CE-001

言語モデルを用いた教師なしマルチソースドメイン適応と距離指標に基づくドメイン選択
Unsupervised Multi-Source Domain Adaptation using Language Model and Domain Choice based on Distance

藤井 巧朗† 濱上 知樹†
Takuro Fujii Tomoki Hamagami

1. はじめに

機械学習システムでは，多くの場合，訓練データとテストデータが同じ分布に従うと仮定するが，実際にはそれらは異なることがよくある．この訓練データとテストデータの分布が異なることをドメインシフトと呼び，性能低下を招く原因となる．ドメインシフトが存在する状況で，テストデータでのパフォーマンスを向上させるようにモデルを学習することをドメイン適応と呼ぶ．

近年，大規模コーパスで事前学習した BERT 等の言語モデルを，教師ありターゲットデータで Fine-Tuning することで，自然言語処理における様々なタスクにおいて SOTA を達成している．しかし，大規模事前学習済み言語モデルでも，訓練データとテストデータの分布が異なる場合にはドメインシフトが生じ，性能が低下する[2]．さらに，教師なしデータに関する最適な Fine-Tuning が存在しないという課題もある．

本研究では，これらの課題に対処するために，言語モデルを用いた教師なしドメイン適応に取り組む．まず，既存手法のドメイン適応の要因に関して，ソース・ターゲットドメイン間距離に着目し，実験と考察を行う．次に，この考察に基づいて，教師ありソースドメイン$\mathcal{D}_S$が 1 個だけ存在するというシングルソース(SS)ドメイン適応設定において，ソース・ターゲットドメイン間距離に着目した手法を提案し，実験により有効性を示す．最後に，複数の教師ありソースドメイン$\mathcal{D}_S^i(i = 1,2,\dots,M)$が存在するというマルチソース(MS)ドメイン適応設定において，ドメイン選択とドメイン間距離に着目した新たなフレームワークを提案し，実験により有効性を示す．

2. ドメイン適応

ドメインの定義は，ある特徴空間$\mathcal{X}$上の確率分布$\mathcal{D} = \{\mathcal{X}, P(X)\}$として定式化される[1]．教師なしドメイン適応設定では，教師ありソースドメイン$\mathcal{D}_S$と教師なしターゲットドメイン$\mathcal{D}_T$が存在する．つまり，ソースドメイン$\mathcal{D}_S$から観測されるデータは$X_S = \{x_S^j, y_S^j\}_{j=1}^{N_S}$であり，ターゲットドメイン$\mathcal{D}_T$から観測されるデータは$X_T = \{x_T^j\}_{j=1}^{N_T}$である．このソースドメインが 1 個だけ存在する場合をシングルソース(SS)ドメイン適応設定，M個のソースドメイン$(\mathcal{D}_S^1, \mathcal{D}_S^2, \mathcal{D}_S^3, \dots, \mathcal{D}_S^M)$が存在する場合をマルチソース(MS)ドメイン適応設定と呼ぶ．

教師なしドメイン適応の最終目標は，教師なしターゲットデータへのパフォーマンスを向上させることである．そして，教師なしドメイン適応の貢献としては，目的ドメインのラベル付きデータを収集するコストの低減，アノテーションプロセスに依存しない合理性などが挙げられる．

† 横浜国立大学大学院理工学府

3. 関連研究

大規模事前学習済み言語モデルの登場により，自然言語処理分野が大きく進展した．これは自然言語処理におけるドメイン適応についても同様であり，大規模事前学習済み言語モデルを用いた教師なしドメイン適応手法について多くの研究がなされている．大規模事前学習済み言語モデルを用いた教師なしドメイン適応手法は主に 2 つに分けられる．1 つ目は，事前学習に着目した手法である．DAPT[5]では，教師なしターゲットドメインのデータを用いて，事前学習済み BERT をさらに Masked Language Model(MLM)で追加の事前学習させる．2つ目は，Fine-Tuningに着目した手法である．UDALM[6]は，ソースデータ上のタスク損失とターゲットデータ上の MLM 損失を同時に最小化することで，モデルはタスクを学習しながら，ターゲットドメインの言語に適応することができる．損失関数は分類損失と MLM 損失の重み付き和$\mathcal{L} = \alpha\mathcal{L}_{cls} + \beta\mathcal{L}_{mlm}$で表される(図 1)．

