Vol.2019-MPS-122 No.10
2019/2/28

深層学習による意味ベクトル空間の構築と意味性認知症の病態シミュレーション

出田 達大[1,a)]　小濱 剛[1,b)]

概要：本研究では，意味性認知症（SD）における病態理解の手がかりを得ることを目的として，深層学習を用いて，視覚情報と言語情報を統合して意味関連タスクを遂行する意味ベクトル空間学習モデルを提案する．SD では，高次感覚情報の統合に関与するとされる側頭極に萎縮が認められ，単語の理解や物品の呼称などの意味的判断に障害が現れることから，SD の諸症状は，感覚情報の統合機能の損傷により発症する可能性が考えられる．提案モデルは，深層学習により視覚情報と言語情報を意味情報へとマッピングするものであり，意味関連タスクを遂行させることで，構成論の立場から，側頭極の機能のモデル化とその動作検証を行った．提案モデルにより，認知症の診断に用いられる認知タスクを学習させ，健常者と同程度のパフォーマンスを確保した上で，意味ベクトル空間の破壊実験を行った結果，提案モデルにおける意味ベクトル空間破壊による認知パフォーマンスの低下は，実際の SD 患者の振るまいとよく一致した．これらのことから，提案モデルは SD の発症機序に対する 1 つの説明を与えるものであることが示唆された．

キーワード：意味性認知症，側頭極，深層学習，意味ベクトル学習，認知タスク．

A simulation study of pathophysiology in semantic dementia based on a semantic vector space learning model using deep learning techniques

Abstract: In order to obtain a basic pathophysiological knowledge of semantic dementia (SD) , we proposed a semantic vector space learning model which integrates visual and language information and performs semantic related tasks using deep learning techniques. Previous studies have shown that in SD patients, there is brain atrophy in the temporal pole which is considered to be involved in the integration of higher-order sensory information, and they consequently show impairments in semantic judgment such as word recognition and object naming. These indicate that the SD symptoms may be caused by damage in the integrative functions of sensory information. The proposed model maps visual and language information to semantic information using deep learning techniques, and we verified the performances of the modeled functions of the temporal pole by executing meaning related tasks from the point of the constructive approach. After learning the cognitive tasks used for diagnosing cognitive dementia, we conducted destructive experiments of semantic vector space in the proposed model. The simulation results show that the decline in task performances are well correlated with the behavior of actual SD patients. This suggests that the proposed model gives one explanation for the pathogenesis of SD.

Keywords: Semantic dementia, Temporal pole, Deep learning, Semantic vector space, Cognitive task.

1. はじめに

2015 年 7 月に，前頭側頭葉変性症の一形態として，国の難病に指定された意味性認知症 (Semantic Dementia: SD) は [1][2][3] ，語義失語，物品の使用法や名称等を同定す能力の低下，表層性失語・失書，相貌失認など，特定の物事の意味や概念の忘却・喪失から派生する様々な症状を呈し，患者の社会生活の遂行を著しく妨げることが知られている．

脳イメージング研究の成果から，SD 患者には側頭葉前方部に位置する側頭極（Temporal Pole: TP）の萎縮や損

[1] 近畿大学大学院生物理工学研究科
Graduate School of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama 649-6493, Japan
[a)] 1733730012a@waka.kindai.ac.jp
[b)] kohama@waka.kindai.ac.jp

傷が見られることが明らかとなっているが [4]，その発症機序や TP の機能的な役割は未だ解明されておらず，有効な治療法やリハビリ方法も確立されていない．そのため，TP における認知的処理過程，および SD 発症機序を詳らかにし，それらに基づいた有効な治療法の確立が望まれる．しかしながら，fMRI に代表される非侵襲計測に基づいた分析では，大局的な脳活動しか知ることができないなどの欠点がある．一方，脳細胞の応答を直接計測する電気生理学的手法では，詳細な脳活動が計測可能ではあるものの，侵襲性が高く，被験者へ極度の負担を強いる上，得られるデータが過度に局所的であるために，TP の全体像の把握が困難である．さらに，SD 患者の絶対数が少ないため，実際の患者を対象とした実験を実施すること自体が容易ではないという問題もある．これらの理由から，SD 発症メカニズムを含めた TP の認知処理過程の解明に対しては，計算機を用いたモデルシミュレーションによるアプローチが有効な手段であると言える．

