依存関係の大きさは意味の関連性を表す

大山百々勢 [1,2]　山際宏明 [1]　下平英寿 [1,2]
[1] 京都大学　[2] 理化学研究所
oyama.momose@sys.i.kyoto-u.ac.jp, hiroaki.yamagiwa@sys.i.kyoto-u.ac.jp,
shimo@i.kyoto-u.ac.jp

概要

単語ベクトルの点群を可視化して高次元空間でどのように単語の意味が表現されているのかを解釈する手段として独立成分分析 (ICA) を用いることができる．ところが現実のデータに適用すると ICA によって得られる成分は互いに無相関であるが独立ではなく，成分間には依存関係がある．本稿では単語ベクトルを ICA で変換して得られる成分間の依存関係を 3 次以上の混合モーメントで定量化し，その依存関係が何を表しているのかを理解することを目標とする．実験を行った結果，依存関係が大きい成分同士は意味の関連が強いことがわかった．また，成分間の依存関係への寄与が大きな単語によって関連性を具体的に解釈することができた．

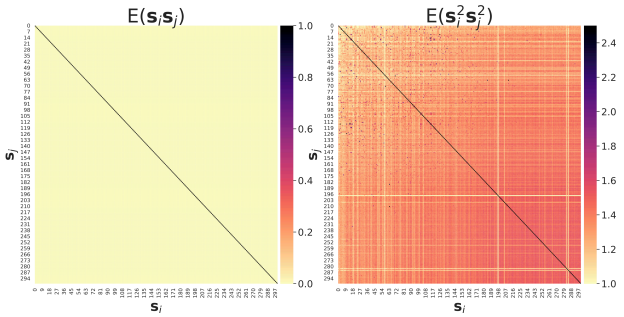

図 1　単語ベクトルを ICA で変換して得られる成分のペア $(\mathbf{s}_i, \mathbf{s}_j)$ の相関係数 $\mathrm{E}(\mathbf{s}_i\mathbf{s}_j)$ (左) と 4 次の混合モーメント $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ (右) を表すヒートマップ．左のヒートマップの非対角成分に着目すると，異なる独立成分間の線形相関は 0 であることがわかる．一方右のヒートマップでは，成分間が独立であるときの値である 1 を超える成分のペアが多く，これらの成分間には非線形な依存関係が存在することが示唆される．

1　はじめに

単語ベクトルは Skip-gram with Negative Sampling (SGNS) [1] のようなシンプルなモデルで得られるものから，複雑なニューラルネットワークに基づく言語モデル [2, 3, 4, 5] の内部表現まで，言語の意味理解と表現において重要な役割を果たしている．しかし高次元空間においてこれらのベクトルがどのように言葉の意味を表現しているのか解釈するのは難しく，様々な議論がされている．

単語ベクトルが高次元空間でどのように意味を表現しているのかを解釈する手段として独立成分分析 (ICA) [6] を用いることができる [7, 8]．ICA はデータを線形変換して可能な限りデータを独立に表現するような座標軸を見つける手法である．ICA を単語ベクトルに適用して得られる各座標軸は，成分値が大きな単語によって解釈することができる．

現実のデータに ICA を適用すると，得られる成分同士は互いに無相関であるが独立ではなく，成分間には依存関係があることが知られている [9, 10, 11]．ICA によって得られる成分間に依存関係が残る主な理由の一つは，ICA では捉えることができない非線形な依存関係が現実のデータには存在し，結果として ICA が仮定するデータの独立性が満たされないことにある．図 1 に単語ベクトルを ICA で変換して得られる成分のペア $(\mathbf{s}_i, \mathbf{s}_j)$ の相関係数 $\mathrm{E}(\mathbf{s}_i\mathbf{s}_j)$ と 4 次の混合モーメント $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ を示す．この図からも，ICA を適用して得られる成分間の相関は 0 であるものの，これらの成分は実際には独立ではなく非線形な依存関係があることがわかる．

本稿では単語ベクトルを ICA で変換して得られる成分間の依存関係を 3 次以上の混合モーメントで定量化し，その**依存関係が何を表しているのかを理解することを目標とする**．

本稿の構成　まず 2 章で単語ベクトルを ICA で変換して得られる独立成分が解釈可能であることをみる．続く 3 章では独立成分間の依存関係を定量化する 3 次以上の混合モーメントについて述べる．4 章では依存関係が大きい成分同士は意味の関連が強いことと，その関連性は依存関係への寄与が大きな単語によって解釈できることを実験で示す．

