自然で不均衡な擬似ラベルからの脱バイアス学習

------------------------------

実世界の観測は、キュレーションされていないデータセットと同様に、自然にロングテール分布している [18, 56]。不均衡分類[10, 24, 59]は、このようなデータの偏りに対処し、モデルがヘッドクラスのインスタンスに支配されることを防ぐ。また、バイアスに対抗できる視覚認識システムの開発は、社会的に大きな影響を与える[34]。既存の方法は、人間の注釈者が収集した不均衡なグランドトゥルースラベルからの脱バイアス学習に焦点を当てているが、我々は、機械学習モデルが生成する疑似ラベルが自然に不均衡であることを発見し、広く偏った学習のための別のソースを作成します！擬似ラベルとは、既存の（教師）モデルがラベルのないデータに対して行った確信度の高い予測のことで、ラベルのないデータへの（生徒）モデルの適応を監督するための学習データの一部となる（図1a）。図1a）。Long Lian$^{1}$ Stella X. Yu$^{1}$ $^{2} Microsoft Research

------------------------------

擬似ラベルとは、ラベル付きソースデータで学習した分類器が、ラベルのないターゲットデータに対して行う確信に満ちた予測である。半教師付き学習などで、ラベルのないデータにモデルを適応させるために広く利用されている。我々の重要な洞察は、擬似ラベルは、バランスのとれたソースデータで学習し、バランスのとれたターゲットデータで評価した場合でも、内在するデータの類似性により、自然に不均衡になることです。もし、この未知の不均衡な分類問題に対処することができれば、擬似ラベルによって生じる偽の多数派へのモデルの偏りを除去することができる。我々は、反実仮想推論と適応的マージンに基づく、擬似ラベルを用いた新規かつ効果的な脱バイアス学習法を提案する： 前者は分類器の応答バイアスを除去し、後者は擬似ラベルのアンバランスに応じて各クラスのマージンを調整する。広範な実験によって検証された結果、我々の脱バイアス学習は、ImageNet-1Kの最新技術を大幅に上回る精度を達成した。0.2%の注釈を用いた半教師付き学習では26%、ゼロショット学習では9%である。コードは https:／／github.com／frankxwang／debiased-pseudo-labeling で公開されています。

------------------------------

図1. 我々は、擬似ラベリングに基づく準教師付き学習(SSL)と伝達型ゼロショット学習(ZSL)を研究している。両タスクは、ラベル付きソースデータから学んだ意味情報を擬似ラベリングによってラベルなしのターゲットデータに伝達する必要がある。驚くべきことに、SSLとZSLの典型的な手法（FixMatch [53]やCLIP [45]）によって生成されたターゲットデータの疑似ラベルは、ソースデータとターゲットデータの両方がクラスバランスされていても、同じドメインからサンプルされていても、非常に偏りがあることがわかった。

------------------------------

生徒モデルは教師モデルそのものであり、学習プロセスは自己訓練とも呼ばれる [4, 5, 29, 53, 65]。疑似ラベリングは、半教師付き学習（SSL）[32, 53]、ドメイン適応[25, 37]、転移学習[1]で広く用いられている。

------------------------------

の課題を解決した。1）ソースとターゲットのドメインが異なるゼロショット転移学習（ZSL）では、事前に学習したCLIPモデル[45]は、CLIPの学習セットはほぼバランスされているものの、キュレーションされバランスされたImageNet-1Kデータセット上で非常に不均衡な予測を行う（図1c）。2)半教師付き学習において、ソースとターゲットのドメインが同じ場合、ラベル付きCIFAR10画像で学習したFixMatch [53]は、ラベル付きとラベルなしのセットはバランスしているが、ラベルなしの画像で非常に偏った擬似ラベルを生成する（図1b）。

------------------------------

これは、人間が作成した真実のラベルと同じように、不均衡なものである。もし我々が、真実のラベルではなく、擬似ラベルに起因するこの未知の不均衡な分類問題に取り組むならば、擬似ラベルに基づくモデル学習を改善し、擬似ラベルが作り出す偽大勢へのモデルバイアスを除去できるだろう。本研究では、真実のラベルを用いた脱バイアス学習[22, 33, 57]で容易に利用できる、実際の分類マージンの分布に関する知識を必要としない、疑似ラベルを用いた新規かつ効果的な脱バイアス学習法を提案する。適応的デビアスモジュールと適応的マージンロスから構成される。前者は反実仮想推論により分類器の応答バイアスを動的に除去し、後者は擬似ラベルの不均衡に応じて各クラスのマージンを動的に調整する。広範な実験により検証された結果、我々の脱バイアス学習は、ImageNet-1Kにおいて、0.2%の注釈を持つSSLで26%、ZSLで9%の最新技術を改善するだけでなく、ドメインシフトに対してより頑健な、様々な疑似ラベリング手法への汎用アドオンであることが示された。ラベルのない生データが自然に不均衡な場合、不均衡な擬似ラベルの問題はさらに深刻であり、モデルはテールクラスのサンプルをヘッドクラスと誤ラベルする傾向がある。脱バイアス学習を適用することで、ロングテール設定下でのSSL性能を大きく向上させることができる。我々の研究は、4つの主要な貢献をしている。1) 擬似ラベルが自然に不均衡になり、偏った学習を引き起こすことを系統的に調査・発見する。2) 疑似ラベルを用いた脱バイアス学習法を提案し、実際の分類マージンを知る必要がない。3) ZSL／SSLを大幅に改善し、本手法が様々な擬似ラベルモデルに適用可能であることを実証する。4) CLIPのような視覚と言語の事前学習済みモデルを適用するための新しい効果的なZSL／SSLパイプラインを確立した。

------------------------------

半教師付き学習は、限られたラベル付きデータからモデルを学習するために、ラベルなしデータを統合するものです。このアプローチには4つの系統がある。1)一貫性に基づく正則化手法により、非ラベル化データに対して摂動による分類不変性損失を課す [36, 51, 55, 64]. 2) 擬似ラベル付けは、モデルの学習データをラベル付きデータから、ラベルは付いていないが確信を持って擬似ラベルを付けたデータに拡張する [4, 5, 29, 31, 53, 65]。3) 転送学習は、まず自己教師付き表現学習（例えば対比学習）によってラベルのない大規模データでモデルを学習し、次に教師付き分類器学習によってラベルのある小規模データで学習する[2, 13]．4) データ中心SSLは、ラベル付きデータは与えられないが、ラベル付きデータの中から最適に選択してラベル付けを行うことができると仮定する[60]。このラベル付きデータの選択という現実的な問題に着目することで、SSLに大きな利点をもたらすことが判明した。CReST[62]は、ラベル付きデータの分布に従って少数クラスに対してより頻繁にサンプリングするクラスリバランスサンプラーを活用することで、クラスバランスの悪いデータに対する既存のSSL手法を改良する。CReSTは、ラベル付きデータがバランスしている場合には機能しない。一方、本アプローチでは、ラベル付き集合の事前分布を仮定していない。SSLでは先行文献が大きな成果を上げているが、SSLにおける暗黙の偏った擬似ラベリングの問題はこれまで知られておらず、十分に分析されていなかったため、しかし、学習効率に大きな影響を与える。本研究では、この重要な問題を解決するために、シンプルかつ効果的なデビアスモジュールを提案することに焦点を当てる。