現在，TP の機能的役割における仮説として有力視されているのは，TP が様々なモダリティ情報のハブとして機能しているとする説である．図 1 に示すように，TP は視覚野，言語野などの各連合野との連絡を持っていることが実験的に示されている [4]．このことから，視覚情報，言語情報などの高次感覚情報が TP において関連付けて統合され，モダリティに依拠しない包括的な『意味』の情報が形成された後，種々の認知的・身体的課題の遂行に利用されていると考えられる．SD 患者が語義失語などの症状を呈する一方で，復唱や文法，計算能力，発話の流暢性など，物事の意味には依拠することなく，規則に従って遂行する能力は保たれているという事実に鑑みても，非常に有望な説であると言える．この仮説が正しいとするなら，ヒトの脳の情報処理方式に則って様々なモダリティ情報を混合し，得られた情報に基づいて何らかのタスクを遂行するような計算処理を実現した場合，各モダリティ情報を元に得られた情報は，TP における意味情報表現を近似的に再現しているはずである．

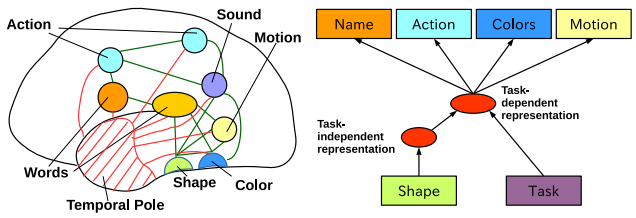

図 1 大脳皮質における意味表現の仮説 [4]
Fig.1: A hypothesis of semantic expression in the cerebral cortex[4].

そこで本研究では，前述の仮説に対する実験的証拠を示すことにより，SD の発症機序を解明するための足がかりとなることを期待して，様々なモダリティ情報を統合して意味情報を表現するような，深層学習による意味ベクトル空間学習モデルを構築し，意味関連タスクのスコアを評価することで，提案モデルが TP のモデルとして適当であるかについて検討を行い，提案モデルの振る舞いから，SD の発生機序，および，TP が有する機能的役割について考察を行う．

2. 提案モデルの具体化

Tsapkini ら [5] により，SD の病状悪化に伴う意味関連タスクスコアの推移が報告されている．Tsapkini らは，萎縮や損傷によって失われた TP の体積と，Naming Task（NT）および Comprehension Task （CT）という，2 種類の意味関連タスクのスコアとの関係について調べ，萎縮や損傷の程度が大きくなるほど，意味関連タスクパフォーマンスも低下することを示した．NT とは，提示された物体の名前を回答させるタスクであり，CT は，予め示された 4 つの物体の中から，言葉で指示された物体を被験者が選択するというタスクである．本研究では，各モダリティ情報を統合して得られた意味情報に基づいて，NT と CT を遂行することが可能な，マルチタスクおよびマルチモーダルな意味ベクトル空間学習モデルを提案し，これら 2 つのタスクの正答率を，提案モデルにおける SD の再現度を評価するための指標として用いる．

NT および CT の遂行には，言語情報と視覚情報を関連付けて統合する必要があるが，視覚情報と言語情報に基づいて所望の解答を得るようなモデルの先行事例として，Dynamic Memory Network plus（DMN+）[10] がある．これは，画像あるいは文字列から抽出した情報を事実として記憶に保持し，文字列による質問を受けて記憶から情報を引き出し，質問の内容に回答するという深層学習モデルである．以下の図 2 に DMN+ の概要を示す．