2　背景：単語埋め込みの独立成分

独立成分分析 (ICA) はデータを線形変換して可能な限りデータを独立に表現する座標軸を見つける手法である．この章では ICA を単語ベクトルに適用して得られる各座標軸は，その成分値が大きな単語によって解釈できることを確認する．

2.1　記号の定義

n 個の単語 $w_1, \cdots, w_n$ の中心化した単語ベクトルを各行に並べた行列を $\mathbf{X} \in \mathbb{R}^{n \times d}$ と表す．ここで d は単語ベクトルの次元である．

行列 $\mathbf{X}$ に対して，ICA は変換後の行列 $\mathbf{S} \in \mathbb{R}^{n \times d}$ の各列が統計的になるべく独立になるような変換 $\mathbf{B} \in \mathbb{R}^{d \times d}$ を求める．ICA の変換は以下の式で与えられる．

$$\mathbf{S} = \mathbf{X}\mathbf{B}$$

$S_{t,i}$ の値は ICA で得られた i 番目の座標軸における単語 w_t のベクトルの成分値を表す[1)]．行列 $\mathbf{S}$ の各列を独立成分と呼び，以下では i 列目の独立成分を $\mathbf{s}_i \in \mathbb{R}^n$ と表記する．

2.2　単語ベクトルの独立成分の解釈可能性

ICA を単語ベクトルに適用して得られる各座標軸は，その成分値が大きな単語によって解釈できることを確認する．

実験設定　実験に使用した単語ベクトルは，Skip-gram with Negative Sampling (SGNS) を用いて学習した．学習に使ったコーパスは text8 コーパス [12] である．単語数は n = 253,854 であり，次元数は d = 300 とした．ICA の計算には Scikit-learn に実装されている FastICA [13] を用いており $\mathbf{S}$ は白色化されている．

実験結果　表 1 は，各座標軸において，ICA で変換した後に正規化した単語ベクトルの成分値が大きな単語[2)]と，その成分値を示したものである[3)]．各座標軸が持つ意味はその軸の成分値が大きな単語によって解釈できる．例えば第 0 軸の成分値が大きな単語は *dishes, sause, fried, ...* であることから，第 0 軸は「食べ物」の意味を持つ軸であると解釈できる．同様に，第 30 軸は「ラジオ放送」，第 60 軸は「河川」，第 90 軸は「アメリカの漫画」，第 120 軸は「季節・天候」，第 150 軸は「核」の意味をそれぞれ持つと解釈できる．

表 1　各座標軸において，ICA で変換した後に正規化した単語ベクトルの成分値が大きな単語とその成分値の表．text8 コーパスにおいて 100 回以上登場する単語のうち，成分値の上位 5 単語をそれぞれ表示した．

0 軸		30 軸		60 軸	
dishes	0.511	stations	0.625	river	0.626
sauce	0.495	fm	0.619	tributaries	0.538
fried	0.485	radio	0.610	rivers	0.505
dish	0.461	broadcast	0.569	navigable	0.409
cooked	0.457	broadcasting	0.540	flows	0.380
90 軸		**120 軸**		**150 軸**	
comics	0.669	winters	0.510	nuclear	0.619
marvel	0.600	summers	0.503	bomb	0.505
superhero	0.571	temperatures	0.472	bombs	0.426
superman	0.560	precipitation	0.471	fission	0.422
dc	0.540	humidity	0.470	plutonium	0.419

3　背景：独立成分間の依存関係

現実のデータに ICA を適用すると，得られる成分同士は互いに無相関であるが独立ではなく，成分間には依存関係があることが知られている [9, 10, 11]．特に ICA では捉えることができない非線形な依存関係が，これらの成分間に存在する．

3.1　依存関係を定量化する尺度

単語ベクトルを ICA で変換し，そこから得られる独立成分 $\mathbf{s}_i$ と $\mathbf{s}_j$ 間の依存関係を定量化する方法を検討する．この定量化には，相互情報量や HSIC [14] のほか，3 次以上の混合モーメント $\mathrm{E}(\mathbf{s}_i^k \mathbf{s}_j^l)$ $(k+l \geq 3)$ を使用することができる．本稿では，4 次の混合モーメントを用いて依存関係を定量化する．

$$\mathrm{E}(\mathbf{s}_i^2 \mathbf{s}_j^2) = \frac{1}{n} \sum_{t=1}^{n} S_{t,i}^2 S_{t,j}^2 \tag{1}$$

