

マルチラベル学習における Deep Neural Network の共有構造の構築法に関する研究

1X17C016-6 石倉滉大

指導教員 後藤正幸

1. 研究背景・目的

近年、膨大な文書データが日々生成されるようになり、それらの内容が複数のカテゴリに属するケースも多い。そのため、1つの文書に複数のラベルが付与されることを想定するマルチラベル分類に関する技術の重要性が高まっている。

そのマルチラベル分類問題を解くためのアプローチの1つとして、ラベル間の関係性を反映し Deep Neural Network (以下, DNN) のネットワークを自動構築する Branched Multi-Task Networks[1] (以下, BMTN) がある。BMTN は、関係性の高いラベル (以下, 類似ラベル) で DNN の中間層を共有し、この中間層の共有構造はラベル間の類似度のクラスタリング結果に基づき決定される。しかし、このクラスタリングでは、分析者が中間層ごとのクラスタ数を事前に設定する必要がある。そのため、ラベル間の関係性を適切に表現できるクラスタ数が決定されているとは限らない。

そこで本研究では、BMTN のクラスタリングにおいて、ラベル間の関係性を適切に表現できるクラスタ数の決定アルゴリズムを提案する。具体的には、ラベル間の類似度行列における固有値が γ 未満の個数をクラスタ数とする手法をアルゴリズムとして示す。これにより関係性の低いラベル (以下, 非類似ラベル) が独立に学習されるため、類似ラベルに特化した学習が可能になる。これにより、分析者に依らない適切なクラスタ数を決定でき、推定精度の向上が期待される。最後に、提案手法を読売新聞記事データに適用し有用性を示す。

2. Branched Multi-Task Networks

BMTN は、類似ラベル同士で DNN の中間層を共有し、ネットワークを構築する手法である。

2.1. DNN の構造

BMTN では、ラベルごとに行う個別学習と、ラベル間の関係性を考慮しマルチラベル分類を行う一括学習とで異なる DNN 構造を用いる。個別学習では全ラベルを個別に学習し、各ラベルに対し同じ DNN 構造を用いる。その際、中間層の数は D 、 d 層目の隠れユニットの数は U_d 、出力層のユニット数は 1 である。一方、一括学習では全ラベルを 1 つのネットワークで学習し、中間層の数は D 、 d 層目のユニットの数はラベル間の類似度の基づくクラスタリングにより決定され、出力層のユニット数はラベル数 T である。

2.2. ラベル間の類似度の算出

BMTN では、全ラベルで個別学習した T 個の DNN へ同一の N 個の検証データを入力して得られる中間表現からラベル間の類似度を計算する。このとき、個別学習の DNN 構造は全ラベルで共通であり、 d 層の中間表現は全ラベルで U_d 次元となる。すなわち、 d 層の中間表現は全ラベルで $N \times U_d$ 行列となり、これを用いて、ラベル t の d 層における中間表現の特徴を表す特徴表現行列を得る。そして、中間層ごとの T 個の特徴表現行列から、ラベル間の類似度を得る。ラ

ベル間の類似度は以下のアルゴリズムにより算出される。

Step1) ラベル $t \in \mathcal{L} := \{1, 2, \dots, T\}$ に関して個別に DNN の学習を行い、 T 個の DNN を構築する。

Step2) Step1 で学習した T 個の DNN に同一の検証データを入力し、全ての層で中間表現 W_{dt} ($d \in \{1, 2, \dots, D\}$) を得る。

Step3) W_{dt} に対し、検証データ間でのピアソンの相関係数より、中間表現の特徴を表す特徴表現行列 S_{dt} を得る。

Step4) S_{dt} の上三角行列を S'_{dt} とする。このとき、ラベル i, j に対し、 S'_{di}, S'_{dj} 間スピアマンの順位相関係数 r_{ij}^d を計算し、 D 個のラベル相関行列 $R_d = (r_{ij}^d)$ を得る。

2.3. スペクトルクラスタリング

2.2 節で算出されたラベル相関行列 R_d をスペクトルクラスタリング法によりクラスタリングを行う。これは、 R_d を固有値分解し、固有値を小さい順から k 個選択して、 k -means 法でクラスタリングを行う手法である。スペクトルクラスタリングは、事前にクラスタ数を分析者が与える必要がある。