TP における高次感覚情報に対する統合過程のモデル化には，脳の情報処理機構との類似性，タスクパフォーマンスの高さ，異なるモダリティ間の統合の容易さなどの理由から，DMN+ のような深層学習モデルが適していると考えられる．また，TP は視覚や言語以外の情報も受け取っているが，SD の症状の中でも特に顕著なものが，語義失語や視覚対象に対する同定能力の減退であること，NT と

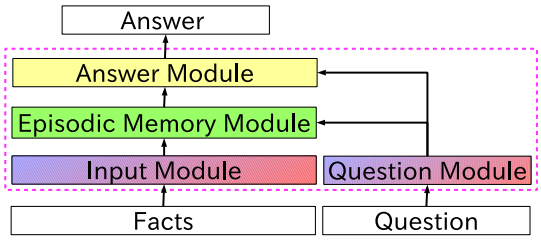

図 2 Dynamic Memory Network+の概要
Fig.2: Schematic diagram of Dynamic Memory Network+

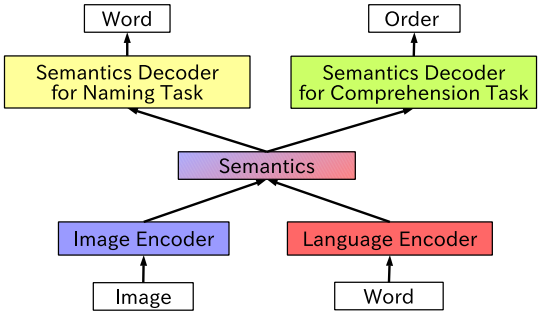

図 3 提案モデル概略
Fig.3: Schematic diagram of proposed model.

CT の遂行には視覚情報と言語情報のみで十分であることから，本研究では視覚情報と言語情報による TP 再現モデルの構築を試みる．

図 3 に提案モデルの概要を示す．基本的なコンセプトとしては，Image Encoder および Language Encoder において抽出された視覚および言語特徴に対して，共通の重み行列を用いて TP における意味情報と対応した意味ベクトル空間（Semantics）への写像を求め，得られた意味ベクトルの情報に基づいて各タスクを遂行させるというものである．

Image Encoder，Language Encoder，およびそれらからの出力を統合して形成された Semantics 層の出力を，指定された意味関連タスクに対する出力へと変換する Semantics Decoder は，任意の深層学習モデルであり，性能向上を目的として自由に差し替えてもかまわない．本研究では，たとえば CT における Semantice Decoder として，先述した DMN+ を用いる．図 3 中の Semantics が図 1 中の Task independent representation 部に，Semantics Decoder が Task dependent representation 部に相当する．Image Encoder への入力は画像情報であり，Language Encoder への入力は，Word2Vec[14] や Seqence to Sequence[16] 等によって得られた言語ベクトルである．

3. シミュレーション実験

3.1 意味関連タスクの学習とパフォーマンスの評価

提案モデルの学習には，猫や船などの計 10 カテゴリで構成される CIFAR-10[17] に対して，Krizhevsky らが選別を行った全 6 万枚のカラー画像を入力として用いた．図 4 に CIFAR-10 のサンプル画像を示す．

このうち 5 万枚を学習用，残りの 1 万枚を評価用とした．入力として用いる言語情報は，文章ではなく単語に限定した．入力単語には，CIFAR-10[17] の画像に対するラベル文字列を使用した．単語の埋め込みに用いた Word2Vec[14] は，画像と言語間の意味的な関連性を記述したメタデータを含んだ大規模画像・言語データセットである Visual Genome[18] が提供する文章データを用い，埋め込み先の次元を 100 次元として学習させた．