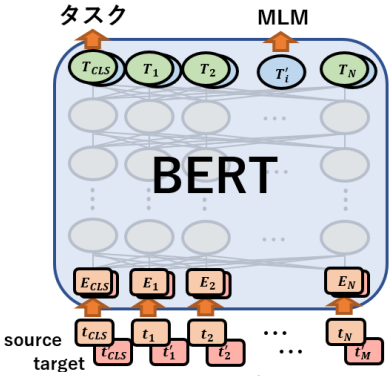

図 1: UDALM の概要

4. データ

データは，Amazon Review[3]から，Automotive(A)，Toys and Games(T)，Kindle Store(K)，Musical Instruments (M)の 4 ドメインを用いた．本データはそれぞれジャンルの商品に対するレビューテキストと 5 段階の評価値が含まれる．以下の実験では，評価値 3 以下のものをネガティブ，4 以上のものをポジティブとして，レビューテキストのネガポジ判定タスクを行う．訓練データには 4000 件，評価データには 1000 件のデータを用いた．また，SS ドメイン適応設定ではソース・ターゲットペアの合計 12 通り，MS ドメイン適応設定では合計 4 通りで，ターゲットドメインデータにおけるネガポジ判定の精度を比較対象とする．

5. 予備実験

5.1 実験概要

本実験は，提案手法で DAPT 及び UDALM を用いる動機付けのための予備実験であり，UDALM 及び DAPT の有効性の検証を行う．これは論文[6]の模倣実験であり，再検証である．

本実験は SS ドメイン適応設定であり，ソース・ターゲット 12 ペアで実験を行う．また，事前学習済み BERT には，bert-base-uncased を用い，ソースドメイン，ターゲットドメイン共にバッチサイズは 4 とし，学習は 50 エポックで，評価は accuracy により行う．

5.2 比較手法

以下の 3 手法を比較することで UDALM，DAPT の有効性を検証する．

Source only (SO) BERT: ソースデータのみを用いて，ネガポジ判定タスクで BERT を Fine-Tuning する．

UDALM: Fine-Tuning の際に，ソースデータでタスクを，ターゲットデータで MLM をマルチタスク学習することでドメイン適応を行う．

DAPT+UDALM (UDALM$_{\mathrm{DAPT}}$): 追加の事前学習 DAPT を行った後に，UDALM を行う．

5.3 実験結果

全ペアにおいて，UDALM が SO BERT を上回り，平均精度も 1.5 ポイント高いことから，UDALM の有効性が検証できた(表 1)．さらに，DAPT+UDALM が 8 ペアで UDALM を上回り，平均 accuracy も 0.6 ポイント高いことから，DAPT の有効性も確認できた．従って，以下の実験で UDALM を用いる動機が得られた．また，以下の実験で UDALM 用いる際は，断りがない限り追加の事前学習である DAPT を事前に行ってあるものとする．

表 1: 12 ドメインペアにおける各種教師なしドメイン適応手法の精度．DAPT 及び UDALM の有効性が確認できる．

$\mathcal{D}_S \to \mathcal{D}_T$	SO BERT	UDALM	UDALM$_{\mathrm{DAPT}}$
T → A	85.6	**87.3**	86.7
K → A	83.7	**86.0**	85.7
M → A	86.1	87.6	**89.0**
A → T	89.8	**91.2**	90.6
K → T	89.0	**90.0**	**90.0**
M → T	90.3	90.7	**91.7**
A → K	80.4	81.0	**83.0**
T → K	80.3	82.9	**84.3**
M → K	81.6	**84.9**	84.2
A → M	88.3	89.3	**90.8**
T → M	88.1	89.4	**90.6**
K → M	86.2	87.6	**88.4**
average	85.8	87.3	**87.9**

