5P-06

入力表現の適応的選択を伴うグラフ畳み込みネットワーク学習

菊地　翔馬[†]　瀧川　一学[†‡§]

[†] 北海道大学大学院情報科学研究科　[‡] 北海道大学化学反応創成研究拠点　[§]JST さきがけ

1　はじめに

グラフは，化合物データ，分子間相互作用ネットワーク，XML 文節データなど様々な知識処理に応用されるデータ構造である．各々の対象がこのようなグラフで表現されたデータに対する回帰・分類問題は，グラフ G とそれに対応する関連値・クラス y がラベル付けされたデータを用いた教師付き学習であり，特に近年，生命科学や物質科学において化合物の物性・活性をデータ駆動方式で予測する手法として注目されている．

教師付き学習は通常は固定次元の多変量ベクトルに対して定義されるため，グラフに対する拡張は自明ではない．近年では，この表現学習の手法として，グラフ構造に対するニューラルネットワーク (Graph Neural Network, GNN) が盛んに研究されている．GNN は，グラフのノードとエッジに対して状態ベクトル (多次元潜在変数) を考え，非線形変換を施す．ノード毎に隣接するノードやエッジ間の情報や関係性の集約や更新を繰り返し，最終的に全ノードの状態ベクトルを集約したものをそのグラフの特徴ベクトルとする [4]．Duvenaud らは，グラフの各ノードに状態ベクトルを付与しグラフ畳み込み演算を適用する手法を提案し，化合物データを用いた実験により手法の有効性を示した [2]．この研究では，化合物の原子をノードとみたて，その初期状態ベクトルとして各原子の化学情報を用いた入力表現が用いられている．しかし，この入力表現として様々な他の選択肢も考えられ，どのように化学的なドメイン知識をエンコードするかは予測精度を左右すると考えられる．

本研究では，[2] で提案されたモデル Neural Fingerprint (NFP) に基づき，複数の入力表現から適応的に選択するよう拡張したモデルを提案する．ノードの初期状態ベクトルとして，具体的にどのような入力表現が適しているかは，対象とするタスクやデータの性質によって異なり，背景知識が無ければ選択することが難しい．そこで，化学情報を用いた入力表現を複数用意し，予測に適切な表現をグラフごとに動的に選択する Soft Attention 層を導入したモデルを提案する．さらに，[2] で実験に用いられた 3 種類のデータで，提案手法と既存手法の予測精度を比較する実験を行う．

2　提案手法

2.1　概要

[2] の実験では用いられた入力表現の種類が 1 つであった．あらかじめ用いる入力表現を決めてしまうと，汎用性にかけてしまうことが考えられる．また， データセットによって適した入力を選ぶにはそのデータセットに対する背景知識が必要になってくる．

そこで本研究では，各々のグラフやタスクごとの背景知識を必要とせず，モデルがグラフごとに各々の入力表現の着目すべき度合いを動的に学習する手法を提案する．具体的には，入力表現ごとの注目度係数 α を計算する Soft Attention 層を追加する．この計算層は注目度係数 α を計算するニューラルネットワークであり，出力 α は入力されたグラフに応じて各入力表現がどのくらい重要であるかを表現する．このような入力に応じた適応的重み学習の導入は，ニューラルネットワークの構造設計において，様々な事例で有効性が示されているデザインパターンである．

2.2　Soft Attention 層の導入と学習

i 番目の入力表現 R_i を用いて NFP で計算されたグラフ G の固定長ベクトルを $\boldsymbol{f}_i(G) \in \mathbb{R}^F$ とする．F は $\boldsymbol{f}_i(G)$ の次元を表し，モデルのハイパーパラメータの一つである．Soft Attention 層への入力は固定長ベクトルの集合 $\{\boldsymbol{f}_1(G), \boldsymbol{f}_2(G), \ldots, \boldsymbol{f}_K(G)\}$ である (K は入力表現の種類数)．Soft Attention 層では，各 $\boldsymbol{f}_i(G)$ に対して，独立したニューラルネットワーク ($NN_i : \mathbb{R}^F \to \mathbb{R}^F$) を用意し，式 (1) で $\boldsymbol{e}_i \in \mathbb{R}^F$ を計算する．計算された K 個の $\boldsymbol{e}_i$ の各 j 番目の要素 $\boldsymbol{e}_{,j} = (e_{1j}, e_{2j}, \ldots, e_{Kj})'$ を softmax 関数に通し，式 (2) で正規化された値 $\alpha_{ij} \in \mathbb{R}$ を得る．これを固定長ベクトル $\boldsymbol{f}_i$ または入力表現 R_i の注目度係数とする．最後に，Soft Attention 層の出力として，各 $\boldsymbol{f}_i(G)$ が注目度係数 $\boldsymbol{\alpha}_i \in \mathbb{R}^F$ で加重平均された固定長ベクトルを式 (3) で総和をとったものが $\boldsymbol{f} \in \mathbb{R}^F$ として計算され，これをグラフ G の固定次元の多変量ベクトルへの埋め込み表現として用いる．モデル式は下記となる．