Image Encoder には ResNet[19][20]，Language Encoder および Naming Task 用の Semantics Decoder には多層パーセプトロン，Comprehension Task 用の Semantics Decoder には DMN+[10] を用いた．実験に使用したモデルの構造の詳細を図 5 に示す．意味ベクトル空間を形成する Semantics 層は，Image Encoder が出力する視覚特徴と，Language Encoder が出力する言語特徴のそれぞれに対して，共通の重み行列を持つ多層のパーセプトロンであり（図 6），両者の特徴量を関連付けて統合する．

各ブロック中に表記している数値は，それぞれの次元数を表している．また，Image の次元である $3 \times 32 \times 32$ とは，高さ 32 × 幅 32 の 3 チャネル画像を意味する．NT では，単一の画像を入力とし，その画像のラベルとして相応しい 100 次元の単語ベクトルを出力する．CT では，

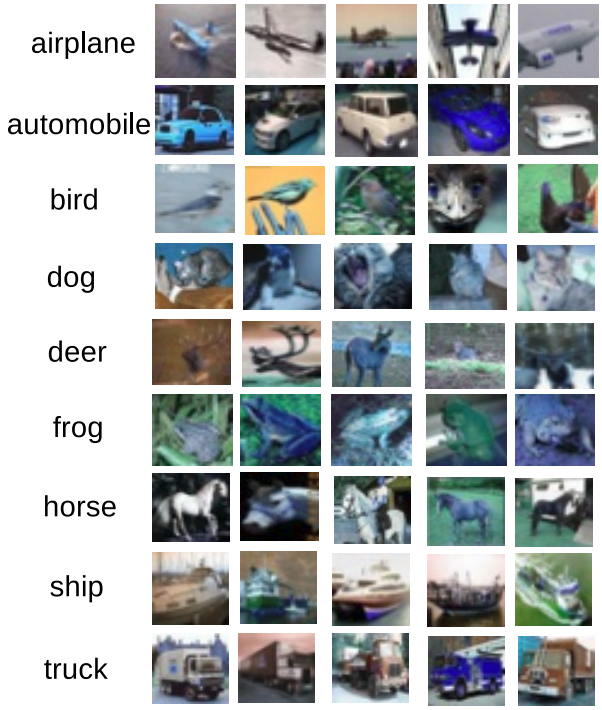

図 4 CIFAR-10 のサンプル画像
Fig.4: Example images of CIFAR-10.

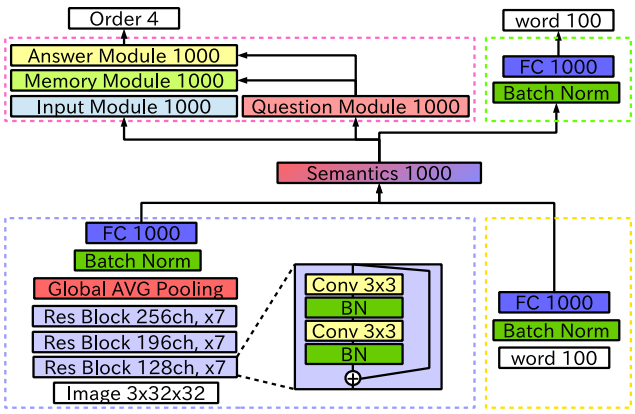

図 5 提案モデルの具体的構造
Fig.5: Detailed structure of proposed model.

表 1　学習に用いたパラメータ
Table1: Parameters used for learning.