6. 距離指標を用いた分析

6.1 分析内容

実験 1 では UDALM の学習によりドメイン適応できることを示したが，ドメイン適応に起因する要素は不明である．ドメイン適応を考える上で，2 ドメイン間の距離，分布のアラインメント，相互情報量などからアプローチする場合があるが，本研究ではドメイン間距離に着目し，UDALM の学習に従って，$\mathcal{D}_S$と$\mathcal{D}_T$のドメイン間距離が近づく/遠ざかるなどの特徴があるのではないかと考えた．

UDALM で学習した BERT を用いて，$\mathcal{D}_S$からのデータX_Sのうち，予測ラベルがポジティブ及びネガティブである特徴表現をそれぞれ$h^j_{S,p} = BERT(x^j_{S,y=p})$，$h^j_{S,n} = BERT(x^j_{S,y=n})$と表し，$\mathcal{D}_T$からのデータ$X_T$についても，$h^j_{Tp} = BERT(x^j_{T,y=p})$，$h^j_{Tn} = BERT(x^j_{T,y=n})$と表す．以下，ドメイン間距離を測定する際に，次式に示すラベル別平均特徴量を用いる．

$$\mu_{Sp} = \frac{\sum h^j_{Sp}}{N_{Sp}} \quad \mu_{Sn} = \frac{\sum h^j_{Sn}}{N_{Sn}} \quad \mu_{Tp} = \frac{\sum h^j_{Tp}}{N_{Tp}} \quad \mu_{Tn} = \frac{\sum h^j_{Tn}}{N_{Tn}} \quad (1)$$

また，分析には，Euclid 距離，Cos 距離，MMD の 3 つを用いる．

6.1.1 Euclid 距離

ラベル別平均特徴量を用いて，2 ドメイン間の Euclid 距離を次式で定義する．

$$D_{euclid}(X_S, X_T) = \|\mu_S - \mu_T\|_2 \quad (2)$$

ただし，分析には，各ドメインにおいて，ポジティブと予測されたデータのラベル別平均特徴量μ_{Sp}, μ_{Tp}間の距離$D_p(\mu_{Sp}, \mu_{Tp})$，ネガティブと予測されたデータのμ_{Sn}, μ_{Tn}間の距離$D_n(\mu_{Sn}, \mu_{Tn})$の和$D_p + D_n$を 2 ドメイン間距離とする．

6.1.2 Cos 距離

内積空間におけるベクトル間の類似度を示す Cos 類似度は$S_{cos} = \frac{\mu_S \cdot \mu_T}{\|\mu_S\|_2 \|\mu_T\|_2}$で表され，Cos 距離は$1 - S_{cos}$により算出される．ここで，Cos 距離は厳密には距離ではないことに注意されたい．また，分析では，Euclid 距離と同様，$D_p(\mu_{Sp}, \mu_{Tp}) + D_n(\mu_{Sn}, \mu_{Tn})$を 2 ドメイン間距離とする．

6.1.3 Maximum Mean Discrepancy (MMD)

MMD は，異なる$\mathcal{D}_S$と$\mathcal{D}_T$がある時，分布$P_S(X), P_T(X)$からそれぞれ独立同一にサンプリングされたデータが同じ分布に従うかを仮説検定に用いられる．ある写像$\phi(\cdot)$を用いて，MMD は次式で定義される．

$$MMD^2 = \|\mathbb{E}[\phi(h_S)] - \mathbb{E}[\phi(h_T)]\|^2_{\mathcal{F}} \quad (3)$$

また，式(3)はカーネル$k(\cdot)$を用いて，次式のように求まる．

$$MMD^2(X_S, X_T) = \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(h_{S,i}, h_{S,j}) - \frac{2}{m \cdot m} \sum_i \sum_{j \neq i} k(h_{S,i}, h_{T,j}) + \frac{1}{m(m-1)} \sum_i \sum_{j \neq i} k(h_{T,i}, h_{T,j}) \quad (4)$$

本分析では，カーネルkに次式で示すガウシアンカーネルを用いる．

$$k(x, x') = \exp\left(-\frac{1}{2\sigma}|x - x'|^2\right) \quad (5)$$

ただし，Euclid 距離や Cos 距離とは異なり，MMD は全特徴量を用いて計算する．

6.2 分析結果 1

分析結果のうちターゲットが M のものを図 2 に示す．Euclid 距離と Cos 距離に関しては増減の特徴がみられないが，MMD(不一致度)は増加しているように見える．以上の結果から，UDALM の学習によるドメイン適応に寄与する可

