修士論文

マルチタスク学習時の有効な補助タスク選定

古田 朋也

2023 年 3 月 5 日

岐阜大学大学院 自然科学技術研究科 知能理工学専攻 知能情報学領域
鈴木研究室

本論文は岐阜大学大学院自然科学技術研究科に
学位（工学）授与の要件として提出した修士論文である.

古田 朋也

指導教員：
鈴木 優 准教授

マルチタスク学習時の有効な補助タスク選定*

古田 朋也

内容梗概

マルチタスク学習とは，主となるタスク以外のタスクも同時に解くことができるようにモデルの学習を進めることによって，特定のタスクのみに限定されない特徴量を獲得し，汎化性能を向上させる学習手法である．マルチタスク学習に関する研究は，効率良く学習を進めるための手法の提案や最適なタスクの組み合わせを効率良く探索する手法の提案など様々行われている．しかし，補助タスクの選定は感覚的に行っている研究が多く，採用する補助タスクの最適な数や選定基準については明らかになっていない．そこで本研究では，マルチタスク学習モデルを実装し，どのような補助タスクをどの程度採用すると推定精度向上に効果的なのかを調査するために実験を行った．実験では，商品タイトルの興味を惹く度合い推定を主タスクとしたマルチタスク学習モデルを複数構築して，推定精度を比較・分析することによって補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴について調査した．本実験を通して，マルチタスク学習モデル構築時に選定する補助タスクの採用数を増やすことによって推定精度向上が期待できること，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強い補助タスクほど推定精度向上に有効であるという知見が得られた．この知見は，マルチタスク学習モデル構築時における補助タスク選定方法の一基準となり得るものであると考える．

キーワード

マルチタスク学習，補助タスク選定，ニューラルネットワーク，テキスト処理

*岐阜大学大学院 自然科学技術研究科 知能理工学専攻 知能情報学領域 修士論文, 学籍番号: 1214525067, 2023 年 3 月 5 日.

目次

図目次 iv

表目次 v

第 1 章 はじめに 1

第 2 章 基本的事項 4

2.1 マルチタスク学習 4

2.2 BERT 5

2.3 決定係数 R^2 5

2.4 RMSE (Root Mean Squared Error) 6

2.5 対応のない 2 標本 t 検定 7

第 3 章 関連研究 9

第 4 章 構築モデル 11

4.1 使用データセット 11

4.2 補助タスク 11

4.3 ラベル付与 12

4.4 モデル概要 14

第 5 章 実験 17

5.1 実験 1: 有効な補助タスクの特徴と採用数についての分析 17

5.1.1 実験手順 17

5.1.2 結果・考察 18

補助タスク採用数についての分析 18

推定精度向上に効果的な補助タスクの特徴についての分析 . 20

5.2 実験 2: 補助タスクがノイズとして扱われているかの検証 22

5.2.1 実験手順 22

5.2.2 結果・考察 23

第 6 章　おわりに 25

謝辞 27

参考文献 28

発表リスト 30

図目次

2.1　自由度 18，$t = 2$ のときの t 分布と p 値の関係 8

4.1　構築するモデルの概要 14

表目次

4.1 データセット内の主タスク用ラベルの内訳 13
4.2 データセット内の各補助タスク用ラベルの内訳 14

5.1 構築したモデルの決定係数平均値 19
5.2 構築したモデルの主タスク RMSE の平均値，p 値 20
5.3 補助タスク採用数と主タスク RMSE の平均値の関係 21
5.4 各補助タスクの効果の大きさと主タスクとの関係 21
5.5 ノイズとして扱う補助タスクを採用したモデルの主タスク RMSE の平均値，決定係数平均値，Sub1 採用時と Sub2 採用時それぞれとの検定による p 値 23

第 1 章　はじめに

マルチタスク学習 [1][2] とは，主となるタスク以外のタスクも同時に解くことができるようにモデルの学習を進めることによって，特定のタスクのみに限定されない特徴量を獲得し，汎化性能を向上させる学習手法である．マルチタスク学習に関する研究は，画像処理やテキスト処理に適用した研究の他にも，効率良く学習を進めるための手法の提案をしている研究や最適なタスクの組合せを効率良く探索する手法の提案をしている研究など様々なものが行われている．しかし，採用する補助タスクについては主タスクと関連していると思われるタスクを感覚的に選定している研究が多く，明確な基準に従って選定されているわけではない．また，採用する補助タスクの数について言及している研究や補助タスクの明確な選定基準を示している研究は調べた限りでは行われていない．そのため，マルチタスク学習モデルを構築する際にどのような補助タスクをどの程度採用するべきなのかは明らかになっていない．選定基準が明確でないため，実際に各タスク用のデータを作成してモデルを構築してみないと補助タスクの良し悪しが判断できない．それでは，補助タスクが効果的ではなかった場合に作成したデータが無駄になってしまう．そこで，本研究ではマルチタスク学習モデルを構築する際にどのような補助タスクをどの程度採用すると推定精度向上に効果的なのかを調査する．どのような補助タスクをどの程度採用すると推定精度向上に効果的なのかを調査して補助タスクの選定基準を定めることによって，構築するモデルの推定精度向上に役立つだけでなく，効果的でないタスクのデータを作成する必要がなくなるためデータ作成の面でも役立つと考える．

本論文では，商品タイトルの興味を惹く度合い推定を主タスクとしたマルチタスク学習モデルを複数構築して，推定精度を比較・分析することによって補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴について調査する実験を行った．実験では，主タスクである興味を惹く度合い推定に加えて 5 つの補助タスクを用意した．複数人にアンケートを取ることによって作成したデータセットを使用して，用意した 5 つの補助タスク全通りの組合せである 32 種類のマルチタスク学習モデルを 10 個ずつ構築した．その際，BERT[3] を用いた興味を惹く度合い推定を行うシングルタスク学習モデルをベースにしてマルチタスク

学習モデルを構築した．その後，構築したモデルのテスト用データ推定精度に対して採用した補助タスクの組合せごとの平均値を算出し，それを比較・分析することによって，補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴について調査した．

補助タスクの採用数についての分析では，補助タスクの採用数と採用数ごとの推定精度の間に相関が見られるかどうかを調査した．推定精度向上に効果的な補助タスクの特徴についての分析では，補助タスク用データと主タスク用データの相関係数の値と補助タスクを採用した際の効果の大きさに相関が見られるかどうかを調査した．それぞれの分析結果から，補助タスクの採用数を増やしたモデルであるほど推定精度が高くなる傾向があること，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強いタスクを採用することによって推定精度が高くなる傾向があることを確認した．

本実験を通して，マルチタスク学習モデル構築時に選定する補助タスクの採用数を増やすことによって推定精度向上が期待できること，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強い補助タスクほど推定精度向上に有効であるという知見が得られた．この知見は，マルチタスク学習モデル構築時における補助タスク選定方法の一基準となり得るものであると考える．

本論文における貢献は以下の通りである．

- マルチタスク学習モデル構築時に選定する補助タスクの採用数の増加に伴う推定精度向上を確認した．
- シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強い補助タスクほど推定精度向上に有効であることを確認した．

本論文の構成は以下の通りである．2 章では本論文にて用いた技術や手法についての基本的事項を述べる．3 章ではマルチタスク学習に関する研究について述べる．4 章では本論文にて行った実験における主タスクの設定やそれに伴い用意した補助タスク，使用したデータセットの作成方法，構築したモデルについて述べる．5 章では実験の目的と内容，結果・考察について述べる．最後に 6 章では本論文のまと

めと今後の課題について述べる.