ゼロショット分類とは、ゼロショットモデルが、新規クラスの画像を、学習中にモデルが見たことのない正しいカテゴリーに分類する問題設定を指す [44, 47, 58]。様々な観点から、いくつかの戦略が検討されている： 1) 手作業で作成された属性 [15, 27]; 2) クラスの意味的記述という形で事前知識を組み込んだ事前学習済み埋め込み [16, 52]; 3) 知識グラフを用いて見たクラスと見ていないクラスの関係をモデル化 [23, 38]; 4) 視覚言語モデルで一般的な視覚概念を学習し、下流の様々な分類課題にモデルをゼロショット転送できる [8, 45].

Long-Tailed Recognition (LTR)は、多くのインスタンスが利用可能な「Many-shot」クラスでの性能を犠牲にすることなく、少数のインスタンスを持つクラスに対して正確な「Few-Shot」モデルを学習することを目的としている。1) 再バランス／再重み付け法 τ -norm [24] は、テールクラスをより重要視することでLTR問題に取り組む。2) マージンベース法 LDAM [10] は、ラベル分布を考慮したマージン損失を提案し、テールクラスのマージンを大きくすることで少数クラスの汎化を向上させる； 3) post-hoc adjustment approachは、LA [35]のようにクラス分布の事前知識に従って学習済みモデルの予測を修正したり、Causal Norm [54]のように運動量の逆説的効果を除去して直接的な因果関係を追求する。 4) ensemble-based approach RIDE [59] は、複数の多様なエキスパートと動的エキスパートルーティング・モジュールを最適化して、ロングテールデータでのモデルの偏りと分散を低減する。クラス分布の事前知識を必要とするか、学習済みモデルに事後的に適用されるLTRに関する先行研究とは全く対照的に、提案するdebiasモジュールは事前知識を必要とせず、これまで知られていなかった偏った疑似ラベルの問題に焦点を当てる。

------------------------------

3. 疑似ラベルは自然にアンバランスになる

------------------------------

不均衡なデータで訓練した場合に生じるバイアスに集中していた以前の研究とは対照的に、我々は、バランスのとれたデータで訓練した場合であっても、擬似ラベルのバイアスに焦点をあてている。本節では、SSLにおけるFixMatch [53]とZSLにおけるCLIP [45]の大成功の裏に隠された、この未知の問題の分析を行う。まず、擬似ラベリングアプローチの背景を説明し、次にそのバイアス問題を分析する。バイアスの原因はクラス間相関の問題であるとする。

------------------------------

図2. FixMatchの擬似ラベルは、学習するラベルなしデータとラベル付きデータがクラスバランスされているにもかかわらず、異なる学習段階において非常に不均衡であることがわかる。一方、DebiasPLは、後半になるとほぼバランスのとれた擬似ラベルを生成する。FixMatchとDebiasPLの確率分布は、すべてのラベルなしデータに対して平均化されたものである。クラス指標は平均確率でソートされている。CIFAR10において、1クラスあたり4個のラベル付きインスタンスで実験を行った。

------------------------------

半教師付き学習のためのFixMatch。FixMatch [53]の核となる技術は擬似ラベリング [29]である。これは、信頼度の高いラベルのないサンプルを学習対象として選択するものである。L 個のラベル付きインスタンスを持つラベル付きデータセット X$\_{L}$ = {(x$\_{i}$ , y$\_{i}$ )}$\_{i=1}^{L}$ と、U個のインスタンスを持つラベルなしデータセット X$\_{U}$ = {(x$\_{i}$ )}$\_{i=L+1}^{L+U}$ があるとする．x$\_{i}$は入力インスタンス、y$\_{i}$ = [y$\_{i}^{1}$ , ..., y$\_{i}^{C}$ ] ⊆ {0, 1}$^{C}$ はC個のクラスを持つ離散アノテーション対象である。X$\_{U}$とX$\_{L}$は、同じ意味ラベルを共有する。最適化目的は2つの項よりなる： L = L$\_{s}$ + λ$\_{u}$ L$\_{u}$ 、すなわち、ラベル付きデータに適用する教師あり損失L$\_{s}$とラベルなしデータに適用する教師なし損失L$\_{u}$、λ$\_{u}$はスカラーハイパーパラメーターである。教師ありの損失L$\_{s}$は、モデルの予測値とグランドトゥルースとの間のクロスエントロピーである: L$\_{s}$ =

------------------------------

弱補正された非ラベル化サンプルからは、信頼できるインスタンスが生成され、強補 正されたサンプルのモデル予測を監視するために使用される。最大の確率が信頼度閾値τ以下になるインスタンスは、信頼できないサンプルとみなされ、破棄される。形式的には、教師なし損失L$\_{u}$は以下のように定式化できる：

------------------------------

ここで、βは強い増強[14]であり、μはミニバッチにおけるラベル付きサンプルとラベルなしサンプルの比率を決定する。

CLIPによるゼロショット学習 CLIP [45]は、クエリバランスを考慮した400Mの画像-テキストペアのデータセットに対して、画像表現をゼロから学習する効率的でスケーラブルな方法である。事前学習時に、画像エンコーダとテキストエンコーダは、対になった（対になっていない）キャプションと視覚的画像の間の類似度を最大（最小）にすることで最適化される。未ラベルデータの擬似ラベル作成には、自然言語によるプロンプトを使用し、ターゲットデータセットへのゼロショット転送を可能にする： CLIPは、ターゲットデータセットのクラスの名前または説明を、潜在的なテキストのペアリングのセットとして使用し（例えば「犬の写真」）、画像とテキストのペアのコサイン類似度に従って、最も可能性の高いクラスを予測する。具体的には、まず、画像の特徴埋め込みと、可能性のあるテキストの集合の特徴埋め込みが、それぞれのエンコーダによって計算される。そして、これらの埋込みのコサイン類似度を評価し、ソフトマックス関数によって確率分布に正規化する。

------------------------------

図3. 1.3MのImageNetインスタンスに対して、事前に訓練されたCLIPを用いた擬似ラベル予測のクラスごとの精度とリコール。リコールが高い大多数のクラスは、精度の低い擬似ラベルを持つことが多い。