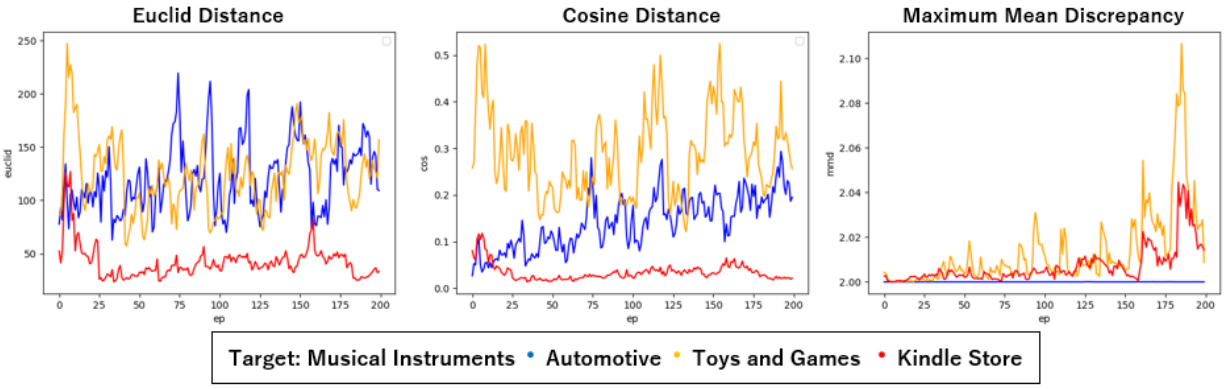

図 2：UDALM の学習中におけるドメイン間距離の挙動．明確な特性は見られなかった．

能性がある要因の 1 つとして，MMD(不一致度)の増加が考えられる．

他方で，Euclid 距離と Cos 距離は特徴がみられないのではなく，距離を正しく測定できなかった可能性がある．これは，本実験では BertForMaskedLM を用いており，出力が約 30000 次元と高すぎることが原因だと考えられる．

7. 提案手法 1(UDALM added Linear)

先の考察から，BERT の出力をさらに低次元にすることでドメイン間距離を正しく測定できるのではないかという仮説の下，UDALM に線形層を追加し，BERT の出力からタスクと MLM を，線形層の出力からタスクを同時にマルチタスク学習するモデルを提案する(図 3)．損失関数は 2 つの分類損失と MLM 損失の重み付き和$\mathcal{L} = \alpha\mathcal{L}_{cls1} + \beta\mathcal{L}_{cls2} + \gamma\mathcal{L}_{mlm}$で表される．また，最終的な分類ラベルは，線形層を通って出力されるものとする．

本手法の着想は，距離分析において BERT の出力が高次元であるために距離を正しく測定することができないという考察に基づくものである．また，キーとなる考えは，BERT はソース・ターゲット共通の特徴空間を，線形層はタスクの特徴空間を担うという点と，線形層による次元圧縮により，BERT で獲得した高次元特徴量のうち，タスクに重要な要素を抽出でき，ドメイン間距離を正しく測定できるようになるという点にある．

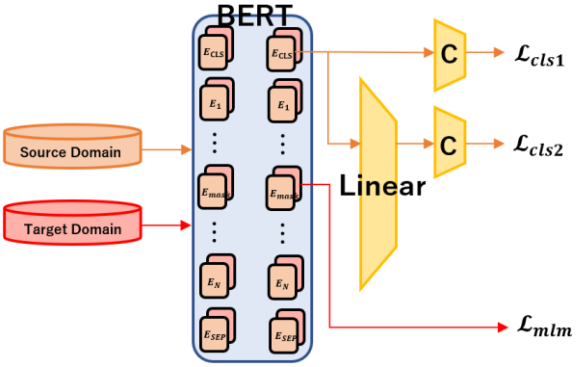

図 3：提案モデル(UDALM added Linear)の概要

7.1 実験概要

本実験では，予備実験で用いた 3 手法をベースラインとしてネガポジの精度を比較することで，提案手法 1 の有効性を示す．ただし，提案手法の線形層の出力は 100 及び 30 次元の 2 通りを試した．さらに，高次元であるためにドメイン間距離が正しく測定できなかったという考察を確かめるべく，項 6 と同様の距離分析も行う．