第 2 章　基本的事項

2.1　マルチタスク学習

マルチタスク学習とは，主となるタスク以外のタスクも同時に解くことができるようにモデルの学習を進めることによって，特定のタスクのみに限定されない特徴量を獲得し，汎化性能を向上させる学習手法である．あるタスクの学習で得た知識を別のタスクに活用するという考え方は転移学習 [4] に類似している．転移学習では，解きたいタスクについての学習を行う際に，別のタスクの学習で得られたモデルのパラメータの値を流用する．それにより，別のタスク得られた知識を活用でき，少ない学習データでの学習が可能になる．転移学習ではあるタスクの学習が完了した後に別のタスクの学習を開始する．それに対して，マルチタスク学習では一つのモデルの中で同時に複数のタスクの学習を進めていく．また，マルチタスク学習では複数のタスクを同時に解くことが可能になるように学習を進めていくため，モデルの出力時には主タスクの結果に加えて補助タスクの結果も同時に出力される．そのため，転移学習とマルチタスク学習では別のタスクで得た知識を活用するという考え方は類似しているが，そのアプローチの方法が異なっている．マルチタスク学習を適用することによって，主タスクの汎化性能の向上が期待できる，一つのモデルで複数のタスクを解くことが可能となる，学習時間や総パラメータ数の削減ができるなどのメリットが得られる．

マルチタスク学習モデルのパラメータ更新時には，各タスクの損失を算出した後，式 2.1.1 のように，すべてのタスクの損失の重み付き和を用いることによって，主タスク以外のタスクも同時に解くことが可能になるように学習を進めていく．

$$L_{Total} = \lambda_1 \times L_1 + \lambda_2 \times L_2 + \cdots + \lambda_n \times L_n \tag{2.1.1}$$

ここでは，モデル内で同時に解くタスクの総数を n，各タスクの損失をそれぞれ $L_1, L_2, \cdots, L_n$ としている．それぞれのタスクの損失と重み $\lambda_1, \lambda_2, \cdots, \lambda_n$ の積をとる．タスクの損失と重みの積を全タスク分足し合わせることによって，L_{Total} を算出し，この L_{Total} を用いてモデルのパラメータ更新を行う．なお，重み付き和を算出する際の重み $\lambda_1, \lambda_2, \cdots, \lambda_n$ の値を大きくすることによって，そのタスクの損失を下げることを重視した学習が可能になる．

2.2 BERT

BERT(Bidirectional Encoder Representations from Transformers) とは，Transformer[5] の Encoder を使用したニューラルネットワークモデルである．モデルの構造を修正せずとも，転移学習することで，様々な自然言語処理タスクに応用できるモデルとなっている．転移学習前の事前学習として MLM(Masked Language Modeling) と NSP(Next Sentence Prediction) を長い文章を含むデータセットを用いて行っている．なお，この事前学習に用いられるデータセットに事前にラベルは付与されていない．

MLM は複数箇所が穴になっている文章に対して穴埋め単語を予測するタスクとなっている．入力されたトークンの 15% に対して [MASK] トークンに置換している．その際，入力されたトークン 15% のうち，80% を [MASK] トークンに，10% をランダムなトークンに，10% をそのままのトークンに置換する．そして，置換された元のトークンを予測するタスクを解いている．NSP は入力された二文が連続した文かどうかを判定するタスクとなっている．入力する二文のペアのうち，50% を連続した文のペア，残りの 50% を連続していない文のペアとしている．これら二種類の事前学習にて獲得したネットワークのパラメータに対して別のタスク用にファインチューニングを行うことによって，高い推定精度を発揮することが期待できるモデルである．

2.3 決定係数 R^2

決定係数 R^2 とは，入力データに対する回帰モデルの当てはまりの良さを表す評価指標である．モデルによって予測された数値と正解の数値を比較して，どの程度一致しているかを表現した数値となっている．予測された数値と正解の数値の差が小さいほどと 1 に近い値をとり，この値が大きいほどモデルがデータにうまく当てはまっていることを意味する．そのため，回帰モデルの良し悪しを判断するための評価指標として用いられる数値の一つとなっている．

決定係数 R^2 の算出は，モデルに入力したデータの総数を n，データ i の正解値を y_i，モデルによるデータ i の予測値を $\hat{y}_i$，正解値 $y_1, y_2, \cdots, y_n$ の平均値を $\bar{y}$ と

して，以下のように行う．

$$R^2 = 1 - \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \tag{2.3.1}$$

2.4 RMSE (Root Mean Squared Error)

RMSE (Root Mean Squared Error) とは，各データに対する予測値と正解値の差の平均値を算出することによって，モデルがあるデータについて予測した場合に出力される数値にどの程度の誤差があるのかを示した数値である．RMSE はモデルによる誤差の数値を示しているため，機械学習モデルの学習時の損失関数や回帰モデルの良し悪しを判断するための評価指標として用いられる．この数値はモデルによる誤差の数値であるため，0 に近いほど良いモデルであると判断できる．

RMSE の算出式を以下に示す．

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.4.1}$$

ここでは，モデルに入力したデータの総数を n，データ i の正解値を y_i，モデルによるデータ i の予測値を $\hat{y}_i$ としている．各データの予測値と正解値の差の二乗を算出して，それらの平均値に対して平方根をとることによって求めることができる．

RMSE の値を評価指標として用いることによって大きな誤差を重要視することができる．これは，算出する際に予測値と正解値の差を二乗していることによって，大きな誤差が RMSE の値に大きく影響するためである．また，二乗誤差の平均値に対して平方根をとることによって，二乗したことによって発生した単位のずれを修正している．それによって，算出された数値を見た際にどの程度の誤差があるのかを理解しやすくしている．例えばモデルの RMSE の値が 0.35 であった場合は，このモデルによる予測にはおよそ ± 0.35 のずれが存在すると読み取ることができる．

2.5　対応のない 2 標本 t 検定

対応のない 2 標本 t 検定は，二つの独立した集団がある場合において，各集団の平均値の間に差があるかどうかを確かめる際に行う検定である．例えば，ある学校で行われたテストの平均点が 1 組と 2 組で差があるかどうか確かめる場合などに用いられる．この検定を行うことによって，観察された結果が偶然起きたことなのか，必然的に起きたことなのかを確かめることができる．

検定を行う際には，まず帰無仮説を設定する．帰無仮説には棄却したい仮説を設定するため，上述のテストの例においては帰無仮説を「1 組の平均点と 2 組の平均点の間に差はない」とする．この帰無仮説を棄却することによって，2 群の間に差があることを示すことができる．また，帰無仮説を棄却するかどうかの基準となる有意水準 α の値を設定しておく．有意水準 α の値は検定によって様々であるが，帰無仮説が成立する確率が 5% となる $\alpha = 0.05$ や 1% となる $\alpha = 0.01$ がよく採用される．