$\mathbf{S}$ は白色化されており，$\mathbf{s}_i$ と $\mathbf{s}_j$ が互いに独立であれば，$\mathrm{E}(\mathbf{s}_i^2 \mathbf{s}_j^2) = 1$ である．$\mathrm{E}(\mathbf{s}_i^2 \mathbf{s}_j^2)$ の値が 1 から離れるほど，$\mathbf{s}_i$ と $\mathbf{s}_j$ は依存関係が大きいと言える．

1)　座標軸は分布の歪度が正になる向きにとり，歪度の絶対値の順に軸番号を定めた．

2)　正規化したベクトルの各成分値は，そのベクトルと各座標軸ベクトルとのコサイン類似度を表す．各座標軸において，正規化した単語ベクトルの成分値が大きなものを選ぶことは，座標軸との向きが近いベクトルを選ぶことに対応する．

3)　ただし，text8 コーパスにおいて 100 回以上出現する単語の中から選んだ．以後，特に断りがなければ図表で示す単語は全て text8 コーパスにおいて 100 回以上出現する単語の中から選んだものである．

表 2　(上段) $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が大きい独立成分のペア．表の中で重複が現れないように上位 6 ペアを示した．(下段) $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が小さい独立成分のペア．上段におけるペア $(\mathbf{s}_i, \mathbf{s}_j)$ の軸番号が小さい成分 $\mathbf{s}_i$ に対して，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_k^2)$ の値が小さい成分 $\mathbf{s}_k$ を表の中で重複が現れないように選んだ．表の単語はそれぞれの成分値が大きな上位 5 単語である．

$\mathrm{E}(\mathbf{s}_{22}^2\mathbf{s}_{59}^2) = 2.247$		$\mathrm{E}(\mathbf{s}_{14}^2\mathbf{s}_{113}^2) = 2.380$		$\mathrm{E}(\mathbf{s}_{70}^2\mathbf{s}_{58}^2) = 2.395$		$\mathrm{E}(\mathbf{s}_{10}^2\mathbf{s}_{125}^2) = 2.431$		$\mathrm{E}(\mathbf{s}_{2}^2\mathbf{s}_{114}^2) = 2.480$		$\mathrm{E}(\mathbf{s}_{63}^2\mathbf{s}_{210}^2) = 2.964$	
22 軸	59 軸	14 軸	113 軸	70 軸	58 軸	10 軸	125 軸	2 軸	114 軸	63 軸	210 軸
instrument	concerto	topological	frac	accusative	imperfect	dna	algae	acid	morphisms	organization	unesco
instruments	fugue	isomorphic	cos	nouns	perfect	proteins	bacteria	hydrogen	homomorphism	international	itu
bass	sonata	banach	equation	genitive	future	rna	fungi	acids	hydrogen	organizations	interpol
guitars	bwv	topology	euler	noun	present	mrna	mitochondria	oh	wavelengths	interpol	observer
tuning	beethoven	isomorphism	infty	adjectives	past	protein	organisms	ch	metadata	standardization	temporary
$\mathrm{E}(\mathbf{s}_{22}^2\mathbf{s}_{1}^2) = 1.065$		$\mathrm{E}(\mathbf{s}_{14}^2\mathbf{s}_{57}^2) = 1.017$		$\mathrm{E}(\mathbf{s}_{70}^2\mathbf{s}_{197}^2) = 0.940$		$\mathrm{E}(\mathbf{s}_{10}^2\mathbf{s}_{18}^2) = 0.980$		$\mathrm{E}(\mathbf{s}_{2}^2\mathbf{s}_{3}^2) = 0.993$		$\mathrm{E}(\mathbf{s}_{63}^2\mathbf{s}_{76}^2) = 1.109$	
22 軸	1 軸	14 軸	57 軸	70 軸	197 軸	10 軸	18 軸	2 軸	3 軸	63 軸	76 軸
instrument	genus	topological	s	accusative	population	dna	actress	acid	al	organization	you
instruments	species	isomorphic	and	nouns	median	proteins	footballer	hydrogen	ibn	international	know
bass	extinct	banach	was	genitive	estimated	rna	musician	acids	muhammad	organizations	me
guitars	birds	topology	in	noun	residing	mrna	actor	oh	abu	interpol	we
tuning	subspecies	isomorphism	by	adjectives	total	protein	singer	ch	qaeda	standardization	want