実験設定は実験 1 と同様で，SS ドメイン適応設定，ソース・ターゲット 12 ペア，バッチサイズはソース・ターゲットドメイン共に 4，学習は 50 エポック，評価は accuracy である．

7.2 実験結果

線形層の出力が 100 次元の場合，実験 1 の最高精度と比較して，8 ペアで上回り，2 ペアで等しく，平均精度も 0.5 ポイント上回るという結果になった(表 2)．さらに，線形層の出力が 30 次元の場合，実験 1 の最高精度と比較して，10 ペアで上回り，平均精度も 0.9 ポイント上回るという結果になった．したがって，共に実験 1 の手法よりも優れていると言える．

さらに，距離分析の結果から，学習するにしたがって，ユークリッド距離は大きくなり，Cos 距離と MMD は小さくなっている(図 4)．従って，線形層により低次元に圧縮することで，30000 次元の特徴量では影響が小さかった重要

表 2：12 ドメインペアにおける提案手法の精度．表 1 と比較して精度が向上．また，100 次元よりも 30 次元の方が高精度．

$\mathcal{D}_S \rightarrow \mathcal{D}_T$	UDALM Linear100	UDALM Linear30
T → A	87.3	**88.7**
K → A	**87.9**	**87.9**
M → A	89.0	**89.6**
A → T	**91.8**	91.3
K → T	90.7	**91.6**
M → T	91.0	**91.4**
A → K	83.1	**83.9**
T → K	**85.0**	84.5
M → K	84.7	**85.0**
A → M	90.3	**90.9**
T → M	**90.7**	90.4
K → M	89.7	**90.0**
average	88.4	**88.8**

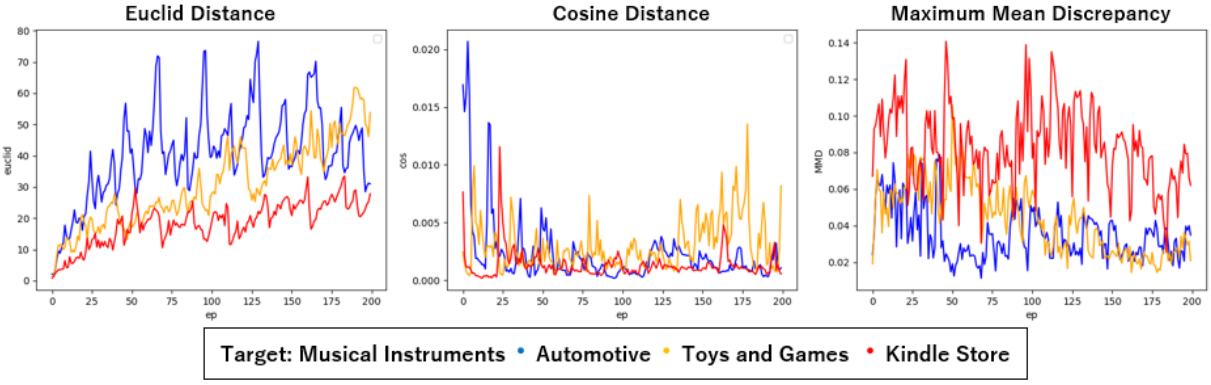

図 4: 提案モデルの学習中におけるドメイン間距離の挙動．Euclid 距離は増加傾向，Cos 距離は減少傾向の特性．

な要素が抽出され，ドメイン間距離がより正確に測定できていることが明らかになった．

8. 提案手法 2(Bandit based Distance Algorithm)