------------------------------

図4. CLIPのゼロショット予測は、様々なデータセットやベンチマークにおいて、非常に偏りがあることがわかります。

------------------------------

3.2. 半教師付き学習における偏り

------------------------------

図2は、様々な学習エポックにおいて、全てのラベルなしデータで平均化されたFixMatch確率分布を可視化したものです。驚くべきことに、ラベル付きデータとラベルなしデータが共にキュレーションされている（クラスバランスが取れている）場合でも、擬似ラベルのクラスバランスは非常に悪く、特に訓練初期に顕著である。訓練が進むにつれて、この状況は続きます。生徒モデルは暗黙のうちに不均衡な擬似ラベルを継承し、今度は教師モデルのバイアスを強化することになります。一度混乱したサンプルに誤った擬似ラベルが貼られると、その誤りを自己修正することはほとんど不可能です。それどころか、モデルを誤解させ、既存のバイアスをさらに増幅させ、より間違った予測を生み出す可能性さえあります。介入しなければ、モデルは修復不可能なバイアスに陥ってしまう。これに対し、図2のように、DebiasPLも最初は不均衡な擬似ラベルに悩まされますが、この状況はかなり緩和され、最終的にはモデルを動的にデビアスすることでほぼバランスの取れた分布を得ることができるのです。

------------------------------

図5. ImageNetの低頻度クラスは、クラスあたりのCLIP予測数が10個と少ないため、通常クラス間相関が強く、高頻度クラスはその逆である。各クラスの画像埋め込みセントロイドと、最も近い9つの「ネガティブ」クラスの埋め込みセントロイドのコサイン類似度を比較します。(拡大図)

------------------------------

飛行機 車 鳥 猫 鹿 犬 蛙 馬 船 トラック

------------------------------

飛行機 車 鳥 猫 鹿 犬 蛙 馬 船 トラック

------------------------------

図6. 擬似ラベルの偏りの原因は、部分的にはクラス間の交絡に起因していると考えられる。例えば、FixMatchは "ship "を "plane "と誤分類してしまうことがあります。FixMatchの擬似ラベルと我々のDebiasPLの擬似ラベルの混同行列を可視化したものである。

------------------------------

3.3. ゼロショット学習におけるバイアス

------------------------------

CLIPは実際にImageNet上で非常に偏った予測を生成しており、これはCLIPの全体的なゼロショット予測精度という驚異的な成功の裏に隠されています。不均衡の問題を除けば、図3に示されるように、多くの高頻度クラスの精度とリコールは、多くの中／少数ショットクラスよりはるかに低い。信頼度スコアに基づいてCLIPの予測値を閾値化することは有効かもしれません。しかし、単に信頼度スコアの閾値を高く設定するだけでは、さらに不均衡な分布になる可能性があります（詳細は付録をご覧ください）。アンバランス率と精度／再現性の間にはトレードオフがある。偏りの大きいゼロショット予測は、ImageNetに限ったことではありません。図4に示すように、EuroSAT [20], MNIST [28], CIFAR10 [26], CIFAR100 [26], Food101 [7] など、多くのベンチマークで広く存在している。

------------------------------

擬似ラベルの偏りの原因を掘り下げるために、クラス間相関の分析を行う。CLIPでは、まず、事前に訓練されたCLIPモデルの画像エンコーダによって抽出された、このクラスに属する正規化画像特徴の平均を取ることによって、クラスごとに1つの画像セントロイドを計算する。

------------------------------

予測頻度上位10位／下位10位のクラスの画像セントロイドと、それに最も近い「混同」クラスのコサイン類似度を可視化したもの。予測混同は、クラスレベルでの画像の類似性を示している。図5から、ImageNetの低頻度クラスは、クラスあたりのCLIP予測数が最も少ない10クラスであり、通常、クラス間の混同が強いことがわかります。

図6aはFixMatchの擬似ラベルの混同行列を示したものです。例えば、"ship "は "plane "と誤分類されることが多い。クラス間相関の分析から、擬似ラベルの偏りの原因は、図6bのようにクラス間交絡（こうらく）に大きく起因すると考えられ、提案するDebiasPLはこれにうまく対処することができます。DebiasPLについては次節で紹介する。

------------------------------

本節では、Debiasased Pseudo-Labeling（DebiasPL）と、それをZSLおよびSSLタスクに組み込む方法を紹介する。なお、今回提案するシンプルかつ効果的なアプローチは、ここで紹介するものに限らず、様々なネットワークやベンチマークに普遍的に適用可能である。

-----------------------------

-我々のDebiasPLアプローチは、異なる分布に従うソースとターゲットデータにさらされた場合でも、限界クラス分布に関する事前知識を活用することなく、生徒モデルに対する偏った疑似ラベルの影響を動的に緩和することを目的としている。この目的を達成するために、反事実推論と適応的な限界損失を持つ適応的デビアスモジュールが提案されている（次に説明）。適応的デビアスw／反実仮想推論。因果推論とは、事実のみを前提として、因果関係のあるグラフモデルで変数間の介入を表現し、反実仮想の結論を導き出すことである [17, 40, 42, 48, 49]. 因果推論は、ほとんどすべての実証研究に蔓延している選択バイアスの除去 [3]、因果的介入を用いた交絡効果の除去 [67]、反事実推論による望ましい直接効果の分離 [6] など、さまざまなタスクで広く研究・適用されてきた。

------------------------------

図7. FixMatchの上部に追加された、提案するAdaptive DebiasingモジュールとAdaptive Marginal Lossの構成図。

------------------------------

図8. 反実仮想推論によるデビアスの因果グラフ。

------------------------------

このことに動機づけられ、望まないバイアス（反実仮想）の影響を動的に緩和するために、反実仮想推論を通じて、不正確な予測を生成する因果関係を組み込む[21, 40-43]。図8の因果関係グラフが提案されたとき、私たちは、偏った予測を生成するための目標を明確にすることができます：A$\_{i}$ → Yに沿って直接的な因果効果を追求することで、Controlled Direct Effect (CDE) [17, 42, 43, 46, 54]として定義します：

------------------------------

すなわち、A = A$\_{i}$ で曝露した場合の反実仮想結果と、A = Â = {A$\_{1}$ , ..., A$\_{n}$ } で曝露し、媒介者を固定レベルDとした場合の反実仮想結果との対比を示す。CDE[17,42]は、モデルの偏りを反実仮想の世界で分離するもので、モデルの偏りは、A = Âの時はY 'の間接効果とみなされるが、A = A$\_{i}$の時はDの値を維持する。 しかし、すべての訓練サンプルを訪問して反実仮想結果を測定するには、著しく計算量がかかる。そこで、ACDE(Aximated Controlled Direct Effect)を用いる。ACDEは、モデルのバイアスが大きく変化しないことを前提とするため、勢いよく更新された反事実結果（式4）は、実際の［Y$\_{i}$ ｜do( Â), do(D) ］の近似として用いることができる。後に擬似ラベリングを行う（すなわち、式1のp(α(x$\_{i}$ ))を置き換える）ために用いられる、反事実推論を用いたデビアスロジットは、定式化することができる：