表 3　6 つの成分ペア $(\mathbf{s}_i, \mathbf{s}_j)$ について，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値への寄与 $S_{t,i}^2 S_{t,j}^2$ が大きな上位 6 単語とその値を示す．表に示す 6 つの成分ペアは 4.2 節で定義している最小全域木 T 上から選んだ．

$\mathrm{E}(s_{63}^2 s_{57}^2) = 2.158$		$\mathrm{E}(s_{16}^2 s_{118}^2) = 2.124$		$\mathrm{E}(s_{16}^2 s_{10}^2) = 1.947$		$\mathrm{E}(s_{13}^2 s_{168}^2) = 1.991$		$\mathrm{E}(s_{13}^2 s_{126}^2) = 1.735$		$\mathrm{E}(s_{13}^2 s_{73}^2) = 1.740$	
16 軸	52 軸	16 軸	118 軸	16 軸	10 軸	13 軸	168 軸	13 軸	126 軸	13 軸	73 軸
blood	infectious	blood	disorder	blood	dna	windows	license	windows	pointer	windows	ip
organs	infection	organs	mental	organs	proteins	os	copyleft	os	return	os	tcp
liver	disease	liver	disorders	liver	rna	unix	gpl	unix	string	unix	protocols
kidney	infections	kidney	symptoms	kidney	mrna	linux	licenses	linux	pointers	linux	protocol
w_t	$S_{t,16}^2 S_{t,52}^2$	w_t	$S_{t,16}^2 S_{t,118}^2$	w_t	$S_{t,16}^2 S_{t,10}^2$	w_t	$S_{t,13}^2 S_{t,168}^2$	w_t	$S_{t,13}^2 S_{t,126}^2$	w_t	$S_{t,13}^2 S_{t,73}^2$
abscess	1932.6	atrophy	2110.2	adenylate	2079.8	qpl	5678.2	bytecode	2107.1	netware	1799.2
multifocal	1440.2	hemiparesis	1877.5	effectors	1842.5	lgpl	4519.9	macros	1167.9	netbios	1543.2
hemorrhagic	1239.3	axonal	1465.9	antisense	1639.9	trolltech	3588.4	userland	1142.9	imap	1414.0
esophagitis	1187.4	dysfunction	1380.2	cyclase	1638.9	gpl	3325.2	gnumeric	1011.9	glut	1239.0
efferent	1160.5	neuropathy	1300.1	myosin	1201.8	gnu	2826.1	preprocessor	918.8	wfw	1179.9
mitral	1143.5	myopathy	1288.3	axons	1144.2	bsd	2822.7	overwriting	818.2	dhcpv	1115.5

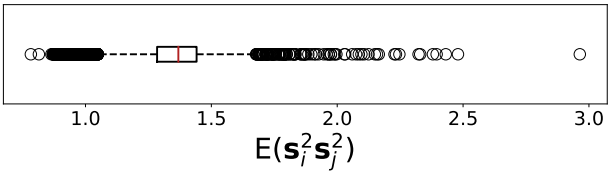

図 2　$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値の箱ひげ図．$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ は対称なので，$i < j$ である 44,850 個の値を使用した．

4　実験：依存関係の解釈と可視化

4.1　依存関係が大きいと意味が関連する

単語ベクトルを ICA で変換して得られる独立成分のペア $(\mathbf{s}_i, \mathbf{s}_j)$ について，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値を計算した．この値に関する詳細は図 2 の箱ひげ図に示されている．この図から，ペアごとに $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が大きいものと小さいものがあることがわかる．

$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が大きい成分のペアと小さい成分のペアをそれぞれ具体的に分析し，その特徴を明らかにする．

結果と考察　表 2 に独立成分ペア $(\mathbf{s}_i, \mathbf{s}_j)$ の $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値と，それぞれの成分値が大きな単語を示す．$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が大きな独立成分ペアはそれぞれの成分が持つ意味が関連していることがわかる．逆に，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が小さいと意味が関連していないこともわかる．