次に，検定統計量である t 値を算出する．対応のない 2 標本 t 検定においては，1 群目の標本サイズを n_1，標本平均を $\bar{x}_1$，母平均を μ_1，2 群目の標本サイズを n_2，標本平均を $\bar{x}_2$，母平均を μ_2 として以下のように算出する．

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s^2(\frac{1}{n_1} + \frac{2}{n_2})}} \tag{2.5.1}$$

ここで，帰無仮説によって母平均 μ_1 と μ_2 の差は 0 となるため，式 2.5.1 は以下のように書き換えることができる．

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(\frac{1}{n_1} + \frac{2}{n_2})}} \tag{2.5.2}$$

上記の式 2.5.2 を用いて検定統計量である t 値を算出する．なお，式 2.5.2 における s^2 は 2 つの標本の不偏分散 s_1^2，s_2^2 をブールしたものであり，以下のように算出する．

$$s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2} \tag{2.5.3}$$

そして，求めた t 値を基にして，帰無仮説が成立する確率を示す p 値を算出する．p 値の算出には自由度 $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ の t 分布を用いる．自由

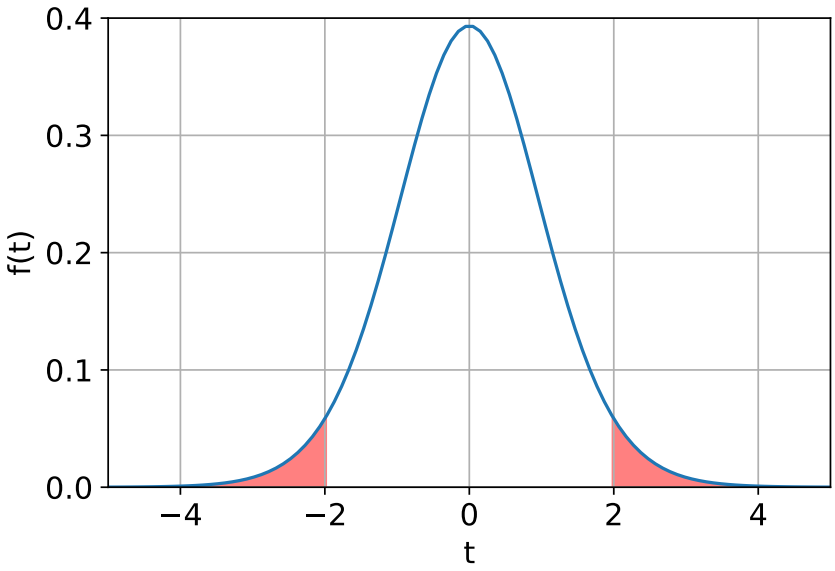

図 2.1　自由度 18，$t = 2$ のときの t 分布と p 値の関係

度が 18，求めた t 値が 2 であった場合のの t 分布と p 値の関係を図 2.1 に示す．ここでは求めた t 値が 2 であるため，$-\infty$ から-2 と 2 から ∞ の範囲における t 分布の面積を算出する．図 2.1 においては赤色に着色してある部分の面積である．この面積が p 値となる．

最後に，事前に設定した有意水準 α の値と求めた p 値を比較することによって帰無仮説が棄却できるかどうかを判断する．有意水準 α と p 値の関係が $\alpha < p$ となった場合は，帰無仮説は棄却できないため，2 群の間に有意な差があるとは言えないという結論になる．一方，有意水準 α と p 値の関係が $\alpha > p$ となった場合は，帰無仮説が棄却できるため，2 群の間に有意な差があるという結論になる．

第 3 章　関連研究

本章では，マルチタスク学習に関連する研究についてを記述する．Lamprinidis ら [6] は，ニュース記事の見出しからニュース記事の人気予測を行う研究にマルチタスク学習を適用している．この研究ではニュース記事の人気予測という主タスクに加えて，品詞情報のタグ付けタスクとニュース記事のトピック予測という二つの補助タスクも同時に学習することによって予測性能の向上を図っている．

Lu ら [7] は，テキスト要約を行う研究にマルチタスク学習を適用している．この研究ではテキスト要約を行う主タスクに加えて，構文解析タスクとテキストのカテゴリ予測という二つの補助タスクも同時に学習することによって予測性能の向上を図っている．

Mulyar ら [8] は，自由記述形式の電子カルテから投薬履歴や治療履歴などの様々な情報を抽出する研究にマルチタスク学習を適用している．この研究では，投薬履歴や治療履歴などの抽出タスクをすべて一つのモデルで解くことによって，予測性能の向上と共にモデル構築コストの削減も図っている．

この他にも，マルチモーダルモデルによる感情分析にマルチタスク学習を適用している Akhtar ら [9] の研究や，係り受け解析を行うモデルにマルチタスク学習を適用している Duong ら [10] の研究などが行われている．このようにマルチタスク学習は画像やテキストなど様々な分野の研究にて適用されている．

また，上記のようなマルチタスク学習を適用して推定精度向上を図っている研究以外に，効率良くマルチタスク学習を行うための研究も行われている．Yu ら [11] は，効率良くマルチタスク学習を行うためのパラメータ更新時の損失重み付き和の算出方法を提案している．各タスクの勾配が逆方向を向いていた場合に勾配の向きを補正することによって効率良くパラメータを更新していく手法となっている．強化学習モデルにマルチタスク学習を適用した実験において，この手法を用いることによって精度が大幅に改善している．

Zhang ら [12] は，タスクごとの学習の難易度に着目した学習手法を提案している．マルチタスク学習モデル構築時にタスクごとに Early Stopping を設定して簡単なタスクの学習を途中で打ち切ることによって，簡単なタスクの過学習を防止し，主タスクの収束を促進する手法となっている．顔のランドマーク検出を主タス

クとするマルチタスク学習モデルにおいて，この手法を適用することによって検出の精度が向上することが確認されている．

Fifty ら [13] は，マルチタスク学習における補助タスクの最適な組合せを効率良く探索する手法を提案している．全タスクを採用したマルチタスク学習モデルの構築時の勾配情報を基に算出したタスク同士の親和性を用いて最適な組合せを決定する手法となっている．この手法は最適な組合せを決定するためにモデルを複数構築する必要がないため，補助タスクの組合せ探索時の大幅なコスト削減になる．

Standley ら [14] は，一つの入力に対して解きたいタスクが複数ある場合における最適なタスクの割り当てを行うフレームワークを提案している．この手法では，関連するタスクは同一ネットワークにて，競合するタスクは別のネットワークにて学習が行われる．そのため，構築するモデルの総数は増加する一方で，最適なタスクの組合せによるマルチタスク学習モデルが構築できるため各タスクの推定精度向上が期待できる．

上記のようにマルチタスク学習に関して様々な研究が行われているが，採用する補助タスクについては主タスクと関連していると思われるタスクを感覚的に選定している研究が多く，明確な基準に従った選定が行われているわけではない．また，採用する補助タスクの数について言及している研究や補助タスクの明確な選定基準を示している研究は調べた限りでは行われていない．そのため，マルチタスク学習モデルを構築する際にどのような補助タスクをどの程度採用するべきなのかは明らかになっていない．そこで本研究では，マルチタスク学習モデルの構築時にどのような補助タスクをどの程度採用すると推定精度向上に効果的なのかを調査するために 5 つの補助タスクを用意して実験を行った．