------------------------------

m∈ [0, 1]はモメンタム係数、f (α(-))は弱補正未ラベル化インスタンスのロジット、p$\_{k}$はソフトマックス関数により得られるインスタンスα(x$\_{k}$ )の確率分布である。 λは間接効果の強さを制御するデビアスファクターを表す。デビアス因子が強すぎると、モデルがデータにフィットしにくくなり、小さすぎるとバイアスをほとんど排除できず、結局、一般化能力が損なわれる。ロジットのスケールは不安定であり、特に初期の学習段階では、式3の第2項にロジットベクトルを直接使うのではなく、確率分布p$\_{k}$を使う。 log関数を使ってp̂をロジットの大きさに合わせて再スケールする。式3はロングテール認識における再重み付けやロジット調整法と関連付けられるが、我々のものは動的に適応的である。

適応的マージナルロス ３節で述べたように、擬似ラベルの偏りはクラス間の混同によって部分的に引き起こされる可能性がある。このため、適応的マージンロスを適用し、偏りの少ないクラスと偏りの多いクラスの間のマージンを大きくすることで、モデルの偏りが大きいクラスのスコアが他のカテゴリを圧倒することがないようにする。また、動的なクラス固有のマージンを強制することで、クラス間の混乱を大きく抑制することができ、これは図6でさらに実証的に証明されている。L$\_{AML}$は次のように定式化できる：

------------------------------

ここで、j∈{1, ..., C}に対して△$\_{j}$ = λ log($\_{p̂j}^{1}$ )、z = f (β(x$\_{i}$ )) とする。L$\_{AML}$を用いて、式1のH( y ˆ$\_{i}$ , f (β(x$\_{i}$ ))を置換する。オプション）確率の低いラベルなしインスタンスは最終的な損失に寄与しない。我々は、ラベル付けされていないインスタンスに、クロスレベルインスタンスグループ識別損失CLD [61]を適用して、その情報を十分に活用することが有益であると考える。

------------------------------

4.2. オルタナティヴとの区別と繋がり

------------------------------

LA、LDAM、Causal Normの紹介は第2項を参照されたい。SSL分布整列（DA）[4]でよく採用される別の手法も比較する。これは、モデルの予測値の実際のマージナル分布が、実際のマージナルクラス分布と一致するように促すことを目的としています。表1を参照してください。1では、分布の不一致とロングテール認識を扱うこれらの代替手法の重要な特性における区別と関連性を、Tab. 2、Tab. 3では、実験結果を比較しています。DebiasPLの成功には、反実仮名推論による擬似ラベルのデビアスと適応的限界損失の適用に勢いよく更新されたp↪Mn\_302を用いることが重要であり、これにより我々の訓練目的は必ずしも真の限界クラス分布を事前知識として用いる必要がないことも可能となる。さらに、クラスあたりの学習サンプルが多いからといって、モデルの偏りが大きくなるとは限らないため、LAやLDAMのようにクラスあたりのサンプル数に基づいてマージンを測定するのではなく、ダイナミックにマージンを調整することで、各クラスに対する偏りの度合いをより適切に評価することができる。サンプル数だけでは偏りの程度を判断することはできません。また、LA／LDAMやCausal Normなど、固定的なマージンや調整を用いる先行研究とは異なり、各クラスの偏りの程度は決して固定値であるべきではなく、動的な変化の過程にあると主張する。偏りの原因は、データだけに起因するものではなく、モデルとデータの相互作用に起因するものである。DAでは、真の限界クラス分布が得られるか、ソースデータとターゲットデータが同じ分布から収集されるかのどちらかのシナリオに限定されることが最大の問題であり、現実の世界では理想的すぎるということである。DebiasPLの有効性と実現可能性を示すために、いくつかのベンチマークで実験が行われた。不均衡なデータに対しては、Tab. 1では，LA[35]をFixMatchに統合した場合，FixMatch w／DebiasPL に比べて大きく遅れをとることがわかる．バランスデータでは，調整または再重み付けベクトルが真のクラス分布に基づいて計算されるため，真の限界クラス分布に依存する既存のロングテール法のほとんどは，大きな変更なしにもはや適用できない（バランスクラス分布は，すべてのクラスに対して同一の処理をもたらす）．

------------------------------

表1. これらの望ましい特性をすべて備えているのは、我々の手法だけである。LA [35], LDAM [10], DA [4], Causal Norm [54], そして我々のDebiasPLを含む、トレーニングデータ分布問題の解決に集中した先行研究との主要特性における比較である。望ましい（望ましくない）特性は緑色（赤色）である。

------------------------------

半教師付き学習では，適応脱バイアスモジュールと適応限界損失を採用することで，図7のように，提案するDebiasPLをFixMatchに統合することができる．SSLの性能をさらに向上させ、視覚言語事前学習モデルの能力を引き出すために、学習時に、CLIPをFixMatch／DebiasPLに組み込むことも可能であり、捨てられたラベルなしインスタンスをCLIPで疑似ラベル付けする。CLIPが自信を持てないインスタンスはノイズが多い可能性があるため、CLIPの信頼度スコアがτ$\_{clip}$より大きい未ラベルのインスタンスのみをCLIPで擬似ラベル付けする。反復ごとに再予測することなく、すべての訓練データに対するCLIPの予測を取得し、辞書に格納することができた。したがって、CLIPモデルの使用によってもたらされる計算オーバーヘッドは無視できる程度である。CIFAR10のような低解像度のデータセットでCLIPを使用すると、CLIPにスケールベースのデータ増強機能がないこともあり、わずかな利益しか得られないため、我々は大規模データセットにおいてのみCLIPを活用する [45]．トランスダクティブゼロショット学習では、視覚言語事前学習モデルから学んだ知識をより良く利用し、下流のZSLタスクに知識を転送する際のドメインシフト問題を緩和するために、FixMatchとCLIPに基づくトランスダクティブゼロショット学習（T-ZSL）を行う新しい枠組みを開発しました。具体的には、ワンホットラベル（すなわち、モデル出力のarg max）を活用し、最大のクラス確率が信頼閾値τ$\_{clip}$（デフォルトでは=0.95）以上になる擬似ラベルを保持することによって、再び擬似ラベルのアイデアを利用する。これらのインスタンスと擬似ラベルは、SSLでは「ラベル付きデータ」とみなされる。この後、元のFixMatchパイプラインに従って、「ラベル付き」データと「ラベルなし」データを合同で最適化することができます。先行研究との公平な比較とシステム全体の簡略化のため、モデルの初期化部分を含め、他のすべての学習レシピと設定は、オリジナルのFixMatch+EMANの設定と一致している。図は付録の通りです。CLIPは非常に偏っているため、バニラFixMatch + CLIPフレームワークは、オリジナルのCLIPゼロショット学習の性能を下回り、偏ったモデルから学習すると既存の偏りがさらに増幅され、より間違った予測を生み出す可能性があるという私たちの以前の仮説が確認された。そこで、教師なし損失L$\_{u}$を、クラス間混乱を緩和するためのAdaptive Marginal Lossと、偏った疑似ラベルを生成するためのAdaptive Debiasで、4.1節と同様に更新することにしました。