具体例　$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が特に大きいペアとして第 22 軸と第 59 軸のペアに注目した．22 軸は「楽器」，59 軸は「オーケストラ」の意味でそれぞれ解釈できる．これにより，両者には意味の関連性があることがわかる．一方で，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値が小さいペアである第 22 軸と第 1 軸では，第 22 軸が「楽器」，第 1 軸が「食べ物」を意味し，これらの間には明確な意味の関連性が見られないことがわかる．

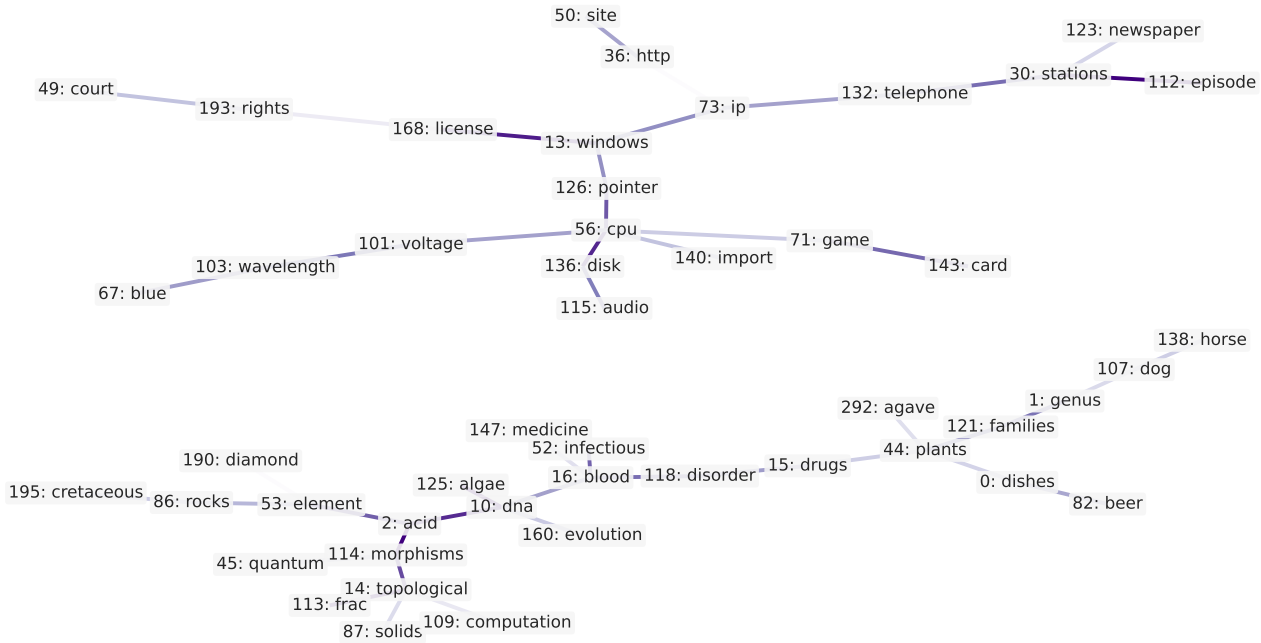

図 3　4.2 節で定義している最小全域木 T における，連結な部分木を 2 つ可視化した．各頂点はそれぞれ ICA で得られる単語ベクトルの独立成分を表す．頂点のラベルには頂点番号を記し，その隣に成分値が最も大きな単語を記した．頂点 (i, j) 間の枝の色は成分間の $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値の大きさを表しており，枝の色が濃いほど大きい．

4.2　依存関係のネットワークを可視化

独立成分 $\mathbf{s}_i$ $(i \in \{0, \cdots, d-1\})$ を頂点とする重みつき完全グラフ G を考える．頂点対 $(\mathbf{s}_i, \mathbf{s}_j)$ の間の辺の重みを $e^{-\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)}$ とする[4]．このグラフ G に対して最小全域木 T を計算した[5]．

結果と考察　図 3 に最小全域木 T における 2 つの部分木を示す．この図を通して 3 つ以上の独立成分の意味のつながりやまとまりといった構造を読み取ることが可能である．

具体例　例えば，*126: pointer, 56: cpu* や *2: acid, 10: dna* といった意味的に関連する成分ペアは T 内の辺で結ばれている．また，*49: court, 193: rights, 168: license* や *147: medicine, 52: infections, 16: blood* といったグループは，辺によって結合された頂点集合として，ひとつの意味のまとまりを形成していることがわかる．