第 4 章　構築モデル

本論文では，商品タイトルの興味を惹く度合い推定を主タスクとしたマルチタスク学習モデルを構築して実験を行う．モデル構築の際には主タスクの学習に加えて補助タスクの学習も同時に行う．モデル構築に使用するデータセットについてを 4.1 節にて，用意した補助タスクについてを 4.2 節にて，補助タスクを考慮して行ったラベル付与についてを 4.3 節にて，構築するモデルの概要についてを 4.4 節にて述べる．

4.1　使用データセット

本論文では，国立情報学研究所から提供されている楽天データセット*を使用した．このデータセットは楽天市場に出品されている商品データ約 2 億 8,300 万件によって構成されている．各商品データには，商品タイトル，商品画像，商品説明文，商品カテゴリなどが含まれている．

上記データセットのうち，食品カテゴリに属する商品のタイトルを用いたデータセットを作成して実験を行った．本論文で構築するのは，商品タイトルの興味を惹く度合い推定を行うモデルである．そのため，商品画像などの見た目に関わる要素が興味を惹くかどうかに大きく影響する商品データを使用するのは不適切であると考えた．そこで，見た目の影響が小さい食品カテゴリに属する商品データを使用することにした．

4.2　補助タスク

本論文では，商品タイトルの興味を惹く度合い推定を行う主タスクに加えて，5 つの補助タスクを用意して実験を行う．先行研究では，主タスクと関連していると思われるタスクを感覚的に選定して補助タスクとする研究が多い．そのため，本研究においても同様の選定方法にて補助タスクを用意した．

*楽天グループ株式会社 (2020): 楽天市場データ．国立情報学研究所情報学研究データリポジトリ．(データセット)．https://doi.org/10.32130/idr.2.1

本論文にて構築するマルチタスク学習モデルの主タスクが商品タイトルの興味を惹く度合い推定であるため，興味を惹く要因となり得る要素についてを推定するタスクである下記 5 つを補助タスクとして用意することにした．

Sub1　販売商品把握の可否の推定
Sub2　呼びかけ表現の有無の推定
Sub3　希少性/限定性を示す表現の有無の推定
Sub4　評価/実績の記載の有無の推定
Sub5　商品の特徴/アピールポイント記載の有無の推定

本論文にて用意した 5 つの補助タスクは，4.3 節にて述べるアンケートにおける投票人数を推定する回帰タスクとしている．

4.3　ラベル付与

用意した商品タイトルに対してアンケートを取り，主タスクのラベルに加えて 4.2 節にて述べた各補助タスクに対応するラベルも付与する．商品タイトルを読んだ際に興味を惹かれるかどうかの基準は人によって様々であるため，一人の意見を用いて作成したデータセットでは汎化性能が低くなってしまう．そのため，商品タイトル一つに対して複数人にアンケートを取りデータセットを作成していく．アンケートでは，作業者に商品タイトルを一つずつ提示して，以下に示す主タスク用と補助タスク用の合計 6 つの質問をする．

i. 主タスク用の質問
   - Main その商品について「詳しく知りたい」または「買ってみようかな」と感じますか?

ii. 補助タスク用の質問
   - Sub1 何を販売しているか理解/把握できますか?
   - Sub2 閲覧者へ呼びかける表現が含まれていますか?
   - Sub3 商品の希少性/限定性 (数量限定，希少部位など) について示す表現が含まれていますか?
   - Sub4 商品の評価/実績 (○○賞獲得，ランキング 1 位など) について示す表

現が含まれていますか?

- Sub5 商品の特徴やアピールポイントについて示す表現が含まれていますか?

上記の質問はすべて'Yes' か'No' の二択で回答する形式となっており，各質問に対して'Yes' と回答した人数を集計して，その人数を各タスクのラベルとして付与する．そのため，本論文で作成したデータセットには，6 種類のタスクそれぞれに対応したラベルが商品タイトル一つに対して付与されている．なお，主タスクの学習に用いる質問の集計に関しては 3 票以上のラベルを同列に扱う．商品タイトルを読んで興味を惹かれるかどうかの基準は人によって様々であり，読んだ人全員の興味を惹く商品タイトルを作成することは困難であると思われる．そのため，読んだ人のおよそ 30% の興味を惹くことができる商品タイトルであれば十分であると考えられる．そして，およそ 30% の人の興味を惹くことができるかどうかがモデルの推定によって分かるだけでも有用性はあるのではないかと考えたため，3 票以上のラベルを同列に扱うことにした．

本論文では，上記手順にて商品タイトル 10,587 件に対してラベルを付与してデータセットを作成した．その際に，商品タイトル一つにつき 9 人に対してアンケートを実施した．その結果，主タスクである興味を惹く度合い推定に用いるラベルの内訳は表 4.1 のようになった．また，各補助タスクに対応するラベルの内訳は表 4.2 のようになった． 表 4.2 の補助タスクの番号は 4.2 節にて述べた補助タスクの番号と対応している．作成したデータセットを使用して，興味を惹く度合い推定を主タスクとするマルチタスク学習モデルを構築する．

表 4.1　データセット内の主タスク用ラベルの内訳

タスク	票数				総数
	0	1	2	3 以上	
Main	6,952	2,616	746	272	10,587

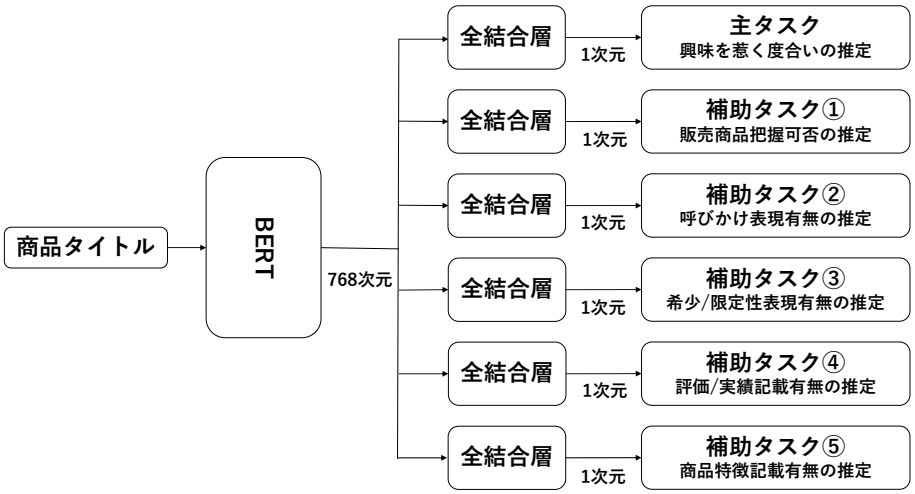

図 4.1　構築するモデルの概要

4.4　モデル概要

本論文では，BERT を用いた興味を惹く度合い推定を行うシングルタスク学習モデルをベースにしてマルチタスク学習モデルを構築する．構築するモデルの概要を図 4.1 に示す．まず，入力として与えられた商品タイトルを BERT モデルへ入力する．その出力となる 768 次元のベクトルを各タスクへ分岐させる．分岐する数は補助タスクの採用数によって異なり，図 4.1 は用意した補助タスク 5 つをすべて採