先の実験では，UDALM に線形層を追加し，BERT の出力から得られる特徴量を次元圧縮することで，重要な要素を抽出することができ，ドメイン間距離を正しく測定できる可能性があることを示した．次に，このドメイン間距離をモデルに組み込んだ手法を提案する(図 5(c))．また，本手法は，ソースドメインが複数存在する MS ドメイン適応設定で用いる．

本手法は，MS ドメイン適応設定において，各エポックでターゲットデータの精度に最も貢献しそうなソースドメインを 1 つ選択し，提案手法 1 で提案したモデルを学習させる．ここで，ドメイン選択問題をバンディット問題として捉え，ドメイン間距離を報酬とすることで，ソースドメインを選択する．例えば Euclid 距離の場合，ドメイン間 Euclid 距離が小さいものを選択し，モデルを学習することで，最終的に全ドメイン間 Euclid 距離が大きくなり，SS ドメイン適応設定よりも精度が向上することが期待できる．また，Cos 距離の場合，ドメイン間 Cos 距離が大きいものを選択し，モデルを学習することで，最終的に全ドメイン間 Cos 距離が小さくなり，SS ドメイン適応設定よりも精度が向上することが期待できる．

8.1 実験概要

本実験では，2 つのベースライン(図 5(a)(b))とネガポジの精度を比較することで，提案手法 2(Bandit based Distance)の有効性を示す．ただし，先と同様，線形層の出力は 100 次元及び 30 次元の 2 通りを試した．また，距離報酬に Euclid 距離，Cos 距離，MMD を用いた．そして，Euclid 距離の場合は距離の小さいドメインを選択し，Cos 距離と MMD の場合は距離の大きいドメインを選択するようにした．

実験設定は MS ドメイン適応設定であり，3 つのソースドメインと 1 つのターゲットドメインの計 4 ペア，学習は 50 エポック，評価は accuracy で行う．ただし，バッチサイズについては，入力するデータ数に大きな差がでないように，Mixture と Bandit based Distance ではソース・ターゲットドメイン共にバッチサイズを 4，Multi ではターゲットドメインのバッチサイズを 4，各ソースドメインのバッチサイズを 1 とした．

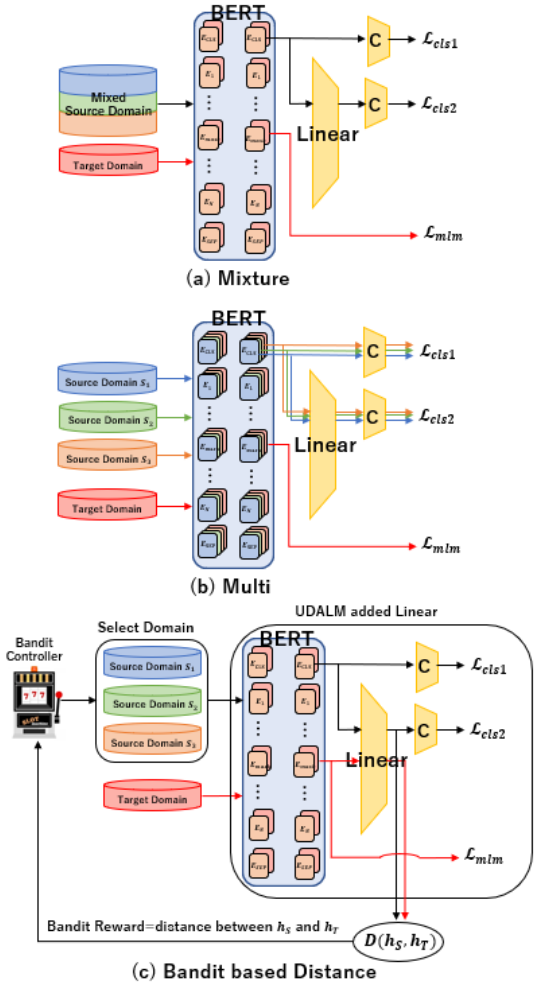

図 5: 実験 3 の比較手法．(a)(b)はベースライン，(c)は提案手法

8.2 比較手法

Multi：各ドメインのデータを UDALM added Linear に入力する．ただし，ソースドメインのバッチサイズを 1，ターゲットドメインのバッチサイズを 4 とした．