Activation Function	ReLU
Learning Rate	0.01
Momentum Coefficient	0.9
L2 Regularization Coefficient	0.0002
Mini-Batch Size	100

4 枚の画像を順次入力し，100 次元の埋め込みベクトルをクエリとして，クエリで指定された物体の画像が何番目に入力されたかを出力する．この時，CT 用の Semantic Decoder として用いられている DMN+ の IM を構成する Bidirectional RNN には，各時刻ごとに，1 枚の画像を表象した意味ベクトルが入力されることとなる．QM には，クエリとなる単語ベクトルを Semantics 層に通して得られた意味ベクトルが入力される．DMN+[10] の Episodic Memory Module のホップ数は 2 とし，重みの初期化には Xavier の初期化 [11] を用い，それ以外のパラメータは正規乱数で初期化した．表 1 には，実験に用いたハイパーパラメータなどの設定を示した．なお，モデルの構築には，Google の提供する TensorFlow[12] というフレームワークを用いた．

図 7 に，確率的勾配降下法により 50 万ステップの学習を行った際の誤差の推移を示す．train loss が学習用データに対する誤差であり，eval loss は学習用データとは異なる評価用データに対する誤差を示している．ただし，eval loss は 1000 ステップの学習ごとに，全 1 万サンプルの中から無作為に抽出した 100 サンプルを用いて算出した誤差の値を示している．学習済みモデルに対する，NT の入出力例を図 8 に，CT の入出力例を図 9 に示す．

図 10 に NT における精度の推移を，図 11 には CT における精度の推移を示した．いずれも，評価用データを対象として，1000 ステップの荷重更新毎に 1 回，全 1 万サンプルの中から無作為に抽出した 100 サンプルを用いて，正答率を算出した結果である．なお，Top K Accuracy とは，ネットワーク出力と単語ベクトルとのコサイン類似度の順位を求め，上位 K 個の中に入力画像に対する正解ラベルが含まれていれば正解であるとみなした時の正答率を表している．図 8 の例では，中央の囲みが出力ベクトルの一部を表しており，これとコサイン類似度が高い単語のうち上位 5 つが右の囲みに示されている．この例では，Top 1 Accuracy，Top 5 Accuracy のいずれにおいても正解とみなされる．

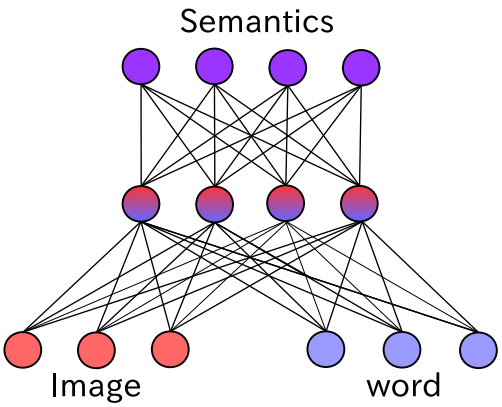

図 6　Semantics 層の具体的構造
Fig.6: Detailed structure of semantic layer.

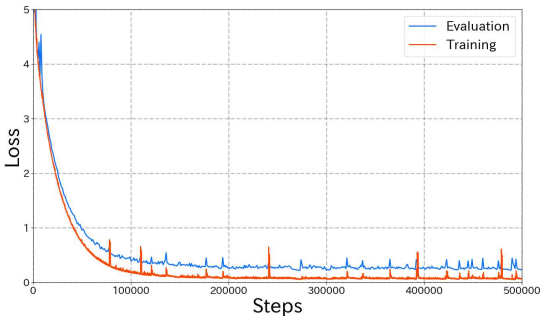

図 7　誤差の推移
Fig.7: Training/Evaluation loss transition.

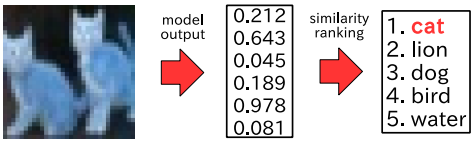

図 8　Naming Task における入出力例
Fig.8: An example of I/O in Naming Task.

図 9　Comprehension Task における入出力例
Fig.9: An example of I/O in Comprehension Task.