------------------------------

本節では、DebiasPLが、1）半教師付き学習とゼロショット学習の両方のベンチマークで最先端の結果を出す、2）普遍的なアドオンとして機能し、様々な手法に一貫した性能向上をもたらす、3）ドメインシフトに対してより強い堅牢性を示す、4）ロングテール、バランス、さらにはハイブリッドデータで性能を向上できる、を示すために実証実験を行う。

------------------------------

データセット CIFAR10[26]、long-tailed CIFAR10 (CIFAR10-LT) [26]、ImageNet1K [50]など、ラベル付きデータの量を変えた複数のSSLベンチマークでDebiasPLの総合評価を実施した。バランスのとれたベンチマークでは、2%以上のラベル付きデータを使用すると、性能はほぼ飽和する。我々は、CIFAR10の0.08%／0.16%／2%、ImageNet-1Kの1%／0.2%という極めて低ショットの設定に注目した。不均衡ベンチマークについては、[62]の設定に従い、DebiasPLをCIFAR10-LTで、あらかじめ定義された様々な不均衡比率γ（γ∈［100、200］）およびラベル付きデータの割合（10%、30%）でテストしています。データセットの詳細については，付録を参照されたい．セットアップ ロングテールCIFAR10とCIFAR10データセットの両方に関するすべての実験について、我々は、ネットワークアーキテクチャWRN-28-2 [19, 66]を使用するために以前の作品 [53, 62]に従っています。また、最適化の総反復回数を半分に減らす以外は、FixMatchの同じハイパーパラメータのセットに従っている。

------------------------------

CIFAR10-LTです： # ラベルの数(パーセント)

------------------------------

CIFAR10: ラベルの数(パーセント)

------------------------------

表2. ラベルなし／ラベルありデータの限界クラス分布に関する予備知識がなくても、CIFARおよびCIFAR-LT SSLベンチマークの両方におけるDebiasPLの性能は、バランスデータ用に設計されるか、ロングテールデータ用に綿密に調整された従来のSOTAを凌駕しています。DibasMatchは、すべてのベンチマークで同じハイパーパラメータのセットで実験されています。§ は、[31]、[53]、[62]からコピーした、対応する手法の最もよく報告された結果を示している。γ: 不均衡率。5つの異なるフォールドで平均した結果を報告する。

------------------------------

表3. DebiasPLは、ImageNet-1Kの半教師付き学習において、様々な割合のラベリングサンプルを用いて、特に極端に少ないショット数の設定で、最先端の結果を出しています。すべての結果は、ResNet-50をバックボーンとして生成されています。ただし，PAWS [2]はノンパラメトリックに生成された擬似ラベルで300エポックの事前学習を行っている．

------------------------------

バックボーンネットワークは、FixMatch w／ EMAN [9]で紹介されたトレーニングレシピに従いますが、これはImageNet-1Kのすべての実験のデフォルトベースラインでもあります。モデルは[9]のようにMoCo v2 + EMANで初期化される。複数のビューを持つ設定の場合、各ラベルなしサンプルに対して、2つの強い増強と2つの弱い増強を実行する。各強オーグメントインスタンスは1つの弱オーグメントインスタンスとペアになり、図7の元の設定と同様に、擬似ラベリングによって2つのペアを共同で最適化する。マルチビューは収束速度を上げ、学習プロセスを安定させることができる。

DebiasPLはシンプルでありながら効果的である。表中のTab. 2およびTab. 3は，DebiasPLがすべての実験済みベンチマークで最先端の性能を発揮し，現在のアプローチを大差で上回っていることを示している．CLIPを用いない場合，DebiasPLはCIFARにおいてCoMatchを上回り，ImageNet-1KにおいてCoMatchと同程度の性能を発揮する．DebiasPLは、シンプルであるというメリットで勝利しています。CLIPの力を活用すれば，DebiasPLの性能を大幅に向上させることができ，ImageNet-1K SSLではCoMatchを4%ほど上回ることができる．DebiasPLはソース／ターゲットのデータ分布に依存しない．表2より 表2より，CIFARとロングテールのCIFAR SSLベンチマークの両方において，統一されたフレームワークと同じハイパーパラメータを用いることで，DebiasPLは，バランスデータ用に設計されたりロングテールデータ用に綿密に調整されたりした従来の最先端手法を上回ることができることがわかる．さらに，Tab. 4では，ラベル付きデータとラベルなしデータが異なる分布に従うシナリオでテストした場合，DebiasPLはベースラインに対してさらに大きな利得（11.4%）を得ることができることを示しています．ラベル付きデータが少ないほど、より大きな利益が得られることが、Tab. 2およびTab. 3では、完全教師あり学習と半教師あり学習の間のギャップがほとんどなくなっています。また、DebiasPLは、Tab.5に示されるように、普遍的なアドオンである。5. DebiasPLを様々なSSL手法に組み込むことで、一貫した性能向上を達成することができる。

------------------------------

表4. DebiasPLは、ラベルなしデータがラベル付きデータと同じ、すなわちロングテール分布の場合、またはラベル付きデータと異なる、すなわちセマンティクス間でバランスよく分布している場合に、SSLの性能を一貫して向上させる。5フォールドで平均化した結果を報告する。

------------------------------

表5. DebiasPLは普遍的なアドオンである。CIFAR10における様々なSSL手法のトップ1精度を、5フォールドで平均して比較した。1クラスあたり4インスタンスがラベル付けされている。

------------------------------

表6. DebiasPLはImageNet-1Kでゼロショット学習の最先端の結果を出し、より大きなモデルやラベルで微調整されたCLIPを凌駕している。†: CoOpとCLIP(few-shot)は、約1.5%のアノテーションデータでファインチューニングされている。

------------------------------

データセット ImageNet-1K [50]を用いて、T-ZSLにおけるDebiasPLの効率性を評価する。また，EuroSAT [20], MNIST [28], CIFAR10 [26], CIFAR100 [26], Food101 [7] を評価データセットとして使用し，ドメインシフトに対する頑健性を示す．セットアップ T-ZSLは、ターゲットデータについて、クラス候補の可能性があるリストが既知であると仮定する。この設定に従い、対象データには意味ラベルを用いない。FixMatchのDebiasPLと同様にCLIPのDebiasPLを適用するが，ラベル付きデータは人間の注釈者ではなくCLIPによって「ラベル付け」されることを除いては，FixMatchのDebiasPLと同様である．具体的には，CLIPの信頼度スコアがτ$\_{clip}$より大きいすべてのラベルなしインスタンスは，CLIPによって疑似ラベル付けされ，「ラベル付き」データとみなされる．バックボーンはResNet50、閾値τ$\_{clip}$は0.95を全てのデータセットで使用した。FixMatch + EMANと同じデフォルトのハイパーパラメータと学習レシピが、公平な比較のために利用される。詳細は付録をご覧ください。DebiasPLはゼロショット学習でSOTAを達成し，人間による部分的なラベル付けデータで微調整されたCLIP [45]やCoOP [68]をも凌ぐ結果を得た．さらに，ResNet50をバックボーンとするDebiasPLは，Tab.6に示すように，15倍以上のバックボーンを持つCLIPを大きく上回ることができる．6. また，DebiasPL w／ CLIP（人間の注釈を一切使用しない）を100エポックでゼロショット学習させた場合の時間コストは，CLIPの全体学習時間の0.01%未満である．図9に示すように，DebiasPLは，意味ラベルにアクセスしないゼロショットCLIPよりも，ドメインシフトに対して強い頑健性を示す．また，DebiasPLは，より大きなドメインシフトを持つデータセットにおいて，より大きな利益（20%以上）を得ることができる．例えば，衛星画像データセットEuroSAT [20]では，25.7%という驚異的な利益を得ることができる．