Mixture：全ソースドメインを混合して 1 つのドメインにし，SS 設定として UDALM added Linear に入力する．

Bandit based Distance：前述した提案手法であり，距離報酬には Euclid 距離，Cos 距離，MMD を用いる．

表 2：4 ドメインペアにおけるベースライン及び提案手法の精度．100, 30 次元の両方で提案手法がベースラインを上回る．

	UDALM added Linear100					UDALM added Linear30				
$\mathcal{D}_S \rightarrow \mathcal{D}_T$	Multi	Mixture	Euclid	Cos	MMD	Multi	Mixture	Euclid	Cos	MMD
T,K,M → A	89.3	**89.6**	89.3	89.2	88.9	88.4	**89.4**	**89.4**	89.2	88.8
A,K,M → T	91.2	91.1	**92.6**	91.6	91.6	90.7	91.5	**92.3**	92.2	91.9
A,T,M → K	83.7	84.7	84.6	**85.4**	85.0	83.7	85.1	**85.2**	84.8	85.1
A,T,K → M	90.6	89.3	**91.1**	91.0	**91.1**	89.9	90.1	91.0	**91.5**	**91.5**
average	88.7	88.7	**89.4**	89.3	89.2	88.2	89.0	**89.5**	**89.4**	89.3

8.3 実験結果

UDALM added Linear の線形層の出力が 100 次元の場合，多くのソース・ターゲットペアにおいて，3 つの距離報酬を適用した Bandit added Distance がベースラインの Multi と Mixture を上回り，平均手法も 3 種類ともベースラインを上回っている(表 3)．さらに，線形層の出力が 30 次元の場合も同様に，3 つの Bandit added Distance はどれもベースラインを上回っている．以上から，提案手法の有効性が確認できた．

また，線形層の出力が 100 次元と 30 次元の両方で，Euclid 距離報酬を適用した Bandit based Distance が最高精度を獲得しており，次いで Cos 距離，MMD という順位になっている．これは，実験 2 の距離分析の結果に従っているといえる．つまり，Euclid 距離，Cos 距離，MMD の順に，ソース・ターゲット間ドメイン間距離の特性が大きく出ていることに一致している．従って，UDALM を用いたドメイン適応には Euclid 距離の増加と Cos 距離の減少が大きな要因となっているといえる．

最後に，線形層の出力が 100 次元と 30 次元の結果を比較すると，全 5 手法において 30 次元の方が高い精度となった．これは実験 2 の結果に一致しており，MS ドメイン適応設定においても 30 次元の方が正しくドメイン距離を測定できると考えられる．

9. おわりに

本稿では，距離分析により，Euclid 距離と Cos 距離がドメイン適応に大きく貢献する要素であることを明らかにし，シングルソースドメイン適応設定における既存手法である UDALM の改良手法，マルチソースドメイン適応設定における距離報酬に基づくドメイン選択アルゴリズムを提案し，その有効性を示した．

今後は，UDALM added Linear の Bert の出力と Linear 層の出力の両方を用いる手法の検討，損失関数に距離指標を導入する手法の検討，Bandit based Distance の距離報酬に Euclid 距離と Cos 距離の両方を利用したアルゴリズムの検討，他タスクへの拡張などを行うことを計画している．

参考文献

[1] Alan Ramponi， Barbara Plank : Neural Unsupervised Domain Adaptation in NLP-A Survey(2020)， Proceedings of the 28th International Conference on Computational Linguistics, pp.6838–6855

[2] A.Fisch, A.Talmor, M.Seo, E.Choi, and D.Chen. MRQA 2019 shared task: Evaluating generalization in reading comprehension. EMNLP, pp. 1–13, 2019.