図 7 から，train loss と eval loss のいずれにおいても，ほぼ単調減少しており，かつ，これらの誤差の推移に乖離が見られないことから，過学習の傾向は認められず，順調に学習が進行したと言える．また，図 10 および図 11 から，NT と CT のいずれにおいても精度は上昇していることから，汎化性能の獲得にも成功していることが見て取れ

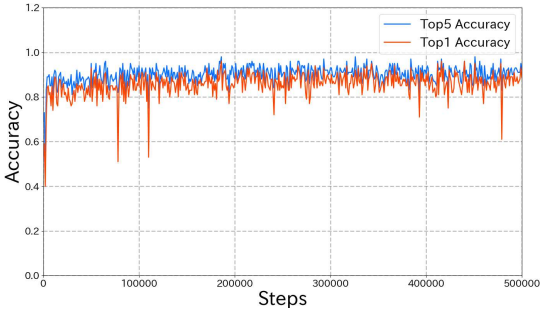

図 10　Naming Task における精度の推移
Fig.10: Accuracy transition in Naming Task.

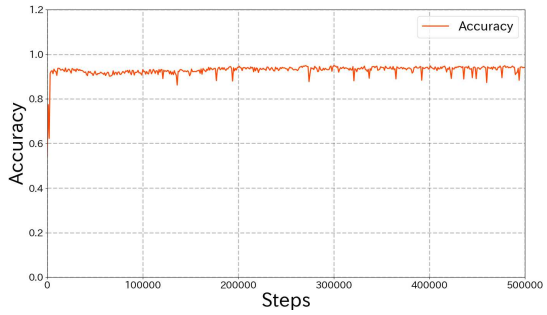

図 11 Comprehension Task における精度の推移
Fig.11: Accuracy transition in Comprehension Task.

る．学習終了後，評価データの全サンプルを用いて精度を求めたところ，NT においては Top 1 Accuracy が 88.4%，Top 5 Accuracy が 92.2%，CT においては 94.1%を記録した．両タスクにおいて 85%以上の高精度を確認できたことから，提案モデルは，健常者と同程度の水準で意味関連タスクの遂行が可能であると判断した．

3.2 意味ベクトル破壊実験

提案モデルにおいて意味ベクトル空間を構成する 1000 次元の Semantics 層において，SD 患者に見られる TP の萎縮や損傷を想定し，一定の割合だけランダムにニューロンを脱落させた際の NT および CT の精度を求めた．脱落させる割合を破壊率とし，0%から 10%刻みで系統的に設定した．ただし，100%破壊すると推論が不可能となるために，上限の程度は 99.9%とした．図 12 には，意味ベクトル空間の破壊率に対する各タスクの精度の推移を示した．

NT と CT のいずれにおいても，10%や 20%のニューロンを脱落させた程度ではほとんど精度に影響が見られなかったことから，提案モデルはロバストな意味情報を獲得していると言える．CT では意味ベクトル空間の破壊が 50%に達するまで精度が低下せず，その後急速にチャンスレベルにまで低下するのに対して，NT では 20%~ 30%破壊された段階で精度が低下しはじめ，緩やかに悪化し続けることが示された．

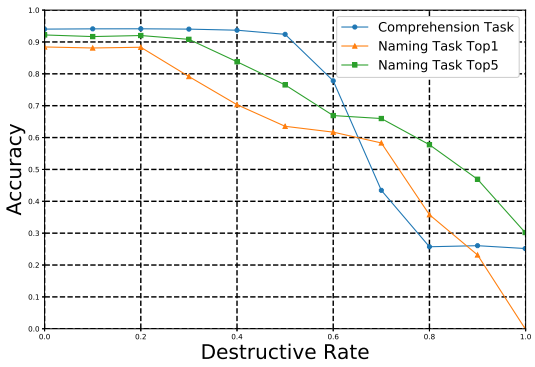

図 12 意味ベクトル空間破壊率に対する精度の推移
Fig.12: Accuracy transition as a function of semantic vector space destruction rate.

