

# マルチラベル分類のための重みつき非対称損失関数

安田有希 宮崎太郎 後藤淳

日本放送協会 放送技術研究所

{yasuda.y-hk,miyazaki.t-jw,goto-j.fw}@nhk.or.jp

## 概要

単一のテキストに複数のラベルを付与するマルチラベルテキスト分類は、自然言語処理で重要なタスクの一つであるが、大きな課題としてラベルの出現頻度が不均衡であることが挙げられる。学習データ中にあまり出現しない低頻度ラベルは、ポジティブサンプル数が極端に少なく、ほとんどの入力にネガティブサンプルである。そのため、モデルは低頻度ラベルをほとんどネガティブサンプル由来の損失から学習することになる。本研究では、ラベルの共起重みと出現頻度重み、ネガティブサンプル由来の損失値を抑制する重みを組み合わせる手法を提案する。実験の結果、Reuters-21578 では、micro-F1 で 0.911 と SOTA を上回る性能を達成した。

## 1 はじめに

マルチラベルテキスト分類 (MLTC) は、自然言語処理の中でも重要な課題の一つであり、今日実世界の様々な場面で適用されている。たとえば、法律文書の分類 [1]、カルテによる自動診断 [2] が挙げられる。MLTC では文書  $x_i \in X$  に対して、事前に定義されたラベルから適切なラベルのサブセット  $y_i \in \{0, 1\}^C$  を付与するタスクである [3, 4]。なお、ここで  $X$  は文書の集合であり、 $C$  は事前に定義されたラベルの総数である。MLTC においてベンチマークとして広く利用されている Reuters-21578 [5] のデータのサンプルを図 1 に示す。

Sample 1	BAHIA COCOA REVIEW Showers continued throughout the week in ...
Labels 1	cocoa
Sample 2	NATIONAL AVERAGE PRICES FOR FARMER-OWNED RESERVE The U.S. Agriculture Department reported the farmer-owned reserve ...
Labels 2	sorghum, oat, barley, corn, wheat, grain

図 1 Reuters-21578 のデータ例

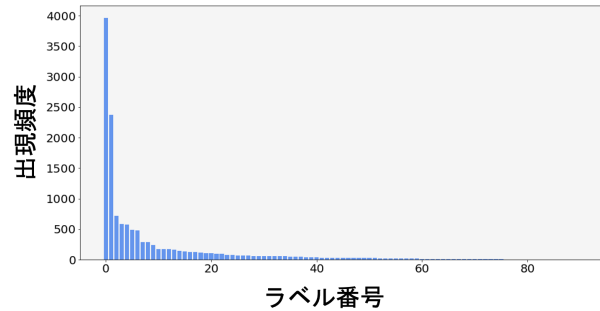


図 2 Reuters-21578 におけるラベル分布

MLTC の特徴として、ラベルの頻度がロングテール分布に従う不均衡データセットが挙げられる。一例として、図 2 に Reuters-21578 の各ラベルの頻度分布を示す。図 2 が示すとおり、低頻度ラベルはポジティブサンプルから学習することは稀であり、殆どの場合ネガティブのサンプルから学習する。これは、*Imbalanced learning* と呼ばれ、機械学習の分類問題における大きな課題の一つである [6, 7]。

この課題を解決するために、画像分類分野では損失値に重み付けを行うような様々な損失関数手法が提案されている [8, 9, 10]。さらに、これらの手法は自然言語処理分野にも流用され、効果を発揮している [11]。本研究では、ラベルの共起情報をもとにした重みとラベルの頻度から構成された重み、ネガティブサンプルの影響を軽減する重みを組み合わせ、不均衡データセットの影響を抑制するような損失関数を提案する。

## 2 関連研究

Class-Balanced Loss (CBL) [8] は、データセット内の各ラベルの出現数に反比例する重みを構築して損失値に重み付けする手法である。つまり、出現数の多いラベルの損失値ほど抑制されるような重みを構築するため、各ラベル間での損失値のバランスが調整される。しかしながら、CBL では同一のラベルにおけるポジティブサンプル由来の損失とネガティブサンプル由来の損失のバランスは考慮されない。マ