表 4.2　データセット内の各補助タスク用ラベルの内訳

タスク	票数										総数
	0	1	2	3	4	5	6	7	8	9	
Sub1	3	5	10	10	35	64	144	525	2,175	7,616	10,587
Sub2	8,070	1,923	422	120	37	11	4	0	0	0	10,587
Sub3	8,227	1,692	329	121	105	57	37	9	9	1	10,587
Sub4	7,539	1,664	481	315	224	153	113	72	22	4	10,587
Sub5	62	3,249	2,089	1,431	1,237	1,027	821	500	147	24	10,587

用した場合のモデル構造となっている．分岐した後，768 次元のベクトルは各タスクに応じた全結合層に入力され，それぞれ 1 次元の数値が出力される．そのため，図 4.1 のモデルの場合は商品タイトルを一つ入力すると 6 つのタスクの推定結果が同時に出力されることになる．なお，本論文にて用意したタスクは主タスクも含めてすべて回帰タスクとなっている．上記のようなマルチタスク学習モデルを 4.3 節に従って作成したデータセットを用いて構築する．

本論文では，訓練済み BERT モデルとして，東北大学の乾・鈴木研究室の訓練済み日本語 BERT モデル[†]を使用した．訓練済みモデルは日本語版 Wikipedia にて事前学習が行われており，語彙数は 32,000 となっている．訓練時には BERT モデルの最終層以外のパラメータは更新しないように設定してファインチューニングを行った．そのため，訓練時に更新されるのは BERT モデルの最終層と各タスクに応じた全結合層のパラメータのみとなっている．

また，マルチタスク学習時のパラメータ更新には採用したタスクの損失すべての重み付き和を用いる．重み付き和を算出する際の重みを大きくすることによって，そのタスクの損失を下げることを重視した学習が可能になる．マルチタスク学習を適用する目的は主タスクの損失を下げることであるため，更新時には主タスクの損失を下げることを重視したい．そこで，本論文ではマルチタスク学習時のパラメータ更新に用いる主タスクと補助タスクの損失の比率が 6 : 4 になるように設定をした．そのため，パラメータ更新に用いる損失すべての重み付き和の算出式を式 4.4.1，式 4.4.2 の通りに設定をしている．

$$L_{Total} = L_{Main} + \frac{2}{3N} L_{ST} \tag{4.4.1}$$

$$L_{ST} = L_{Sub1} + L_{Sub2} + \cdots + L_{SubN} \tag{4.4.2}$$

上式において，パラメータ更新に用いる損失すべての重み付き和を L_{Total}，主タスクの損失を L_{Main}，補助タスクの損失の合計値を L_{ST} としている．また，$L_{Sub1}, L_{Sub2}, \cdots, L_{SubN}$ はそれぞれの補助タスクにおける損失となっている．補助タスクの損失の合計値である L_{ST} を $\frac{2}{3N}$ 倍することによって，主タスクと補助タスクの損失の比率が 6 : 4 になるように調整している．ここで，N はモデル構築

[†]https://github.com/cl-tohoku/bert-japanese

時に採用した補助タスクの総数であり，本論文では $1 \leq N \leq 5$ の範囲の整数値となる．上式のように計算した L_{Total} を用いて学習時のパラメータ更新を行い，モデルを構築する．

第 5 章　実験

本論文にて実験を行う目的は二つある．一つ目は，マルチタスク学習モデル構築時に採用する補助タスクの数によってモデルの推定精度が変動するのかどうかを調査することである．二つ目は，マルチタスク学習モデル構築時にどのような補助タスクを採用すると推定精度向上に効果的なのかを調査することである．

5.1　実験 1: 有効な補助タスクの特徴と採用数についての分析

本実験では，用意した補助タスク全通りの組合せによるマルチタスク学習モデルを構築してテスト用データの推定精度を比較・分析することによって，補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴についてを調査する．

5.1.1　実験手順

まず，4.3 節にて作成したデータセットを用いて，4.4 節にて述べたマルチタスク学習モデルを構築する．本論文では，4.2 節にて述べた 5 つの補助タスク全通りの組合せである 32 種類のマルチタスク学習モデルを構築する．モデル構築に用いるデータセットのうち，8 割を訓練用データ，1 割を評価用データ，1 割をテスト用データとして扱い，訓練用データに対しては同じデータを複製することによって主タスクのラベル間の偏りを解消した．ここで分割した訓練用データ，評価用データ，テスト用データを用いて，すべてのモデルを構築する．そのため，どのような補助タスクの組合せのモデルを構築する場合であっても訓練用データ，評価用データ，テスト用データの中身は同一となっている．学習時の最大エポック数は 10,000 として，主タスクの Validation Loss が 100 エポック改善しない場合は Early Stopping を適用して学習を停止するように設定をした．そのため，最終的なモデルのパラメータは主タスクの Validation Loss が最も低かった場合のものとなる．また，BERT モデルのパラメータ更新設定やパラメータ更新に用いる損失すべての重み付き和の設定は 4.4 節にて述べた通りである．

補助タスクの組合せ一つに対して 10 個ずつモデルを構築して，テスト用データ推定時の主タスク RMSE の平均値を算出する．また，補助タスクを全く採用しないシングルタスク学習モデルの RMSE 平均値との間に差があるか否かの検定を行い，p 値を算出する．その際には，帰無仮説を「構築したマルチタスク学習モデルの主タスク RMSE の平均値とシングルタスク学習モデルの主タスク RMSE の平均値との間に差はない」として対応のない 2 標本 t 検定を行った．そして，得られた各組合せによる結果を比較・分析することによって，補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴について調査し，マルチタスク学習モデルにおける補助タスクの選定方法について考察する．

5.1.2 結果・考察

補助タスク各組合せによる構築モデルの決定係数平均値を表 5.1 に，主タスク RMSE の平均値，検定時の p 値を表 5.2 に示す．表 5.1，5.2 に記載されている補助タスクの番号は 4.2 節にて述べた補助タスクと対応している．表 5.1，表 5.2 に示した補助タスクの組合せ全 32 通りの結果を基に，補助タスクの採用数による推定精度の変動や推定精度向上に効果的な補助タスクの特徴についてを分析する．

補助タスク採用数についての分析

表 5.2 にて補助タスクの採用数に注目すると，タスク同士の組合せによって多少の差はあるが採用数が多いモデルであるほど主タスク RMSE の平均値や p 値が小さくなる傾向がありそうなことが読み取れる．補助タスクの採用数と RMSE 平均値の関係について分析した結果を表 5.3 に示す．ここでは，補助タスクの採用数ごとに表 5.2 中の主タスク RMSE の平均値を算出している．そして，補助タスクの採用数と採用数ごとの RMSE 平均値の相関係数を算出している．表 5.3 を見ると，補助タスクの採用数と採用数ごとの RMSE 平均値の相関係数は-0.7847 となっており，この二つの間には強い相関が見られることが分かった．この結果より，補助タスクの採用数を増やすことによって推定精度向上が期待できると言えるのではないかと考える．