------------------------------

図9. 様々なデータセットでゼロショット学習を行った場合、DebiasPLはドメインシフトに対してより強い頑健性を示す。ResNet-50をバックボーンネットワークとして実験している。CLIPの結果は公式コードで再現しています。

------------------------------

本論文では、これまで知られていなかった偏った擬似ラベルの問題についての研究を行う。本論文では、真のデータ分布に関する事前知識を利用することなく、偏った擬似ラベルが生徒モデルに与える影響を動的に緩和する、シンプルかつ効果的な手法DebiasPLを提案する。DebiasPLは、半教師付き学習、トランスダクティブ・ゼロショット学習の両タスクにおいて、従来の最先端技術を大幅に上回る性能を発揮し、ドメインシフトに対してより強い頑健性を示す。謝辞 この研究は、Etegent TechnologiesのLowShot DetectionとSemi-supervised Detectionに関する米国政府の資金により、一部支援されたものである。

------------------------------

[1] Andrew Arnold, Ramesh Nallapati, and William W Cohen. (注1) 転送学習のための手法の比較検討。In Seventh IEEE international conference on data mining workshops (ICDMW 2007), pages 77-82. IEEE, 2007. 1 [2] Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Armand Joulin, Nicolas Ballas, and Michael Rabbat. (2)は、(3)は、(4)は、(5)は、(6)を表している。2, 7 [3] Elias Bareinboim and Judea Pearl. 因果推論における選択バイアスを制御する。人工知能と統計学、100-108ページにて。PMLR, 2012. 4 [4] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch： 分布マッチングとオーグメンテーションアンカーリングによる半教師付き学習。In International Conference on Learning Representations, 2019. 1, 2, 5, 6, 7 [5] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch： 半教師付き学習への全体的なアプローチ. Advances in Neural Information Processing Systems, 32:5049-5059, 2019. 1, 2, 7 [6] Michel Besserve, Arash Mehrjou, Rémy Sun, and Bernhard Schölkopf. Counterfactuals uncover the modular structure of deep generative models. In International Conference on Learning Representations, 2019. 4 [7] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101 - mining discriminative components with random forests. In European Conference on Computer Vision, 2014. 4, 8 [8] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners.arXiv preprint arXiv:2005.14165, 2020. 2 [9] Zhaowei Cai, Avinash Ravichandran, Subhransu Maji, Charless Fowlkes, Zhuowen Tu, and Stefano Soatto. 自己教師付き学習と半教師付き学習のための指数移動平均正規化. IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 194-203, 2021 に掲載されている。7 [10] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. ラベル分布を考慮したマージンロスで不均衡なデータセットを学習する。In Advances in Neural Information Processing Systems, pages 1567-1578, 2019. 1, 2, 6 [11] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. クラスタ割り当ての対比による視覚的特徴の教師なし学習. In Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS), 2020. 7 [12] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 視覚表現の対比学習のためのシンプルなフレームワーク。In International conference on machine learning, pages 1597-1607. PMLR, 2020. 7[13] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. (1)自己教師付きモデル(Big self-supervised model)は強い半教師付き学習者である.arXiv preprint arXiv:2006.10029, 2020. 2, 7 [14] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment： 検索空間を縮小した実用的な自動データ補強。In Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 702-703, 2020. 3 [15] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. オブジェクトをその属性で記述する。In 2009 IEEE conference on computer vision and pattern recognition, pages 1778-1785. IEEE, 2009. 2 [16] Andrea Frome, Greg Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. Devise： 深層視覚意味埋め込みモデル(A deep visual-semantic embedding model). 2013. 2 [17] Sander Greenland, James M Robins, and Judea Pearl. 因果関係推論における交絡と折りたたみ可能性。統計科学、ページ 29-46, 1999. 4, 5 [18] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis： 大語彙のインスタンス分割のためのデータセット。In Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 5356-5364, 2019. 1 [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 画像認識のための深層残差学習(Deep residual learning for image recognition). In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016. 7 [20] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat： A novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7):2217-2226, 2019. 4, 8 [21] ポール・W・ホランド(Paul W Holland). 統計学と因果関係推論。アメリカ統計学会誌, 81(396):945-960, 1986. 5 [22] Sham M Kakade, Karthik Sridharan, and Ambuj Tewari. 線形予測の複雑さについて： リスクバウンズ、マージンバウンズ、正則化。Advances in neural information processing systems, 21, 2008. 2 [23] Michael Kampffmeyer, Yinbo Chen, Xiaodan Liang, Hao Wang, Yujia Zhang, and Eric P Xing. ゼロショット学習のための知識グラフ伝搬を再考する。In Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 11487-11496, 2019. 2, 8 [24] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. ラベル分布を考慮したマージンロスで不均衡なデータセットを学習する。In International Conference on Learning Representations, pages 1567-1578, 2020. 1, 2[25] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. 教師なしドメイン適応のための対照的適応ネットワーク．In Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 4893- 4902, 2019. 1 [26] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 4, 6, 8 [27] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. ゼロショット視覚オブジェクトのための属性に基づく分類