ルチラベル分類では、ラベル空間が大きいタスクを取り扱うことも多いため、学習サンプルにおいてモデルが計算する損失値のほとんどが、ネガティブラベル由来の損失値である。つまり、ラベル間の重み調整だけでは偏ったネガティブサンプルの影響を取り除けず、ポジティブサンプルに対しても意味のある学習が難しくなってしまう。

この課題を解決するために Distribution-Balanced Loss (DBL)[9] が提案されている。DBL は、損失値を調整する重みと Negative-Tolerant Regularization (NTR) 手法を統合して構成される。重みはバッチ内のラベル出現数の逆数とデータ全体のラベル出現数の逆数の比を求めることで算出される。また、NTR はモデルの出力からその逆数とハイパーパラメータによって算出される値を引くことによってネガティブサンプルの学習を抑制することを実現している。すなわち、モデルの出力が一定の値を下回ると急激に損失値を低下させることができる。ただし、DBL では学習のために事前定義しなければならないハイパーパラメータが非常に多く、データ量が多い傾向にある MLTC では探索のコストが非常に高い。

学習中のネガティブサンプルの影響を緩和するための別の手法として、非対称的な形を持った重みを損失値に適用する Asymmetric Loss (ASL)[10] が提案されている。ASL のコンセプトとして、以下の2点が挙げられる。それは Focal Loss (FOC)[12] における減衰率  $\gamma$  を正負のサンプルで分離する Asymmetric Focusing 手法と学習が十分に進んだネガティブサンプルの損失値を完全に破棄する Asymmetric Probability Shifting 手法である。つまり、ラベルの正負をもとにした損失値への重み付けと意味のあるサンプルからの学習の選択を同時に行う。ASL では、ネガティブサンプルに適用される  $\gamma_-$  をポジティブサンプルに適用される  $\gamma_+$  よりも高く設定することで、ネガティブサンプルに対するモデルの出力が小さい場合に損失値を抑制する一方で、ポジティブサンプルの損失値は不当に抑制されないような学習を実現している。ただし、ASL はネガティブサンプルの影響の抑制に高い効果を発揮する一方で、データセット内で多く出現する高頻度ラベルの影響を抑制できない。

### 3 提案手法

近年の多くの損失関数の提案は MLTC で多大な成果を収めているが、頻度もしくはラベルの正負とい

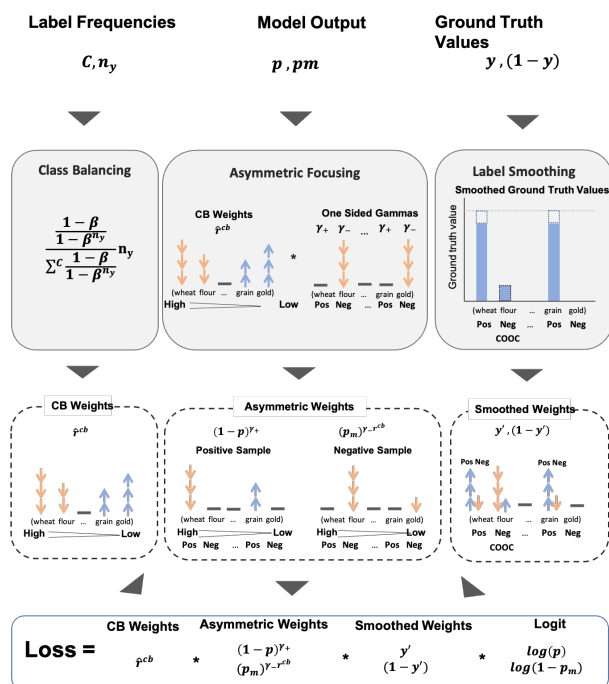


図3 本研究における提案手法の概要図

う一面的な観点でしか重みが構築されていない。そこで、本研究ではラベル間の頻度の差を緩和する重みとネガティブサンプルの影響を抑制する重み、低頻度ラベルのサンプル数の少なさを緩和する重みを統合した手法を提案する。データの中で多数派である低頻度でネガティブなサンプルに対しては  $\gamma$  の値を上げることで損失値を抑制する。一方で、少数派である高頻度ラベルでネガティブなサンプルに対しては  $\gamma$  の値を下げることで損失値の抑制を緩和する。さらに、トレーニングデータから抽出されたラベルの共起表現をもとにした Label Smoothing を行うことによって低頻度ラベルのサンプル数の少なさを補い、モデルの過学習を抑制することを狙う。本提案手法の概要図を図3に示す。

