がん由来サンプルと GTEX のサンプルを合わせて非がんクラ スとした 2 クラス分類問題とした。また、クラス間のサンプルバ ランスの観点から、タスク Tiを甲状腺がん、タスク T2を胃がん、 タスク T3を肺腺がん、タスク T4を大腸がんとした。各遺伝子デ ータの特徴量は、4 つのタスクで共通する特徴量のみに限定 し、18,829 個とした。各タスクともに、80%を学習および検証デ ータ、残りの20%をテストデータとした。各データセットのクラス 数や特徴量数等の構成を表 2 に示す。4 つのタスクを逐次的 に学習し、最終的な予測モデルを用いて各タスクのデータに 対する予測精度により評価を行った。なお、機械学習モデル は deep neural network とし、タスク Ti は optuna を用いて最適 なモデルを決定した⁸。RO-MAS では、以降のタスクでは、モ デル構造は固定とし、learning rate と weight decay のみ grid search により探索し、最適なモデルを決定した。比較のため、 従来法として、継続学習手法を用いず、直前のタスクにて生 成したモデルを初期値とする finetuningを行った。ハイパーパ ラメータの探索範囲を表 3 に示す。

表1 ハイパーパラメータの探索範囲								
(MNIST データセット)								
Hyper parameter Optimization				nization r	ange			
T ₁ tuning and finetuning								
Lea	rning ra	te	10-	¹ , 10 ⁻² , 1	0 ⁻³			
RO-MAS								
Lea	Learning rate			10 ⁻¹ , 10 ⁻² , 10 ⁻³				
Weig	Weight decay			10 ⁻² , 10 ⁻³ , 10 ⁻⁴ , 10 ⁻⁵ , 10 ⁻⁶				
Common								
Numb	Number of layers			2				
Numb	Number of units			256				
Ba	Batch size			200				
Numb	Number of epoch			20				
Activat	Activation function			ReLU				
Op	otimizer		SGD					
表 2 MNIST データセットおよび								
	遺伝子発現データセットの構成							
				Number of				
	clas-	variab-	train	valid.	test			
	ses	les						
MNIST	10	784	48,000	12,000	10,000			
甲状腺 がん	2	18,829	489	163	163			
胃がん	2	18,829	364	121	121			
肺腺 がん	2	18,829	526	175	177			
大腸 がん	2	18,829	460	153	153			

3.実験結果

MNIST データセットにおけるハイパーパラメータの探索結 果を表 4 に示す。また、MNIST データセットにおいて、タスク T₁から T₄まで逐次的に学習し決定した最終的な予測モデル を用いた各タスクのデータに対する予測結果を表 5 に示す。

遺伝子発現データセットにおけるハイパーパラメータの探索結果を表 6 に示す。また、遺伝子発現データセットにおいて、タスク T₁から T₄まで逐次的に学習し決定した最終的な予測モデルを用いた各タスクのデータに対する予測結果を表 7 に示す。

表3 ハイパーパラメータの探索範囲									
(遺伝子発現データセット)									
Typer parameter Optimization range									
Number of	avers		1 - 5						
Number of	units		4 — 128						
Learning	rate		10 ⁻¹ — 1	0 ⁻⁵					
	Finetuning								
Learning rate 10 ⁻¹ , 10 ⁻² , 10 ⁻³									
RO-MAS (grid search)									
Learning	rate	10 ⁻¹ , 10 ⁻² , 10 ⁻³							
Weight de	ecay	10 ⁻² , 10 ⁻³ , 10 ⁻⁴							
	C	ommon							
Batch si	ze		200						
Number of e	epoch	20							
Activation fu	Inction	ReLU							
Optimiz	Optimizer			SGD					
表 4	ハイパーノ	ペラメータ	の探索結	果					
	(MNIST	データセ	ット)						
Hyper		Mod	el of task						
Method	eter od Conventional RO-MAS								
Task	$T_1 T_2$	T ₃	T ₄ T ₂	$T_3 T_4$					
Learning rate		10 ⁻¹		10-4					
Weight decay		_		10 ⁻⁶					
表 5	予測結果	(MNIST	データヤッ	· ト)					
Method		Accu	uracy (%)	.,					
	T 1	<i>T</i> ₂	<i>T</i> ₃	T 4					
Conventional	16.03	9.04	12.09	98.04					
RO-MAS	97.92	93.25	93.76	95.74					
表6~	ハイパーバ	ラメータの	の探索結果	Ę					
	(遺伝子発	現データ	セット)						
Hyper		Mode	l of task						
parameter Method	C	onventior	nal R(
Task	T ₁ T	$\overline{f}_2 = T_3$	T_4 T_2	$T_3 T_4$					
Number of		2 .0							
layers	1		-						
Number of	43		_						
Units Learning									
rate 0.094		10 ⁻³							
Weight	_	1	∩- 2	10 ⁻³					
decay			0	10					
表7予	表7 予測結果 (遺伝子発現データセット)								
Method		Accı T	uracy (%) -						
Convertion	<i>I</i> ₁	12	13	14					
	46.01	69.92 72.47	84.18 85.00	95.45					
RO-MAS	95.71	73.17	85.88	95.45					