------------------------------

のカテゴライズを行う。IEEE transactions on pattern analysis and machine intelligence, 36(3):453-465, 2013. 2 Yann LeCun. 手書き数字データベースmnist. http:／yann. lecun. com／exdb／mnist／, 1998. 4, 8 Dong-Hyun Lee et al. Pseudo-label： ディープニューラルネットワークのためのシンプルで効率的な半教師付き学習法. In Workshop on challenges in representation learning, ICML, volume 3, 2013. 1, 2, 3 Ang Li, Allan Jabri, Armand Joulin, and Laurens van der Maaten. ウェブデータから視覚的なn-gramを学習する。In Proceedings of the IEEE International Conference on Computer Vision, pages 4183-4192, 2017. 8 Junnan Li, Caiming Xiong, and Steven Hoi. Comatch： 対照的なグラフ正則化を用いた半教師付き学習。 arXiv preprint arXiv:2011.11183, 2021. 2, 7 Bin Liu, Zhirong Wu, Han Hu, and Stephen Lin. (注釈付きデータが限られている場合のラベル伝播のための深いメトリックトランスファー). In Proceedings of the IEEE／CVF International Conference on Computer Vision Workshops, pages 0-0, 2019. 1 Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. 畳み込みニューラルネットワークのための大マージンソフトマックスロス。In ICML, volume 2, page 7, 2016. 2 Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 機械学習におけるバイアスとフェアネスに関するサーベイ。ACM Computing Surveys (CSUR), 54(6):1- 35, 2021. 1 Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. ロジット調整によるロングテール学習。In International Conference on Learning Representations, 2021. 2, 6, 7 宮戸猛、前田真一、小山正則、石井伸. 仮想敵対的訓練：教師あり・半教師あり学習のための正則化法.IEEE transactions of pattern analysis and machine analysis and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence, 41(8):1979- 1993, 2018. 2 ナ・ジェミン、チョン・ヒチョル、チャン・ヒョンジン、ファン・ウォンジュン. Fixbi： 教師なしドメイン適応のためのドメイン空間の架け橋.Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 1094-1103, 2021. 1 Nihal V Nayak and Stephen H Bach. (注1)本論文は、(注2)本論文の一部であり、(注3)本論文は、(注4)本論文の一部である。2, 8 Mohammad Norouzi, Tomas Mikolov, Samy Bengio, Yoram Singer, Jonathon Shlens, Andrea Frome, Greg Corrado, and Jeffrey Dean. 意味埋込の凸結合によるゼロショット学習。In ICLR, 2014. 8 Judea Pearl. 統計学における因果推論： An overview. 統計学サーベイランス, 3:96-146, 2009. 4, 5 Judea Pearl. 因果関係。ケンブリッジ大学出版局, 2009. 5 Judea Pearl. 直接効果と間接効果.arXiv preprint arXiv:1301.2300, 2013. 4, 5 ジュデア・パール、ダナ・マッケンジー. The book of why: the new science of cause and effect. ベーシックブックス, 2018. 5 ジェフリー・ペニントン、リチャード・ソッチャー、クリストファー・D・マニング. Glove： Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532-1543, 2014. 2 Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision.arXiv preprint arXiv:2103.00020,2021. 1, 2, 3, 6, 7, 8 Lorenzo Richiardi, Rino Bellocco, and Daniela Zugna. 疫学における媒介分析：方法、解釈、バイアス。International journal of epidemiology, 42(5):1511-1519, 2013. 5 Bernardino Romera-Paredes、Philip Torr. ゼロショット学習への恥ずかしいほどシンプルなアプローチ。In International conference on machine learning, pages 2152-2161. PMLR, 2015. 2 ドナルド・B・ルービン. 潜在的な結果を用いた因果推論： デザイン、モデリング、決定。Journal of the American Statistical Association, 100(469):322-331, 2005. 4 ドナルド・B・ルービン. 因果推論の本質的な概念：驚くべき歴史と興味深い未来。Biostatistics & Epidemiology, 3(1):140-155, 2019. 4 Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211- 252, 2015. 6, 8 Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Deep semi-supervised learning のための確率的変換と摂動による正則化. arXiv preprint arXiv:1606.04586, 2016. 2 Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. クロスモーダル転送によるゼロショット学習。In Advances in Neural Information Processing Systems, pages 935-943, 2013. 2 Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch： 一貫性と信頼性を備えた半教師付き学習の簡略化. Advances in Neural Information Processing Systems, 33, 2020. 1, 2, 3, 7, 8 Kaihua Tang, Jianqiang Huang, and Hanwang Zhang. 良いものを残し、悪いものを取り除く勢いのある因果関係によるロングテール分類. 神経情報処理システムの進歩, 33, 2020. 2, 5, 6 アンティ・タルヴァイネン、ハッリ・ヴァルポラ。平均的な教師はより良いロールモデルである： Weight-averaged consistency targets improve semi-supervised deep learning results. arXiv preprint arXiv:1703.01780, 2017. 2 Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8769-8778, 2018. 1 Feng Wang, Weiyang Liu, Haijun Liu, and Jian Cheng. Additive margin softmax for face verification. arXiv preprint arXiv:1801.05599, 2018. 2

------------------------------

[58] Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. ゼロショット学習のサーベイ： 設定、方法、およびアプリケーション。ACM Transactions on Intelligent Systems and Technology (TIST), 10(2):1-37, 2019. 2

------------------------------

ステラ・ユー 多様な分布を意識した専門家のルーティングによるロングテール認識. In International Conference on Learning Representations, 2021. 1, 2

------------------------------

クロスレベルインスタンスグループ識別による特徴学習。Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 12586-12595, 2021. 5

------------------------------

ファン・ヤン Crest： アンバランスな半教師付き学習のためのクラス・リバランス自己学習フレームワーク。In Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 10857-10866, 2021. 2, 7

------------------------------

とQuoc V Le. Unsupervised data augmentation for consistency training. arXiv preprint arXiv:1904.12848, 2019. 7

------------------------------

クオック・レ 一貫性トレーニングのための教師なしデータ増強. 神経情報処理システムの進歩, 33, 2020. 2

------------------------------

ノイズの多い学生による自己学習でイメージネット分類が向上する。Proceedings of the IEEE／CVF Conference on Computer Vision and Pattern Recognition, pages 10687-10698, 2020. 1, 2

------------------------------

とQianru Sun. 弱教師付き意味分割のための因果的介入. 神経情報処理システムの進歩, 33, 2020. 4

------------------------------

1.Introduction

Real-world observations, as well as non-curated datasets, are naturally long-tail distributed [18, 56]. Imbalanced classification [10, 24, 59] tackles such data biases to prevent models from being dominated by head-class instances. Developing visual recognition systems capable of counteracting biases also has significant social impacts [34].

While existing methods focus on debiasing from imbalanced ground-truth labels collected by human annotators, we discover that pseudo-labels produced by machine learning models are naturally imbalanced, creating another source for widespread biased learning!

Pseudo-labels are highly confident predictions made by an existing (teacher) model on unlabeled data, which then become part of the training data for supervising the (student) model adaptation to unlabeled data (Fig. 1a). When the student model is the teacher model itself, the learning process is also known as self-training [4,5,29,53,65]. Pseudo-labeling is widely used in semi-supervised learning (SSL) [32, 53], domain adaptation [25, 37], and transfer learning [1].

We examine pseudo-label distributions in two common tasks. 1) In zero-shot transfer learning (ZSL) where the source and target domains are different, a pretrained CLIP model [45] produces highly imbalanced predictions on the curated and balanced ImageNet-1K dataset, although the training set of CLIP is approximately balanced (Fig. 1c). More than 3500 instances are predicted as class 0, 3 times the actual number of samples in class 0. 2) In semi-supervised learning where the source and target domains are the same, FixMatch [53] trained on labeled CIFAR10 images generates highly biased pseudo-labels on unlabeled images, although both the labeled and unlabeled sets are balanced (Fig. 1b).