#### 3.1 Class Balancing モジュール

Class Balancing モジュールではトレーニングデータ全体のラベル分布から各ラベルに適用する重みを計算する。本提案手法では、CBL で提唱されているラベルの出現数をもとにした重み  $r^{cb}$  から重み  $\hat{r}^{cb}$  を構築する。 $\hat{r}^{cb}$  は、CBL と同様に損失値全体の重み付けを行うと同時に、後に説明する Asymmetric Focusing モジュールにおいて  $\gamma$  を調整する役割も持

つ。  $\hat{r}^{cb}$  は以下のように定義される。

$$r^{cb} = \frac{(1-\beta)}{1-\beta^{n_y}}, \quad (1)$$

$$\hat{r}^{cb} = \frac{r^{cb} n_y}{\sum^i r^{cb}}. \quad (2)$$

ここで、  $n_y$  は各ラベルの出現数、  $0 \leq \beta < 1$  は重みの強度を調整するハイパーパラメータである。

### 3.2 Asymmetric Focusing モジュール

Asymmetric Focusing モジュールでは ASL と同様に、ポジティブサンプルとネガティブサンプルのそれぞれに適用する重み  $(1-p)^{\gamma_+}$ ,  $(p_m)^{\gamma_-}$  を計算する。このうち、ネガティブラベルに適用される  $\gamma_-$  と Class Balancing モジュールで構築された重み  $\hat{r}_{cb}$  を掛け合わせる。提案手法 Asymmetric Loss with Weight Combination (ASLWC) における損失は以下のように定義される。

$$Loss = \hat{r}^{cb} [-yL_+ - (1-y)L_-], \quad (3)$$

$$L_+ = (1-p)^{\gamma_+} \log(p), \quad (4)$$

$$L_- = (p_m)^{\gamma_-} \log(1-p_m), \quad (5)$$

$$p_m = \max(p-m, 0). \quad (6)$$

ここで、  $0 \leq \gamma_+ < \gamma_-$  はハイパーパラメータである。また、式 6 は Asymmetric Probability Shifting 手法であり、モデルの出力確率が非常に低い場合、ネガティブサンプルによる影響を完全に破棄するための処理である。

### 3.3 Label Smoothing モジュール

Class Balancing と Asymmetric Focusing による重み調整では、低頻度ラベルに対応する入力サンプルの少なさは本質的に解決されない。本提案手法では Label Smoothing 手法 [13] をもとに、事前確率としてラベルの共起情報を設定した Label Smoothing を行う。本提案手法における Label Smoothing は以下のように定義される。

$$y' = (1-\alpha)y + \alpha u(y), \quad (7)$$

$$u(y) = \text{norm}\left(\sum_{i=1, y_i=1}^C PPMI_{ij}\right), \quad (8)$$

ここで、  $0 < \alpha < 1$  は Smoothing の度合いを決定するハイパーパラメータである。さらに、norm は正規化関数であり、Min-Max Normalization 手法を用いる。また、PPMI はトレーニングデータの各ラベルの共起回数から計算された正の自己相互情報量行列

を示す。通常の Label Smoothing では式 7 の項  $u(y)$  は一様分布  $\frac{1}{C}$  が定義されるが、本提案手法の Label Smoothing では、ポジティブラベルとデータセット内で共起するラベルを関連ラベルとみなし、関連ラベルにのみ Ground Truth Value を分配する。これによって、本来は正負の判定をするためだけの式 3 内の項  $y$ 、 $(1-y)$  を共起情報が加味された重みとして扱うことができる。

## 4 実験

提案手法の有効性を検証するために、ベースライン手法との比較実験を行った。ベースライン手法の損失関数として Binary Cross-Entropy (BCE)、FOC[12]、CBL[8]、CBL-NTR[11]、DBL[9]、ASL[10] を用いた。このうち、CBL-NTR と DBL は Reuters-21578 において State-of-the-art である。