$$\boldsymbol{e}_i = NN_i(\boldsymbol{f}_i(G)) \quad (1)$$

$$\alpha_{ij} = \mathrm{softmax}(\boldsymbol{e}_{,j})_i = \exp(e_{ij}) / \textstyle\sum_{k=1}^{K} \exp(e_{kj}) \quad (2)$$

$$\boldsymbol{f} = \textstyle\sum_{i=1}^{K} \boldsymbol{\alpha}_i \circ \boldsymbol{f}_i(G) \quad (3)$$

ただし、$\circ$ は要素毎の積の演算を表す．この注目度係数を学習することで，データにとって重要な入力表現を動的に選択することができる．学習は，通常の NFP と同じく全体が自動微分可能な計算グラフとなることから，同様に Back Propagation で計算できる．

3　実験

提案手法の有効性を検証するために，既存手法との比較実験を行った．実験では、提案手法の予測性能により，Soft Attention 層の導入の有効性を検証する．

3.1　使用するデータ

今回使用したデータセットは，[2] 内で使われたものと同様の 3 つのデータセットを用いた．

- Dataset 1: 1,144 個の化合物の溶解度 [log Mol/L] のデータセット [1]
- Dataset 2: 熱帯熱マラリア原虫 $P.falciparum$ の硫化

Learning Graph Convolution Networks with Adaptive Selection of Input Representations

Shoma KIKUCHI[†], Ichigaku TAKIGAWA[†‡§]
[†]Graduate School of IST, Hokkaido University
[‡]WPI-ICReDD, Hokkaido University
[§]JST PRESTO
{kicchi-s, takigawa}@ist.hokudai.ac.jp

物耐性に対する 10,000 個の試験管内での半数効果濃度 EC_{50}[nM] のデータセット [3]

- Dataset 3: 29,978 個の有機分子の光起電力効率 [%] の密度汎関数法による計算値データセット [5]

3.2 実験設定

ここでは，ノードに与える入力表現を 2 種類準備し，原子の物理的性質を表す入力表現 1 (表 1), 化学的性質を表す入力表現 2 (表 2) を用いた．入力表現 1 は，[2] の実験でも用いられていた表現であり，入力表現 2 は化学的機能に着目した特徴を表す代表的な表現である．それぞれの入力表現に含まれる特徴を one-hot または binary 表現で表し，連結したものをノードに付与する入力表現とした．各特徴量の説明，サイズを表 1，表 2 に示した．今回は，この 2 つの入力表現からデータによって適応的に選択する実験を行った．初めに，2 つの入力表現をそれぞれ単独で用いた NFP を訓練し，予測精度を求める．次に，提案手法でモデルを訓練し，予測精度を求め，精度を比較し，提案手法の有効性を確認する．しかし，両者には入力する情報量の差があるため，公平ではない．そこで，入力表現 1,2 をさらに連結したものを用いて NFP を訓練し，同時に比較し，入力表現を重み付けして選択できているかを確認する．全てのモデルについて，エッジに付与する初期状態ベクトルの表現は統一し，各特徴量の説明，サイズは表 3 に示した．各モデルは，表 4 の通りにデータを training, validation, test 用に分割し，test データの二乗平均平方根誤差 (RMSE) を比較した．エポック数は 1000 回，バッチサイズは 128，その他の NFP，全結合層のハイパーパラメータ，最適化関数は [2] の実験と同等である．Soft Attention 層内の NN_i は全て入力層と出力層を含めて 4 層の多層ニューラルネットワークにし，ノード数は入力層から 50,100,50,50 とした．つまり，固定長ベクトルの長さ F は 50 となる．