That is, pseudo-labels created by machines are naturally imbalanced, just like ground-truth labels created by humans. If we address this previously unknown imbalanced classification problem arising from pseudo-labels instead of groundtruth training labels, we could improve model learning based on pseudo-labels and remove the model bias towards false majorities created by pseudo-labels.

jorities created by pseudo-labels.

We propose a novel and effective debiased learning method with pseudo-labels, without any knowledge about the distribution of actual classification margins that are readily available to debiased learning with ground-truth labels [22, 33, 57]. It consists of an adaptive debiasing module and an adaptive marginal loss. The former dynamically removes the classifier response bias through counterfactual reasoning, whereas the latter dynamically adjusts the margin of each class according to the imbalance of pseudo-labels.

実世界の観測は、キュレーションされていないデータセットと同様に、自然にロングテール分布している [18, 56]。不均衡分類[10, 24, 59]は、このようなデータの偏りに対処し、モデルがヘッドクラスのインスタンスに支配されることを防ぐ。また、バイアスに対抗できる視覚認識システムの開発は、社会的に大きな影響を与える[34]。

既存の方法は、人間の注釈者が収集した不均衡なグランドトゥルースラベルからのデビアスに焦点を当てているが、我々は、機械学習モデルが生成する疑似ラベルが自然に不均衡であることを発見し、広く偏った学習のための別のソースを作成します！

擬似ラベルとは、既存の（教師）モデルがラベルのないデータに対して行った確信度の高い予測のことで、ラベルのないデータへの（生徒）モデルの適応を監督するための学習データの一部となる（図1a）。生徒モデルが教師モデルそのものである場合、この学習プロセスは自己学習とも呼ばれる[4,5,29,53,65]。疑似ラベルは半教師付き学習（SSL）[32,53]、ドメイン適応[25,37]、転移学習[1]で広く用いられている。

我々は、2つの一般的なタスクにおける擬似ラベルの分布について検討する。1）ソースとターゲットのドメインが異なるゼロショット転移学習（ZSL）において、事前に学習したCLIPモデル[45]は、CLIPの学習セットはほぼバランスしているが、キュレーションされバランスしたImageNet-1Kデータセット上で非常にバランスの悪い予測を行う（図1c）。2）ラベル付きCIFAR10画像で学習したFixMatch[53]は、ラベル付きとラベルなしがバランスしているにもかかわらず、ラベルなし画像で非常に偏った擬似ラベルを生成する（図1b）。

つまり、機械が作成した擬似ラベルは、人間が作成した真実のラベルと同じように、自然に不均衡になるのです。このように、真実のラベルではなく、擬似ラベルに起因する未知の不均衡な分類問題に対処すれば、擬似ラベルに基づくモデル学習を改善し、擬似ラベルによって生じる偽の多数派へのモデルの偏りを除去することができる。

擬似ラベルが作り出す偽の多数派に対するモデルの偏りを取り除くことができる。

本論文では、擬似ラベルを用いた新規かつ効果的なデビアス学習法を提案する。擬似ラベルを用いたデビアス学習では、実際の分類マージンの分布に関する知識が必要であるが、これは、グラウンドトゥルースラベルを用いたデビアス学習 [22, 33, 57] で容易に利用可能である。適応的デビアスモジュールと適応的マージンロスから構成される。前者は反実仮想推論により分類器の応答バイアスを動的に除去し、後者は擬似ラベルの不均衡に応じて各クラスのマージンを動的に調整する。

Validated by our extensive experiments, our simple debiased learning not only improves the state-of-the-art on ImageNet-1K by 26% for SSL with 0.2% annotations and 9% for ZSL, but is also a universal add-on to various pseudo-labeling methods with more robustness to domain shift. The imbalanced pseudo-labeling issue is even more severe when the unlabeled raw data is naturally imbalanced, and the model tends to mislabel tail-class samples as head-class. By applying debiased learning, we improve SSL performance under long-tailed settings by a large margin.

Our work makes four major contributions. 1) We systematically investigate and discover that pseudo-labels are naturally imbalanced and create biased learning. 2) We propose a simple debiased learning method with pseudo-labeled instances, requiring no knowledge of their actual classification margins. 3) We improve the ZSL/SSL state-of-the-art by a large margin and demonstrate that our debiasing is a universal add-on to various pseudo-labeling models. 4) We establish a new effective ZSL/SSL pipeline for applying vision-and-language pre-trained models such as CLIP.

我々の広範な実験によって検証された結果、我々の単純な 脱バイアス学習は、ImageNet-1Kにおける最新技術を、0.2%のアノテーションを持つSSLで26%、ZSLで9%改善するだけでなく、ドメインシフトに対してより頑健で、様々な疑似ラベリング方法に普遍的に付加することが可能であることがわかった。ラベルのない生データが自然に不均衡になる場合、不均衡な擬似ラベルの問題はさらに深刻であり、モデルはテールクラスのサンプルをヘッドクラスと誤ラベルする傾向がある。脱バイアス学習を適用することで、ロングテール設定下でのSSL性能を大きく向上させることができる。

我々の研究は、4つの主要な貢献をしている。1) 擬似ラベルが自然に不均衡になり、偏った学習を引き起こすことを系統的に調査・発見する。2) 疑似ラベルを用いた脱バイアス学習法を提案し、実際の分類マージンを知る必要がない。3) ZSL/SSLを大幅に改善し、デビアス学習が様々な擬似ラベリングモデルに対して普遍的なアドオンであることを証明する。4) CLIPのような視覚と言語の事前学習済みモデルを適用するための、新しい効果的なZSL/SSLパイプラインを確立する。

3. Pseudo-Labels are Naturally Imbalanced

In contrast to previous work that concentrated on biases caused by trained on imbalanced data, our focus is on pseudo-label biases, even when trained on balanced data. In this section, we provide an analysis of this previously unknown issue hidden behind the tremendous success of FixMatch [53] on SSL and CLIP [45] on ZSL, both of which require the use of “pseudo-labeling” to transfer knowledge learned in source data to target data.

We first describe the backgrounds for pseudo-labeling approaches and then analyze their bias issue. We attribute the cause of bias to the inter-class correlation problem.

3. 疑似ラベルは自然にアンバランスになる

不均衡なデータで訓練した場合に生じるバイアスに集中していた以前の研究とは対照的に、我々は、バランスのとれたデータで訓練した場合であっても、擬似ラベルのバイアスに焦点をあてている。本節では、SSLにおけるFixMatch [53]とZSLにおけるCLIP [45]の大成功の裏に隠された、この未知の問題の分析を行う。

我々はまず、擬似ラベリングアプローチの背景を説明し、次にそのバイアス問題を分析する。バイアスの原因はクラス間相関の問題であるとする。