### 4.1 データセット

本実験では、ベンチマークとして Reuters-21578[5] と AAPD[14]、20 Newsgroups[15] を用いた。それぞれのデータセットの統計情報を表 1 に示す。なお、Reuters-21578 に関して、本実験では modApté 分割 [16] を使用して、トレーニングデータとテストデータに分割した。また、Reuters-21578、20 Newsgroups はともに検証データが事前に定義されていないため、学習データを 9:1 の比率でランダムに分割し、検証データを構築した。

表 1 データセット概要

データセット	文書	ラベル	平均単語	平均ラベル
Reuters-21578	10,788	90	127.59	1.23
AAPD	55837	54	159.74	2.41
20 Newsgroups	18,846	33	273.16	2.74

Reuters-21578 は、1987 年にロイターのニューswire に掲載された経済分野のニュース記事から構成されており、ラベルの分布が非常に不均衡という特徴がある。AAPD は、科学論文のリポジトリである arXiv に公開された論文のアブストラクトから構成されており、実験に利用されたデータセットの中で中程度の不均衡さを持つ。20 Newsgroups は複数のドメインから構成されるニュース記事を集めたデータセットで、ラベル分布が均衡のとれたデータセットである。

表 2 実験結果

Methods/Metrics	Reuters-21578		AAPD		20 Newsgroups	
	macro-f1	micro-f1	macro-f1	micro-f1	macro-f1	micro-f1
	total / low	total / low	total / low	total / low	total / low	total / low
BCE	0.586 / 0.269	0.903 / 0.421	0.499 / 0.277	0.701 / 0.356	<b>0.885 / 0.821</b>	<b>0.899 / 0.822</b>
FOC	0.549 / 0.173	0.899 / 0.281	0.491 / 0.251	0.703 / 0.337	0.879 / 0.816	0.894 / 0.816
CBL	0.589 / 0.196	0.899 / 0.327	0.500 / 0.276	0.704 / 0.340	0.882 / 0.818	0.897 / 0.820
CBL-ntr	0.611 / 0.283	0.899 / 0.407	0.554 / 0.352	0.721 / 0.426	0.882 / 0.817	0.897 / 0.818
DBL	0.623 / 0.317	0.898 / 0.445	0.556 / 0.363	0.717 / 0.423	0.880 / 0.813	0.894 / 0.813
ASL	0.644 / 0.345	0.904 / 0.487	0.596 / 0.433	0.733 / 0.465	0.884 / 0.821	0.898 / 0.822
ASLWC	<b>0.669 / 0.383</b>	<b>0.911 / 0.547</b>	<b>0.601 / 0.449</b>	<b>0.736 / 0.483</b>	0.882 / 0.818	0.897 / 0.819

## 4.2 実験設定

本実験では、分類モデルとしてニューラルネットワークを用いた。RoBERTa[17]で入力サンプルを文章の特徴量に変換するエンコーダ部分と、文章の特徴量から各ラベルのスコアを出力する2層のFeed Forward Networks (FFNN) から構成されるモデルを使用した。モデルはPytorch[18]とTransformers[19]を用いて実装し、最適化手法としてRoBERTaにはAdamW[20]を適用し、FFNNにはRAdam[21]を適用した。各比較手法で共通のハイパーパラメータの学習エポックは50エポック、ドロップアウト率は0.1、RoBERTaの学習率は $2 \times 10^{-5}$ 、FFNNの学習率は $2 \times 10^{-3}$ 、出力閾値は0.5であった。また、各手法の固有のハイパーパラメータに、それぞれFOCは $\gamma = 2$ 、CBLは $\beta = 0.9$ 、DBLは $\beta = 10.0$ 、 $map - \gamma = 0.9$ 、ASLは $\gamma_+ = 1$ 、 $\gamma_- = 4$ 、ASLWCは $\alpha = 0.3$ 、 $\beta = 0.3$ 、 $\gamma_+ = 1$ 、 $\gamma_- = 4$ と設定した。本実験では、各手法の実験は異なるランダムシードで5回評価を行い、表2、表3には評価値の平均を記載した。