[3] Julian McAuley,Amazon product data,University of Calfornia,San Diego,2018

[4] Yongchun Zhu, Fuzhen Zhuang, Deqing Wang : Aligning Domain-Specific Distribution and Classifier for Cross-Domain Classification from Multiple Sources(2022)，The Thirty-Third AAAI Conference on Artificial Intelligence，pp.5989-5996

[5] Suchin Gururangan，Ana Marasovic，Swabha Swayamdipta，Kyle Lo，Iz Beltagy，Doug Downey，Noah A. Smith：Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks(2021)，Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp.8342–8360

[6] Constantinos Karouzos，Georgios Paraskevopoulos，Alexandros Potamianos：UDALM: Unsupervised Domain Adaptation through Language Modeling(2021)，Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.2579–2590

[7] *Dustin Wright，Isabelle Augenstein：Transformer Based Multi-*Source Domain Adaptation(2020)，Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp.7963–7974,

[8] Han Guo，Ramakanth Pasunuru，Mohit Bansal ：Multi-Source Domain Adaptation for Text Classification via DistanceNet-Bandits(2020)，Proceedings of the AAAI Conference on Artificial Intelligence，pp.7830-7838

[9] Shai Ben-David，John Blitzer，Koby Crammer，Alex Kulesza,，Fernando Pereira, and Jennifer Wortman Vaughan：A theory of learning from different domains(2010)，Machine learning, 79(1-2):151–175.

[10] K.Beyer，J.Goldstein，R.Ramakrishan，U.Shaft：When is Nearest Neighbor Meaningful ? (1999) Poc.7th International Conference on Database Theory – ICDT’99. LNCS. Vol. 1540. pp.217-235

A. 付録

A.1 ドメイン間類似度分析

簡易的なドメイン間の類似性の分析を行う．ドメインに関与するものとして，テキストの長さ，出現単語が挙げられる．

まず，ドメイン毎のテキスト長の分布を図 6 に示す．ただし，BERT の入力は最大で 512 トークンなので，512 トークンより長い分は切り捨てた．図 4(a)より，A と K のピークの位置がずれており，T と M は A と K の中間的な分布となった．つまり，A と K の類似度が低く，T と M はそれぞれ他ドメインとの類似度が高いという結果になった．

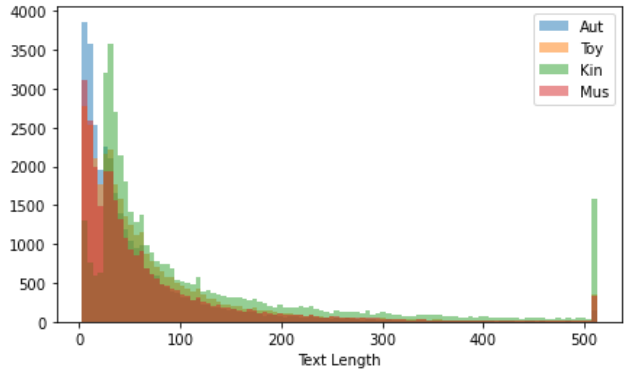

図 6: ドメイン間類似度分析結果．テキスト長

次に，出現上位 10000 単語の重複率を図 7 に示す．Kin のみ他ドメインとの重複率が 7 割を下回っており，その他は 7 割 5 分を上回っていることが分かる．したがって，Kin のみ他ドメインとの類似度が低いという結果になった．

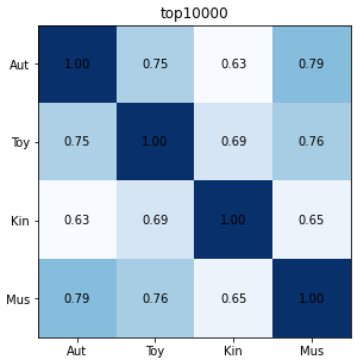

図 7: ドメイン間類似度分析結果．単語重複率

この分析の妥当性は，実験 1 の結果(表 1)から確認できる．ターゲットドメインが K の場合，accuracy が 85 を上回っているものがない一方で，他ドメインは 9 割前後をマークしており，明らかに K への適応度が低いことが分かる．さらに，K 以外のターゲットドメインに関して，ソースドメインが K の場合，他のソースドメインと比較して，accuracy が最も低いことが分かる．以上のことから，K は他ドメインとの類似度が低いために，適応性能が低くなったと考えられる．これは，テキスト長及び単語重複率の分析結果と一致しているため，この分析手法がドメイン類似度の指標の 1 つとなり得ると言える．