4.3　意味の関連性を具体化

独立成分のペア $(\mathbf{s}_i, \mathbf{s}_j)$ において，$S_{t,i}^2 S_{t,j}^2$ が大きな単語 w_t は $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2) = \frac{1}{n}\sum_{t=1}^{n} S_{t,i}^2 S_{t,j}^2$ に対する寄与が大きい．この節では独立成分ペアの $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値への寄与が大きな単語を調べる．

結果と考察　表 3 に示すのは，4.2 節で定義した最小全域木 T から選んだ成分ペアと寄与が大きな単語例である．$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値への寄与が大きな単語 w_k は両方の成分に関連する意味を持ち，成分ペアの関連性がより具体的になることがわかる．

具体例　例として，第 16 軸と第 52 軸のペアにおいて $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の値への寄与が顕著な単語を挙げると，*abscess* (膿瘍)，*multifocal* (多焦点の)，*hemorrhagic* (出血)，*esophagitis* (食道炎) などが含まれる．これらの単語は第 16 軸が表す「人体」と第 52 軸が表す「感染症」の意味の関連性を具体的に反映するものである．

5　おわりに

本稿では，単語ベクトルを ICA で変換して得られる成分間の依存関係を $\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ で定量化した．実験により，依存関係の大きさが独立成分間の意味の関連性を反映することがわかった．また，成分間の依存関係への寄与が大きな単語を調べることで，成分間の関連性を具体的に解釈することができた．今後は，「意味の関連性」についての定性的な議論を拡張し，より客観的かつ定量的な尺度を用いた分析方法を探求することを目指す．

4)　このグラフ G で最小全域木 T を計算することは，$\mathrm{E}(\mathbf{s}_i^2\mathbf{s}_j^2)$ の総和を最大化する全域木を求めることに相当する．

5)　G の部分グラフとして様々なものが考えられるが，残す枝の数と結果の解釈性のバランスを考慮し，各成分ごとに相対的な重要度の高い枝が選ばれやすい最小全域木を選択した．

謝辞

本研究は JSPS 科研費 22H05106, 23H03355 および JST CREST JPMJCR21N3 の助成を受けたものです．

参考文献

[1] Tomás Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In **Advances in Neural Information Processing Systems**, pp. 3111–3119, 2013.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. **arXiv preprint arXiv:1907.11692**, 2019.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 1877–1901. Curran Associates, Inc., 2020.

[5] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. **arXiv preprint arXiv:2307.09288**, 2023.

[6] Aapo Hyvärinen and Erkki Oja. Independent component analysis: Algorithms and applications. **Neural networks**, Vol. 13, No. 4-5, pp. 411–430, 2000.

[7] Tomáš Musil and David Mareček. Independent components of word embeddings represent semantic features. **arXiv:2212.09580**, 2022.

[8] Hiroaki Yamagiwa, Momose Oyama, and Hidetoshi Shimodaira. Discovering universal geometry in embeddings with ICA. In **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 4647–4675, Singapore, December 2023. Association for Computational Linguistics.

[9] Aapo Hyvärinen, Patrik O. Hoyer, and Mika Inki. Topographic independent component analysis. **Neural Computation**, 2001.

[10] Hiroaki Sasaki, Michael Gutmann, Hayaru Shouno, and Aapo Hyvarinen. Correlated topographic analysis: estimating an ordering of correlated components. **Machine Learning**, 2013.

[11] Hiroaki Sasaki, Michael Gutmann, Hayaru Shouno, and Aapo Hyvarinen. Estimating Dependency Structures for non-Gaussian Components with Linear and Energy Correlations. In Samuel Kaski and Jukka Corander, editors, **Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics**, Vol. 33 of **Proceedings of Machine Learning Research**, pp. 868–876, Reykjavik, Iceland, 22–25 Apr 2014. PMLR.

[12] Matt Mahoney. About the test data, 2011. http://mattmahoney.net/dc/textdata.html.

[13] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. **IEEE Transactions on Neural Networks**, Vol. 10, No. 3, pp. 626–634, 1999.

[14] Arthur Gretton, Olivier Bousquet, Alex Smola, and Schölkopf Bernhard. Measuring statistical dependence with hilbert-schmidt norms. In **The International Conference on Algorithmic Learning Theory**, 2005.