本実験にて，マルチタスク学習モデル構築時に選定する補助タスクの採用数を増

やすことによって推定精度向上が期待できるという知見が得られた．補助タスクの採用数を増やすことによって推定精度向上が期待できる一方で，データ作成時にかかるコストも増加してしまう．そのため，補助タスクの採用数を増やす際にはデータ作成時にかかるコストも見極めることが重要になると考える．

表 5.1　構築したモデルの決定係数平均値

構築したモデル	決定係数 (R^2)					
	Main	Sub1	Sub2	Sub3	Sub4	Sub5
Single-Task	-0.4328	-	-	-	-	-
Multi-Task(Sub1)	0.0790	0.1659	-	-	-	-
Multi-Task(Sub2)	0.0755	-	0.3242	-	-	-
Multi-Task(Sub3)	0.0622	-	-	0.5843	-	-
Multi-Task(Sub4)	0.0878	-	-	-	0.5856	-
Multi-Task(Sub5)	0.1890	-	-	-	-	0.5688
Multi-Task(Sub1,2)	0.1214	0.0395	0.2667	-	-	-
Multi-Task(Sub1,3)	0.0714	0.0239	-	0.5754	-	-
Multi-Task(Sub1,4)	0.1302	0.0574	-	-	0.5485	-
Multi-Task(Sub1,5)	0.1654	0.0451	-	-	-	0.5343
Multi-Task(Sub2,3)	0.1034	-	0.2842	0.5261	-	-
Multi-Task(Sub2,4)	0.1225	-	0.2620	-	0.5392	-
Multi-Task(Sub2,5)	0.1397	-	0.2464	-	-	0.4970
Multi-Task(Sub3,4)	0.0960	-	-	0.5464	0.5586	-
Multi-Task(Sub3,5)	0.1698	-	-	0.5210	-	0.5100
Multi-Task(Sub4,5)	0.1221	-	-	-	0.5433	0.4765
Multi-Task(Sub1,2,3)	0.1520	0.0355	0.1987	0.5161	-	-
Multi-Task(Sub1,2,4)	0.1362	0.0238	0.2223	-	0.5347	-
Multi-Task(Sub1,2,5)	0.1766	0.0527	0.2320	-	-	0.4939
Multi-Task(Sub1,3,4)	0.1334	-0.0075	-	0.5422	0.5376	-
Multi-Task(Sub1,3,5)	0.1086	0.0298	-	0.5350	-	0.4906
Multi-Task(Sub1,4,5)	0.1641	0.0332	-	-	0.5208	0.4826
Multi-Task(Sub2,3,4)	0.1210	-	0.2132	0.4922	0.5062	-
Multi-Task(Sub2,3,5)	0.1533	-	0.2290	0.4955	-	0.4607
Multi-Task(Sub2,4,5)	0.1653	-	0.2009	-	0.5026	0.4512
Multi-Task(Sub3,4,5)	0.1710	-	-	0.5053	0.4946	0.4363
Multi-Task(Sub1,2,3,4)	0.1443	0.0010	0.2054	0.5114	0.5221	-
Multi-Task(Sub1,2,3,5)	0.1560	0.0163	0.1659	0.5016	-	0.4705
Multi-Task(Sub1,2,4,5)	0.1454	0.0205	0.1767	-	0.5183	0.4544
Multi-Task(Sub1,3,4,5)	0.1810	0.0322	-	0.4996	0.4978	0.4612
Multi-Task(Sub2,3,4,5)	0.1751	-	0.1791	0.5065	0.4974	0.4275
Multi-Task(Sub1,2,3,4,5)	0.1948	0.0072	0.1031	0.5007	0.4873	0.4288

推定精度向上に効果的な補助タスクの特徴についての分析

表 5.2 にて，採用しているタスクに注目すると Sub5 を補助タスクとして採用しているモデルの主タスク RMSE の平均値や p 値が他のモデルよりも小さくなっていることが読み取れる．各補助タスクの推定精度向上効果の大きさと主タスクとの関係について分析した結果を表 5.4 に示す．ここでは，まず表 5.2 を基に各補助タスクを採用した場合の主タスク RMSE の平均値と採用しなかった場合の主タスク

表 5.2　構築したモデルの主タスク RMSE の平均値，p 値

構築したモデル	RMSE(Main)	p-value
Single-Task	0.8684	-
Multi-Task(Sub1)	0.7262	0.1185
Multi-Task(Sub2)	0.7278	0.1217
Multi-Task(Sub3)	0.7330	0.1353
Multi-Task(Sub4)	0.7228	0.1102
Multi-Task(Sub5)	0.6815	**0.0448**
Multi-Task(Sub1,2)	0.7094	0.0830
Multi-Task(Sub1,3)	0.7293	0.1259
Multi-Task(Sub1,4)	0.7058	0.0768
Multi-Task(Sub1,5)	0.6912	0.0562
Multi-Task(Sub2,3)	0.7166	0.0969
Multi-Task(Sub2,4)	0.7089	0.0823
Multi-Task(Sub2,5)	0.7022	0.0705
Multi-Task(Sub3,4)	0.7196	0.1029
Multi-Task(Sub3,5)	0.6897	0.0535
Multi-Task(Sub4,5)	0.7088	0.0829
Multi-Task(Sub1,2,3)	0.6969	0.0633
Multi-Task(Sub1,2,4)	0.7036	0.0728
Multi-Task(Sub1,2,5)	0.6869	0.0502
Multi-Task(Sub1,3,4)	0.7045	0.0748
Multi-Task(Sub1,3,5)	0.7127	0.0949
Multi-Task(Sub1,4,5)	0.7095	0.0662
Multi-Task(Sub2,3,4)	0.7095	0.0834
Multi-Task(Sub2,3,5)	0.6965	0.0625
Multi-Task(Sub2,4,5)	0.6915	0.0559
Multi-Task(Sub3,4,5)	0.6891	0.0531
Multi-Task(Sub1,2,3,4)	0.6998	0.0681
Multi-Task(Sub1,2,3,5)	0.6950	0.0613
Multi-Task(Sub1,2,4,5)	0.6996	0.0671
Multi-Task(Sub1,3,4,5)	0.6849	**0.0483**
Multi-Task(Sub2,3,4,5)	0.6874	0.0511
Multi-Task(Sub1,2,3,4,5)	0.6790	**0.0424**

RMSE の平均値を算出し，その差を取った．算出した RMSE 平均値の差が大きいタスクほど，採用した場合の効果が大きいタスクであると言える．その結果より，Sub5 のタスクが最も効果が大きいこと，Sub3 のタスクが最も効果が小さいこと，Sub1,2,4 のタスクが同程度の効果を発揮していることが分かる．