4. 考察

MNIST データセットを用いた実験において、従来手法では、タスク T4以外のタスクの予測精度が17%以下と低い結果となった。最新の予測モデルは、過去に学習した知識を忘却していることがわかる。RO-MAS 手法では、全タスクとも予測

精度が93%以上と高い結果となった。つまり、最新の予測モ デルは、全タスクにおいて知識を保持できていると考えられる。

一方、遺伝子データセットを用いた実験において、従来手 法では、タスク T4以外のタスクの予測精度は直近に学習した タスクほど高い傾向にあった。RO-MAS 手法では、従来法と 比較してどのタスクも高い予測精度を達成しているものの、タ スク T1、T4、T3、T2の順で予測精度が高くなっていた。また、 MNIST データセットにおける RO-MAS 手法の結果と異なり、 予測精度の差が最大で22.5%あった。つまり、知識を保持で きているタスク(タスク T₁)と保持できていないタスク(タスク T₂、 T3)が存在していると考えられる。RO-MASは、まずタスクT1を 学習し生成した予測モデル M1において、変化すると出力が 大きく変わるモデルパラメータを特定し、以降のモデル更新 にて変化しないよう知識係数 Inにより制約を加える。そのため、 タスク T2 において、知識係数 I1 により制約が加えられたモデ ルパラメータが予測精度の向上に関わっている場合、タスク T2 の予測精度を高くすることが難しくなると考えられる。つまり、 タスク間で予測に重要な特徴量が異なる場合、後から学習す るタスクは適切なモデルパラメータが得られず、高い予測精 度が得られないと考えられる。ここで、各タスクにおける重要 な特徴量について考える。遺伝子発現データセットにおける タスク Ti は甲状腺がんである。甲状腺がんのドライバー変異 として RET、 BRAF、 RAS、 PAX8/PPARG、 PAX8/GLIS、 APC が良く知られている⁹⁾。タスク T2 である胃がんは、ピロリ菌の有 無や PIK3CA、 RHOA、 p53 や HER2 の変異、 JAK2PD-L1 や PD-L2 の過剰発現が発生に関連があるとされている¹¹⁾。タス クT3である肺腺がんは、EGFR変異、ROS1融合遺伝子変異、 ALK融合遺伝子およびRET融合遺伝子などが発生原因とし て広く知られている¹¹⁾。タスクT4である直腸がんと結腸腺がん を含む大腸がんは、RAS、BRAF、PIK3CA、APC などの遺伝 子変異が関係していることが広く知られている11)。このように、 BRAF、p53、PIK3CAや EGFR などの多くのがんで共通して 重要な遺伝子は4種のがんでも共通している。ただし、特に 胃がん(タスクT2)は他の3つのがんと共通した重要な遺伝子 が少ない。次点で肺腺がん(タスク T3)が他の3つのがんと共 通した重要な遺伝子が少ない。以上より、各タスクで共通する 重要な遺伝子から、タスクT2およびT3が知識を保持できてい ないタスクとなっている理由の1つであると考えられる。