3. 提案手法

3.1. 着想

BMTN では、一括学習の DNN 構造を決定するクラスタリングにおいて、分析者が各層のクラスタ数を与える必要がある。このため、ラベル間の関係性を適切に表現可能なクラスタ数が決定されているとは限らない。そこで本研究では、BMTN において事前にクラスタ数を与えずに、これらを自動決定しクラスタリングを行う手法を提案する。これにより、ラベル間の関係性を考慮した適切な DNN のネットワークを構築でき、推定精度の向上が期待される。

代表的なスペクトルクラスタリングのクラスタ数決定法に、Eigengap 法 [2] と R_d の固有値が 0 の個数をクラスタ数とする方法 (以下, 固有値 0 法) がある。Eigengap 法 [2] は非類似ラベルを積極的に分割する手法で、クラスタ数が多い傾向がある。そのため、パラメータ数が増大し、過学習やネットワーク共有構造に対する解釈が困難になる可能性がある。一方、固有値 0 法は確実に類似していないラベルのみを分割する手法で、クラスタ数が少なくなる傾向がある。

3.2. 提案アルゴリズム

本研究では、3.1 節の固有値 0 法を BMTN に導入し、クラスタ数を中間層ごとに自動決定する。ここで、全ラベル相関が高い場合、 R_d は正定値行列となり、 T 個の固有値は全て正の値をとる。一方、ラベル相関が 0 である変数グループが λ 個ある場合、 R_d のランクが $T - \lambda$ に落ち、正の固有値の数が $T - \lambda$ となる。すなわち、0 となる固有値の数と厳密にラベル相関が 0 である変数グループの数は等しいことを意味する。しかし、実際には真のラベル相関が 0 であっても、有限のデータから推定したラベル相関は微小な値を持つ

てしまうため、閾値を設定し、その閾値 γ よりも小さい固有値を 0 とみなす必要がある。

R_d から適切なクラスタ数を決定するアルゴリズムを以下に示す。ここで、 $R_d = (r_{ij}^d)$ ($i, j \in \mathcal{L}$) から得られる部分正方行列を $R_d^T := (r_{ij}^d)$ ($i, j \in \mathcal{I}$, $\mathcal{I} \subseteq \mathcal{L}$) と定義し、 \mathcal{S}_d を DNN の d 層目のクラスタのラベル集合族とする。

Step1) $d = 1$, $\mathcal{S}_0 = \{\mathcal{L}\}$ とする。

Step2) $\mathcal{I} \in \mathcal{S}_{d-1}$ に対し、 $|\mathcal{I}| = 1$ の場合は $\Lambda_d^T = 1$ とする。 $|\mathcal{I}| > 1$ については、部分正方行列 R_d^T を固有値分解し、固有値 γ 未満の数を Λ_d^T とする。

Step3) $|\mathcal{I}| > 1$ となる全ての $\mathcal{I} \in \mathcal{S}_{d-1}$ に対し、クラスタ数 Λ_d^T として、 R_d^T に対してスペクトルクラスタリングを適用する。これにより、 T 個のラベルは $\Lambda_d = \sum_{\mathcal{I} \in \mathcal{S}_{d-1}} \Lambda_d^T$ 個のクラスタに分割される。

Step4) Step3 で得られた各クラスタに所属するラベル集合 $\mathcal{I} \subseteq \mathcal{L}$ を要素とする集合族を \mathcal{S}_d とする。つまり、 \mathcal{S}_d は Λ_d 個の集合をもつ。

Step5) $d < D$ ならば、 $d = d + 1$ として、Step2へ戻る。

Step6) $\Lambda = (\Lambda_1, \dots, \Lambda_D)$ を各層のクラスタ数とする。

これにより、一括学習の DNN 構造が決定され、 d 層目のユニット数は $U_d \Lambda_d$ となる。

4. 評価実験

4.1. 実験条件

実験データは 2015 年の読売新聞記事を用いる。このデータの分類カテゴリは 2 階層あり、上位階層をカテゴリ、下位階層をラベルと呼ぶ。実験データの概要を表 1 に示す。

表 1: 分析データセット概要

対象カテゴリ	文書数	ラベル数
全カテゴリ	217,697	69
政治・経済	106,079	23
経済・社会	93,610	23
政治・社会	102,965	20