4. 考察

SD 患者の意味関連タスクに関する研究報告 [7][8][9] において，SD 患者では，NT よりも，CT やその同類のタスクの方が，重症化してから顕著にスコアの低下が見られることが示されている．また，Tsapkini ら [5] によれば，SD 患者における CT のパフォーマンスは，TP の損傷がある程度大きくなってから低下が見られることや，損傷の拡大に伴う低下の度合いは NT よりも急速であることが示されている．図 12 に示した結果は，これらの従来知見と符合するものであることから，提案モデルは生理学的な妥当性を備えているものである言え，SD の病態に関して一つの説明を与えるものであると言える．

提案モデルにおける，意味ベクトル空間の破壊に伴う NT と CT のパフォーマンスの相違には，次のような理由が考えられる．NT では，多層パーセプトロンにより意味ベクトルを単語ベクトルへとマッピングしているため，意味ベクトル崩壊の影響が直接的に出力へと現れる．一方，CT では，decoder に DMN+ を用いており，Attention 機構と GRU によって，クエリ単語と似た意味を持つ画像を強調した上で，それらが入力された順序の記憶を形成している．意味ベクトル空間を持つ Semantics 層のニューロンに脱落が生じても，その後の意味ベクトル同士の類似度に基づいて Attention が計算されるために，画像間の区別が曖昧になるほどの次元の減衰に達するまでは，画像間の Attention 重みの配分が大きく変化せず，画像の入力順も保持されていると考えられる．意味ベクトル空間次元の減退が進行し，画像と単語の類似度が算出困難となるか，選択肢として提示された画像間の意味的な違いが弁別不能となる，あるいは，その両方の要因が重なることで，CT の精度が急落すると解釈できる．

提案モデルと類似する decoder の機構が大脳皮質に存在するか否かは不明であるが，少なくとも，TP において視覚情報と言語情報が関連付けられて符号化されているのであれば，その損傷によって，SD 患者が示す意味的認知症の要因が作り出される可能性は示されたと言える．

本論文で使用した CIFAR-10[17] の画像セットは，画像の解像度が低く，カテゴリも少数であるために，ヒトの TP が受け取る情報量には到底及ばない．高解像度でカテゴリ数も多く，文章ラベルも付随したデータセットである VisualGenome[18] を用いた再学習や，BERT[22] モデル等による Language Encoder の性能向上などを実現し，より複雑な意味関連タスクを再現することで，SD の病態理解につながる包括的な知見が得られるものと考えている．

5. まとめ

本研究では，SD の病態を解明するための手がかりを得ることを目的として，視覚情報と言語情報を統合すること

により意味ベクトル空間を学習して，意味関連タスクを遂行する深層学習モデルを構築し，構成論的立場から，SD 患者における認知機能の低下が生じるメカニズムについて論じた．SD 患者は，側頭葉先端に位置する TP に器質的変容が生じ，単語の理解や物品の呼称などの意味的判断に障害が現れることから，視覚情報や言語情報の統合機能に障害を受けた結果，SD の症状が表出する可能性が考えられる．提案モデルでは，深層学習を用いて視覚情報と言語情報とを意味情報へとマッピングし，意味関連タスクを遂行させることで，TP が担う機能のモデル化とその動作検証を行った．画像情報と単語情報の組み合わせを学習させた結果，意味関連タスクとして用いた NT と CT のいずれにおいても，90%前後の正答率を示した．これを健常な状態とみなし，意味ベクトル空間の破壊実験を行った結果，NT では，破壊率が 20%を超えると，それ以降徐々に精度が低下していくのに対し，CT では 50%を超えてから 80%までの間に，急激にチャンスレベルまで精度が低下することが示された．この関係は，実際の SD 患者の振るまいとよく一致したことから，提案モデルは SD の発症機序に対する 1 つの説明を与えるものであることが示唆された．