続いて、RO-MAS 手法を用いた際に、MNIST データセット では、全タスクにおいて知識を保持できたものの、遺伝子発 現データセットではタスクT2およびT3の知識を保持できなか ったことに注目する。MNIST データセットにおける各タスクの データは、T1から一定の法則による並び替え処理が行われた データであるため、各タスク間の相関が存在する。各タスク間 に相関が存在する場合、各タスク固有の最終層のモデルパラ メータの変更により、知識係数により制約されるモデルパラメ ータの変更を最小限に抑えることが可能であると考えられる。 そこで、遺伝子発現データの各タスクのデータ間の分布の類 似度を Jensen-Shannon divergence を用いて各特徴量ごとに 算出した 12)。各タスク間の特徴量の組み合わせごとに算出し た類似度の平均を表8に示す。値が大きいほど分布が類似し ていないことを表す。タスク T1 が他の 3 つのタスクと 1.4 倍以 上類似していないことがわかる。タスク T2 はタスク T3 および T4 と同程度に類似したデータセットであると言える。本結果から、 最初にタスクT1の知識を保持したことが、タスクT2およびT3 が知識を保持できていないタスクとなっている理由の1つにな り得ると考えられる。つまり、逐次的に学習させるタスクの順番 は相関が高いまたは重要な特徴量が共通しているデータセッ トにて実施することで、全タスクにおいて高い予測精度を達成 可能であることが示唆される。

表8 各タスク間の特徴量ごとに算出した類似度の平均値 (遺伝子発現データセット)

	T 1	<i>T</i> ₂	T ₃	<i>T</i> ₄
<i>T</i> ₁	_	3.445×10^{-2}	2.893×10^{-2}	3.586×10^{-2}
T 2	—	—	1.604×10^{-2}	1.564×10^{-2}
T 3	—	—	_	2.026×10^{-2}
T 4	_	_	_	_

5. 結論

Run Once Memory Aware Synapses (RO-MAS)継続学習 手法により、一度学習したタスクのデータを再学習することな く、学習したすべてのデータに対して高精度な予測を実現可 能であることが確認できた。今後、学習するタスクの順序に依 らず高精度な予測を達成可能とする手法の改善や、ヘルスケ アに関連するデータ利活用を促進する手法の開発が期待さ れる。

参考文献

- R. M. French. Catastrophic forgetting in connectionist networks. Trends in cognitive sciences, 1999; 3:4:128-135.
- A. Chaudhry, M. Ranzato, M. Rohrbach, M. Elhoseiny. Efficient lifelong learning with A-GEM. ICLR 2019. arXiv preprint arXiv:1812.00420, 2019.
- [https://arxiv.org/abs/1812.00420 (cited 2021-Aug-23)]. 3) De Lange, M., Aljundi, R., Masana, et al. Continual learning: A
- comparative study on how to defy forgetting in classification tasks. arXiv preprint arXiv:1909.08383 2019. [https://arxiv.org/abs/1909.08383 (cited 2021-Aug-23)].
- 4) R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, T. Tuytelaars. Memory aware synapses: Learning what (not) to forget. ECCV 2018; 139-154.
- Y.-C. Hsu, Y.-C. Liu, Z. Kira. Re-evaluating continual learning scenarios: A categorization and case for strong baselines. arXiv preprint arXiv:1810.12488, 2018.
- [https://arxiv.org/pdf/1810.12488.pdf (cited 2021-Aug-23)]. 6) Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner.
- Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278-2324, 1998.
- Ian J Goodfellow, Mehdi Mirza, Da Xiao, Aaron Courville, Yoshua Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. arXiv preprint arXiv:1312.6211, 2013.

[https://arxiv.org/abs/1312.6211 (cited 2021-Aug-23)].

- Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. KDD 2019.
- 9)加藤良平.遺伝子変異による甲状腺癌の再分類.日本内分泌・ 甲状腺外科学会雑誌,2020:37:2:68-77.
- 10) The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014; 513: 202-209.
 [https://www.netwo.com/orticles/netwo12480 (cited)]

[https://www.nature.com/articles/nature13480 (cited 2021-Aug-23)].

- NPO 法人キャンサーネットジャパン. もっと知って欲しいがんのバ イオマーカーのこと. 2016. [https://www.cancernet.jp/upload/w_baio161214.pdf (cited 2021-Aug-23)].
- 12) Jianhua Lin. Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory 1991; 37-(1): 145-151.