各データについてランダムに選んだ 8 割を学習用、1 割を検証用、1 割をテスト用とした。評価指標はラベルが完全一致した場合の正解率とし、比較手法は全共有モデル、BMTN、Eigengap 法とする。DNN の中間層は経験的に 4 層とし、各中間層のユニット数は全カテゴリデータセットで (10000,5000,2000,300)、その他では (5000,2000,1000,300) とし、初期値は he-normal[3] で与えた。BMTN のクラスタ数は、V.Simon ら [1] に従い 1 層目は 1、それ以外は 2 とし、提案手法では $\gamma = 1$ として分析を行った。

4.2. 分析結果と考察

(1) 推定精度の検証

テストデータに対する正解率を表 2 に、各中間層のクラスタ数を表 3 に示す。

表 2 より、提案手法はどのデータに対しても比較手法よ

表 2: 提案手法と比較手法の正解率

対象カテゴリ	全共有	BMTN	Eigengap 法	提案手法
全カテゴリ	0.419	0.375	0.375	0.551
政治・経済	0.618	0.599	0.593	0.645
経済・社会	0.565	0.530	0.502	0.606
政治・社会	0.655	0.631	0.607	0.671

表 3: 各中間層のクラスタ数 $\Lambda(1,2,3,4$ 層目)

対象カテゴリ	全共有	BMTN	Eigengap 法	提案手法
全カテゴリ	1,1,1,1	1,2,3,4	39,59,59,60	1,1,1,8
政治・経済	1,1,1,1	1,2,3,5	14,16,18,18	1,1,4,23
経済・社会	1,1,1,1	1,2,4,6	8,18,18,19	1,1,4,23
政治・社会	1,1,1,1	1,2,4,5	14,18,18,18	1,1,5,20

り正解率が高いことから、推定精度が向上していることが分かる。また、表 3 より、提案手法のクラスタ数は BMTN、Eigengap 法に比べて、入力層に近い層のクラスタ数が少ない。ここから、提案手法のクラスタ数は、DNN の入力層に近い層では全ラベルに対し普遍的な特徴抽出が行われていること、及び出力層に近くなるにつれてラベルに特化した特徴抽出が行われていることを表現していると考えられる。ただし、全カテゴリデータセットでは、4 層目のクラスタ数が 8 と少ない。これは、中間層の数が 4 層では不足している可能性があると考えられる。

(2) クラスタリング結果

提案手法の経済・社会データセットにおける DNN 共有構造を図 1 に示す。出力層のノードの枠内にはラベル名を記載し、白色の枠は経済、灰色は社会に所属していることを表す。

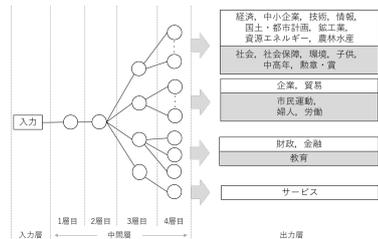


図 1: 経済・社会データにおける DNN 構造

図 1、表 3 より、4 層目のクラスタ数がラベル数と一致したことから、特筆して類似したラベルはないことが分かる。ラベル間の類似度が小さい場合でも、財政、金融、教育で 3 層目まで中間層が共有され、これらはカテゴリを跨ぎ類似ラベルであると認識されている。したがって、ラベル間の類似度が小さい場合においても、分析者の事前知識では気づかないラベル間の関係性を表現でき、推定精度を向上させたといえる。このことより、分析者の事前知識を用いてラベル間の関係性を反映したネットワークを構築することは難しく、提案手法のように構造を自動決定できる手法は有用である。

5. まとめと今後の課題

本研究では、BMTN のクラスタリング手法を改良し、中間層ごとの最適なクラスタ数を自動決定することで、推定精度を向上させる分析アルゴリズムを提案した。また、実際の記事データに提案手法を適用し、その有用性を示した。

今後の課題として、個別学習時の計算量を削減するためにラベル間の類似度算出を改良することなどが挙げられる。

参考文献

- [1] V.Simon, *et al* .“Branched Multi-Task Networks :Deciding What Layers To Share,”*arXiv*, <https://arxiv.org/abs/1904.02920>, 2019.
- [2] John, Christopher, *et al* .“Spectrum:Fast density-aware spectral clustering for single and multi-omic data,”*Bioinformatics*, Vol.36(4) ,pp.1159–1166, 2020.
- [3] HE, Kaiming, *et al* .“Delving deep into rectifiers :Surpassing human-level performance on imagenet classification,”*stering for single and multi-omic data*,”*Proc.of IEEE ICCV*, pp.1026–1034, 2015.