参考文献

[1] 厚生労働省：認知症とは 認知症の基礎~ 正しい理解のために~（オンライン），入手先 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000139666.html>（参照 2019-01-30）．
[2] 日本神経学会：認知症疾患診療ガイドライン 2017（オンライン），入手先 <https://www.neurology-jp.org/guidelinem/nintisyo_2017.html>（参照 2019-01-30）．
[3] 厚生労働省：平成 27 年 7 月 1 日嗜好の指定難病（告示番号 111~ 306） 前頭側頭葉変性症（オンライン），入手先 <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000079293.html>（参照 2019-01-30）．
[4] Patterson, K., Nestor, P.J. and Rogers, T.T., : Where do you know what you know? The representation of semantic knowledge in the human brain, *Nature Reviews Neuroscience*, Vol.8, No.12, pp.976—987(2007)
[5] Tsapkini, K., Frangakis,C.E. and Hillis, A.E., : The function of the left anterior temporal pole: evidence from acute stroke and infarct volume, *Brain:Journal of Neurology*, Vol.134, pp.3094-3105(2011)
[6] Patterson, K., Ralph, M.A.L., Jefferies, E., et al., : Presemantic cognition in semantic dementia: six deficits in search of an explanation, *Journal of Cognition Neuroscience*, Vol.18, No.2, pp.169-183(2006)
[7] Jefferies, E., Patterson, K., Jones, R.W. and Ralph, M.A.L., : Comprehension of concrete and abstract words in semantic dementia, *Neuropsychology*, Vol.23, No.4, pp.492-499(2009)
[8] Woollams, A.M., Ralph, M.A.L., Plaut, D.C. and Patterson, K., : SD-squared: on the association between semantic dementia and surface dyslexia, *Psychological Review*, Vol.114, No.2, pp.316-339(2007)
[9] Rogers, T.T., Graham, K.S. and Patterson, K., : Semantic impairment disrupts perception, memory, and naming of secondary but not primary colours, *Neuropsychologia*, Vol.70, pp.296-308(2015)
[10] Xiong, C., Merity, S. and Socher, R., : Dynamic memory networks for visual and textual question answering, arXiv:1603.01417(2016)
[11] Glorot, X., Bengio, Y. :Understanding the difficulty of training deep feedforward neural networks, Aistats, volume 9, pages 249-256, 2010.
[12] Google： TensorFlow, https://www.tensorflow.org/ 入手先 <https://www.tensorflow.org/>(参照 2019-01-31).
[13] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., : Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv:1412.3555(2014)
[14] Mikolov, T., Chen, K., Corrado, G. and Dean, J., : Efficient estimation of word representations in vector space, arXiv:1301.3781(2013)
[15] Schuster, M. and Paliwal, K.K., : Bidirectional recurrent neural networks, *IEEE Transaction on Signal Processing*, Vol.45, No.11, pp.2673-2681(1997)
[16] Sutskever, I., Vinyals, O. and Le, Q.V., : Sequence to sequence learning with neural networks, *Advances in Neural Information Processing Systems*, Vol.27, pp.3104-3112(2014)
[17] Krizhevsky, A., Nair, V., and Hinton, G., CIFAR-10 and CIFAR-100 datasets(online), available from <https://www.cs.toronto.edu/~kriz/cifar.html>(accessed 2019-01-30)
[18] Krishna, R., Zhu, Y., Groth, O., et al., : Visual genome: connecting language and vision using crowdsourced dense image annotations, arxiv.org/abs/1602.07332(2016)
[19] He. K., Zhang, D., Ren, S. and Sun, J., : Deep residual learning for image recognition, arXiv:1512.03385(2015)
[20] Zagoruyko, S. and Komodakis, N., : Wide residual networks, arXiv:1605.07146(2016)
[21] Hoffman, R.E., Grasemann. U., Gueorguieva, R., el al., : Using computational patients to evaluate illness mechanisms in schizophrenia, *Biol Psychiatry*, Vol.69, No.10, pp.997-1005(2011)
[22] Devlin, J., Chang, M., Lee, K. and Toutanova, K., : BERT: pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805(2018)