効果が大きいタスクと小さいタスクの違いを調査するために，使用したデータセット上での主タスクのラベルと各補助タスクのラベルの相関係数の算出を行った．また，採用した補助タスクがシングルタスク学習モデルにてどの程度学習可能なのかを測定するため，補助タスクのみを解くシングルタスク学習モデルを構築してテスト用データ推定時の決定係数を算出した．このシングルタスク学習モデルはマルチタスク学習モデル構築時と同じ訓練用データ，評価用データ，テスト用データを使用しており，学習時の最大エポック数は 10,000，Validation Loss が 100 エポック改善しない場合は Early Stopping を適用して学習を停止するように設定をして構築した．表 5.4 の決定係数の値はモデル 10 個分の平均値となっている．そして，RMSE 平均値の差 (表 5.4 の①) とデータセット上での主タスクとの相関係数の値 (表 5.4 の②) の相関係数を算出した．算出した相関係数は-0.6822 となり，強い相関があるとは言い難い結果となった．しかし，補助タスクのみを解くシング

表 5.3　補助タスク採用数と主タスク RMSE の平均値の関係

採用数	0	1	2	3	4	5
RMSE	0.8684	0.7183	0.7081	0.6983	0.6934	0.6790
相関係数	-0.7847					

表 5.4　各補助タスクの効果の大きさと主タスクとの関係

		Sub1	Sub2	Sub3	Sub4	Sub5
採用/不採用時の主タスク RMSE の平均値	In	0.7010	0.7007	0.7027	0.7004	0.6930
	Out	0.7162	0.7160	0.7140	0.7164	0.7239
①RMSE 平均値の差 ($In - Out$)		-0.0153	-0.0153	-0.0112	-0.0160	-0.0388
② データセット上での主タスクとの相関係数		0.1205	0.2944	0.2810	0.3951	0.4824
③ シングルタスク学習時　決定係数 (R^2)		0.2380	0.4734	0.6560	0.7076	0.6512
① と ② の相関係数		-0.6822				
① と ② の相関係数 (Sub3,4,5 のみ)		-0.9311				

ルタスク学習モデルの推定精度が低い Sub1 と Sub2 を除外して同様の相関係数を算出したところ-0.9311 となり，強い相関が見られた．この結果から，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強いタスクほどマルチタスク学習時の補助タスクとして推定精度向上に有効と言えるのではないかと考える．

本実験にて，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば主タスクとの相関が強い補助タスクほど推定精度向上に有効であるという知見が得られた．しかし，本論文では一つのデータセットを使用した実験しか行っていない．そのため，別のデータセットにおいても同様の結果が得られるかどうかを確かめるために，別のデータセットを使用した同様の実験を行う必要がある．また，本実験にて構築したマルチタスク学習モデルに採用しているタスクはすべて回帰タスクとなっている．本実験にて得られた知見である主タスクとの相関の強さを基にした補助タスク選定方法が，分類タスクを解くマルチタスク学習モデルにおいても推定精度向上に有効であるかどうかは不明である．そのため，分類タスクを解くモデルの場合においても同様の実験を行い，上記について確かめる必要がある．

5.2 実験 2: 補助タスクがノイズとして扱われているかの検証

表 5.4 において，シングルタスク学習モデルの推定精度が共に低い Sub1 と Sub2 は主タスクとの相関に差があるにもかかわらず同程度の推定精度向上効果を発揮している．この結果から，シングルタスク学習モデルの推定精度が低い補助タスクは学習時にノイズのようにしか扱われていないのではないかと考えられる．そこで，ノイズとして扱うタスクを補助タスクとして採用したマルチタスク学習モデルを構築して推定精度を比較する追加実験を行うことによって上記について検証した．

5.2.1 実験手順

ノイズとして扱うタスクとして，0 から 3 の範囲のランダムな数値を予測する回帰タスクを設定した．このタスクを補助タスクとして採用したマルチタスク学習モ

デルを構築する．モデル構築時には，5.1 節のモデル構築時と同じ訓練用データ，評価用データ，テスト用データを使用しており，学習時の最大エポック数は 10,000，Validation Loss が 100 エポック改善しない場合は Early Stopping を適用して学習を停止するように設定した．また，BERT モデルのパラメータ更新設定やパラメータ更新に用いる損失すべての重み付き和の設定は 4.4 節にて述べた通りである．上記の設定でモデルを 10 個構築してテスト用データ推定時の主タスク RMSE の平均値を算出する．

モデル構築後，Sub1 を採用した場合の RMSE 平均値，Sub2 を採用した場合の RMSE 平均値との間に差があるか否かの検定を行い，p 値を算出する．その際には，帰無仮説を「Sub1 を採用した場合の主タスク RMSE の平均値とノイズとして扱うタスクを補助タスクとして採用した場合の主タスク RMSE の平均値との間に差はない」，「Sub2 を採用した場合の主タスク RMSE の平均値とノイズとして扱うタスクを補助タスクとして採用した場合の主タスク RMSE の平均値との間に差はない」として，それぞれ対応のない 2 標本 t 検定を行った．この検定を行うことによって，シングルタスク学習モデルの推定精度が低い補助タスクは学習時にノイズのように扱われているのか否かについて検証する．

5.2.2 結果・考察

ノイズとして扱う補助タスクを採用したマルチタスク学習モデルにおけるテストデータ推定時の主タスク RMSE の平均値，決定係数平均値，検定時の p 値を表 5.5 に示す．表 5.5 の結果を見ると，主タスク RMSE の平均値は 0.7073 となっている．表 5.4 上段のタスク採用時の主タスク RMSE の平均値を見ると，Sub1 は 0.7010，Sub2 は 0.7007 となっており，ノイズとして扱うタスクを採用した場合と

表 5.5　ノイズとして扱う補助タスクを採用したモデルの主タスク RMSE の平均値，決定係数平均値，Sub1 採用時と Sub2 採用時それぞれとの検定による p 値

採用した補助タスク	RMSE(Main)	決定係数 (R^2)		p-value	
		Main	Noise	Sub1 In	Sub2 In
Multi-Task(Noise)	0.7073	0.1262	-0.4201	0.4793	0.3565

ほぼ同等の推定精度となっている．そして，p 値においては，Sub1 採用時との検定では 0.4793，Sub2 採用時との検定では 0.3565 といずれも有意な差があるとは言えない結果となっている．以上から，Sub1 と Sub2 の補助タスクが推定精度向上に効果的に作用しているとは言い難く，ノイズのように扱われてパラメータが適度に均されたことによって，結果的に主タスクの過学習を抑制しているだけである可能性が否めない．

上記より，シングルタスク学習モデルの推定精度が低いタスクであると，どのようなタスクを補助タスクとして採用したとしてもノイズとして扱われてしまうため，推定精度向上には同程度の効果しか発揮できないのではないかと考える．また，シングルタスク学習において十分な推定精度を発揮できるタスクである Sub4 であっても補助タスクとしては Sub1,2 と同程度の効果しか発揮できておらず，Sub3 に関しては Sub1,2 よりも小さな効果しか発揮できていない．この結果から，シングルタスク学習において十分な推定精度を発揮できるタスクであっても，補助タスクとして採用した際にノイズのように扱われるタスクと同程度の効果を発揮するには Sub4 程度の主タスクとの相関の強さが必要になるのではないかと考える．そのため，シングルタスク学習において十分な推定精度を発揮できるが主タスクとの相関が低いタスクを採用するのであれば，ノイズとして扱われてしまうタスクを採用してパラメータを適度に均すことを選択した方が高い推定精度を期待できるのではないかと考える．