4 結果と考察

表 5 に全ての実験結果を示す．最も RMSE の低かったものは太字で表した．全てのデータに対して，提案手法が最も良い結果となった．単独の入力表現を用いて既存手法を訓練した実験よりも，提案手法の実験の方が RMSE が低く，精度を上げることができたので，Soft Attention 層導入がの有効であったと考えられる．さらに，情報量の差をなくした 2 つの表現を連結した実験と比べても，提案手法の結果が優れているため，与える単純なドメイン知識の量以上の改善と考えられる．

5 おわりに

本研究では，入力されるデータの性質を予測するのに適した入力表現を動的に選択する Soft Attention 層の導入をした拡張モデルを提案した．既存のモデルとの予測精度の比較実験において，提案手法の予測精度が上回り，各グラフに対して動的に入力表現を選択する Soft Attention 層の導入の有効性が確認できた．一方，定量的には有効性が示されたが，各化合物について具体的にどちらの入力表現がどれくらい重要であるか，ということは未解析である．注目度係数の可視化は，予測モデルの学習結果の解釈性の理解を手助ける重要な研究であるため，他の入力表現を選択肢で増やす検討とともに，今後の課題としたい．

謝辞

本研究は JSPS 科研費 17H01783，15H05711, 17K19953 および JST さきがけの助成を受けたものである．

Feature	Description	Size
原子番号	48 種類の元素とそれ以外の元素（one-hot）	49
次数	結合している原子の個数（one-hot）	6
総水素数	結合している水素の個数（one-hot）	5
価電子	価電子の個数（one-hot）	6
ベンゼン環	ベンゼン環の有無（binary）	1
合計		67

表 1: 原子の物理的性質を表す入力表現 1

Feature	Description	Size
ドナー	ドナーであるか（binary）	1
アクセプター	あくセプターであるか（binary）	1
芳香族	芳香族であるか（binary）	1
ハロゲン	ハロゲンであるか（binary）	1
酸性	酸性であるか（binary）	1
塩基性	塩基性であるか（binary）	1
合計		6

表 2: 原子の化学的性質を表す入力表現 2

Feature	Description	Size
結合の種類	単結合，二重，三重，ベンゼン環（one-hot）	4
共役系	共役系であるか（binary）	1
環状	環状構造の一部であるか（binary）	1
合計		6

表 3: 原子間の結合の特徴

	Dataset 1	Dataset 2	Dataset 3
size	1144	10000	29978
training	700	7000	20000
validation	200	1900	6000
test	100	1000	3000

表 4: 各データの分割数

単位	Dataset 1 [1] [log Mol/L]	Dataset 2 [3] [EC_{50}] in nM	Dataset 3 [5] %
平均 [2]	4.29 ± 0.40	1.47 ± 0.07	6.40 ± 0.09
NFP+入力表現 1	1.09 ± 0.04	1.10 ± 0.03	1.89 ± 0.00
NFP+入力表現 2	1.26 ± 0.05	1.12 ± 0.02	2.89 ± 0.02
NFP+入力表現 1,2	1.14 ± 0.04	**1.09 ± 0.02**	1.87 ± 0.04
提案法+入力表現 1,2	**1.00 ± 0.13**	**1.09 ± 0.00**	**1.68 ± 0.02**

表 5: 各モデルの予測精度（RMSE）の比較

参考文献

[1] Delaney,J. S.: ESOL: Estimating Aqueous Solubility Directly from Molecular Structure, *J. Chem. Inform. Comput. Sci.*, 44(3), 1000–1005 (2004)

[2] Duvenaud, D. K., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R. P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints, in *Adv. Neural. Inf. Process. Syst (NIPS2015)*. 28, 2224–2232 (2015)

[3] Gamo, F.-J., Sanz, L. M., Vidal, J., Cozar, de C., Alvarez, E., Lavandera, J.-L., Vanderwall, D. E., Green, D. V., Kumar, V., Hasan, S., et al.: Thousands of chemical starting points for antimalarial lead identification, *Nature*, 465(7296), 305 (2010)

[4] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E.: Neural Message Passing for Quantum Chemistry, in *Proceedings of the 34th International Conference on Machine Learning (ICML2017)*, 1263–1272 (2017)

[5] Hachmann, J., Olivares-Amaya, R., Atahan-Evrenk, S., Amador-Bedolla, C., Sánchez-Carrera, R. S., Gold-Parker, A., Vogt, L., Brockway, A. M., and Aspuru-Guzik, A.: The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid, *J. Phys. Chem. Lett.*, 2(17), 2241–2251 (2011)