第 6 章　おわりに

本論文では，マルチタスク学習モデル構築時にどのような補助タスクをどの程度採用すると推定精度向上に効果的なのかについて調査を行った．マルチタスク学習に関する研究は，効率良く学習を進めるための手法の提案や最適なタスクの組合せを効率良く探索する手法の提案など様々行われている．しかし，補助タスクの選定に関しては感覚的に選定している研究が多く，採用する補助タスクの最適な数や選定基準については明らかになっていない．そこで，採用する補助タスクの最適な数や選定基準について調査するために，商品タイトルの興味を惹く度合い推定を主タスクとしたマルチタスク学習モデルを構築して実験を行った．実験では，主タスクである興味を惹く度合い推定に加えて 5 つの補助タスクを用意した．複数人にアンケートを取ることによって作成したデータセットを使用して，5 つの補助タスク全通りの組合せである 32 種類のマルチタスク学習モデルを構築した．その後，構築したモデルごとのテスト用データ推定精度を比較・分析することによって，補助タスクの採用数による精度の変動や推定精度向上に効果的な補助タスクの特徴について調査した．

補助タスクの採用数についての分析においては，補助タスクの採用数ごとに主タスク RMSE の平均値を算出して，補助タスクの採用数を増やすことによる推定精度の変動を調査した．補助タスクの採用数と採用数ごとの RMSE 平均値の相関係数を算出したところ-0.7847 となり，これら二つの間に強い相関が見られた．推定精度向上に効果的な補助タスクの特徴についての分析においては，各補助タスクを採用した場合の主タスク RMSE の平均値と採用しなかった場合の主タスク RMSE の平均値の差を算出して，各補助タスクの推定精度向上効果と効果の大きい補助タスクの特徴を調査した．補助タスク用データと主タスク用データの相関係数の値と補助タスクを採用した際の推定精度向上効果の相関係数を算出したところ-0.6822 となり，強い相関があるとは言い難い結果であった．しかし，シングルタスク学習において十分な推定精度を発揮できないタスクを除外して同様の相関係数を算出したところ-0.9311 と強い相関が見られた．それぞれの分析結果から，補助タスクの採用数を増やしたモデルであるほど推定精度が高くなる傾向があること，シングルタスク学習において十分な推定精度を発揮できて主タスクとの相関が強いタスクを

採用することによって推定精度が高くなる傾向があることを確認した．また，シングルタスク学習において十分な推定精度を発揮できないタスクであるとノイズとして扱われてしまうため，どのようなタスクを補助タスクとして採用した場合でも同程度の効果しか発揮できないという傾向も見られた．

本実験を通して，マルチタスク学習モデル構築時に選定する補助タスクの採用数を増やすことによって推定精度向上が期待できること，シングルタスク学習において十分な推定精度を発揮できるタスクであるという前提の下であれば，主タスクとの相関が強い補助タスクほど推定精度向上に有効であるという知見が得られた．この知見は，マルチタスク学習モデル構築時における補助タスク選定方法の一基準となり得るものであると考える．

しかし，補助タスクの採用数を増加させることで推定精度向上が見込める一方でデータ作成時にかかるコストも増加してしまう．そのため，補助タスクの採用数を増やす際にはデータ作成時にかかるコストも見極めることが重要になると考える．また，本論文では一つのデータセットを使用した実験しか行っておらず，すべてのタスクが回帰タスクとなっている．本実験にて得られた知見である主タスクとの相関の強さを基にした補助タスク選定方法が，分類タスクを解くマルチタスク学習モデルにおいても推定精度向上に有効であるかどうかは不明である．そのため，別のデータセットを使用した同様の実験，分類タスクを解くマルチタスク学習モデルを使用した実験を行うことによって上記についてを確かめる必要がある．

謝辞

本研究を進めるにあたって、指導教員である鈴木優准教授にたくさんのご指導，ご助言をいただきました．学部時代の時も含めて約 3 年半お世話になりました．秘書の井尾さんと佐野さんにはアルバイトや学会発表など様々な手続きをするにあたって，お世話になりました．また，同じ鈴木研究室の皆様には本研究について参考になる様々なご意見をいただきました．仲良くなるのにかなり時間がかかった気がしますが，楽しかったです．本論文を書き終えることができたのは，支えてくださった皆様のおかげです．心より感謝申し上げます．

参考文献

[1] Rich Caruana. Multitask learning. *Machine learning*, Vol. 28, No. 1, pp. 41–75, 1997.

[2] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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.

[4] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264. IGI global, 2010.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[6] Sotiris Lamprinidis, Daniel Hardt, and Dirk Hovy. Predicting news headline popularity with syntactic and semantic knowledge using multi-task learning. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 659–664, 2018.

[7] Yao Lu, Linqing Liu, Zhile Jiang, Min Yang, and Randy Goebel. A multi-task learning framework for abstractive text summarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 9987–9988, 2019.

[8] Andriy Mulyar, Ozlem Uzuner, and Bridget McInnes. Mt-clinical bert: scaling clinical information extraction with multitask learning. *Journal of*

the American Medical Informatics Association, Vol. 28, No. 10, pp. 2108–2115, 2021.

[9] Md Shad Akhtar, Dushyant Chauhan, Deepanway Ghosal, Soujanya Poria, Asif Ekbal, and Pushpak Bhattacharyya. Multi-task learning for multi-modal emotion recognition and sentiment analysis. In *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. 370–379, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[10] Long Duong, Trevor Cohn, Steven Bird, and Paul Cook. Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser. In *Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (volume 2: short papers)*, pp. 845–850, 2015.

[11] Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in Neural Information Processing Systems*, Vol. 33, pp. 5824–5836, 2020.

[12] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pp. 94–108. Springer, 2014.

[13] Chris Fifty, Ehsan Amid, Zhe Zhao, Tianhe Yu, Rohan Anil, and Chelsea Finn. Efficiently identifying task groupings for multi-task learning. *Advances in Neural Information Processing Systems*, Vol. 34, pp. 27503–27516, 2021.

[14] Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *International Conference on Machine Learning*, pp. 9120–9132. PMLR, 2020.

発表リスト

[1] 古田朋也，鈴木優『本文中単語の影響を考慮した読者の興味をひく記事タイトルの判定』東海関西データベースワークショップ 2021, 2021.
[2] Tomoya Furuta，Yu Suzuki『A Fact-checking Assistant System for Textual Documents』IEEE 4th International Conference on Multimedia Information Processing and Retrieval (IEEE MIPR 2021), 2021.
[3] 古田朋也，鈴木優『読者の興味を惹くかどうかと本文との整合性を考慮した記事タイトル作成支援』第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM2022), 2022.
[4] 古田朋也，鈴木優『Wikipedia 校閲時の事実確認作業における誤り箇所の自動推定』電子情報通信学会論文誌 D: 情報・システム, 2022.
[5] 古田朋也，鈴木優『マルチタスク学習を用いた商品タイトルの興味を惹く度合いの推定』東海関西データベースワークショップ 2022, 2022.
[6] 古田朋也，鈴木優『マルチタスク学習モデルにおける有効な補助タスクの選定』第 15 回データ工学と情報マネジメントに関するフォーラム (DEIM2023), 2023.