## 4.3 評価指標

本実験では、評価指標としてMacro-f1、Micro-F1を用いた。なお、出現頻度が下位三分の一であるラベルを低頻度ラベルとして定義し性能を比較した。

## 4.4 結果

本実験の結果を表2に示す。Reuters-21578とAAPDにおいて、すべてのラベルを用いた指標で本提案手法が比較手法を上回ったことが確認された。なお、Reuters-21578においてはSOTAであるCBL-NTRとDBLを超える精度が確認された。一方で、20 Newsgroupsにおいて提案手法を含む重み調整の手法がBCEの精度を超えなかったことが認め

られた。なお、すべてのデータセットにおける各ラベルグループごとの結果は付録Aに記載した。

## 4.5 議論

Reuters-21578とAAPDはどちらもラベルの頻度に偏りがあり、その影響をモデルに与えやすいデータセットである。一方で、20 Newsgroupsはラベルの不均衡の度合いが弱く、相対的には低頻度ラベルであっても1000件程度のサンプル数が確保されている。このことから、表2を鑑みると本提案手法はラベルの不均衡の度合いがある程度強いデータに有効である可能性がある。

表2の低頻度ラベルの結果から、Reuters-21578とAAPDにおいて本提案手法が低頻度ラベルの精度を向上させている。このことから、本提案手法において低頻度ラベルの精度向上が全体のラベルの分類精度の向上につながったのではないかと推察できる。つまり、本提案手法によってラベルの不均衡による影響を他の提案手法に比べて緩和できている可能性がある。

## 5 おわりに

マルチラベルテキスト分類では、ラベルの出現頻度に偏りがあるデータにおいて、低頻度ラベルの精度が低いという課題が存在する。そこで、本研究ではラベルの共起情報をもとにした重みと出現頻度の重み、ネガティブサンプルの損失を抑制する重みとを組み合わせた手法を提案した。実験結果より、提案手法がラベル分布の偏りが強いデータセットにおいてベースラインを上回ることが確認された。また、提案手法が低頻度ラベルの精度を改善していることからImbalanced learningの問題を緩和できる可能性がある。今後の展望として、より幅広いベンチマークデータセットでの検証と既存の大規模モデルへの適用による手法の頑健性の検証が挙げられる。

## 参考文献

- [1] Ilias Chalkidis, Emmanouil Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. Extreme multi-label legal text classification: A case study in EU legislation. In **Proceedings of the Natural Legal Language Processing Workshop 2019**, pp. 78–87, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [2] Clare Amanda and King D. Ross. Knowledge discovery in multi-label phenotype data. In **Principles of Data Mining and Knowledge Discovery**, pp. 42–53, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- [3] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. **IEEE Transactions on Knowledge and Data Engineering**, Vol. 26, No. 8, pp. 1819–1837, 2014.
- [4] Tsoumakas Grigorios and Katakis Ioannis. Multi-label classification: An overview. **International Journal of Data Warehousing and Mining**, Vol. 3, No. 3, pp. 1–13, 2007.
- [5] Philip J. Hayes and Steven P. Weinstein. CONSTRUE/TIS: a system for content-based indexing of a database of news stories. In **Proceedings of the The Second Conference on Innovative Applications of Artificial Intelligence**, IAAI '90, pp. 49–64. AAAI Press, 1990.
- [6] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. **IEEE Transactions on Knowledge and Data Engineering**, Vol. 21, No. 9, pp. 1263–1284, 2009.
- [7] Wang Yaqing, Yao Quanming, Kwok T. James, and Ni M. Lionel. Generalizing from a few examples: A survey on few-shot learning. **ACM computing Surveys**, Vol. 53, No. 3, pp. 1–34, 2020.
- [8] Cui Yin, Jia Menglin, Lin Tsung-Yi, Song Yang, and Belongie Serge. Class-balanced loss based on effective number of samples. In **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**, pp. 9268–9277, June 2019.
- [9] Wu Tong, Huang Qingqiu, Liu Ziwei, Wang Yu, and Lin Dahua. Distribution-balanced loss for multi-label classification in long-tailed datasets. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, **Proceedings of the European Conference on Computer Vision**, pp. 162–178, Cham, 2020. Springer International Publishing.
- [10] Ridnik Tal, Ben-Baruch Emanuel, Zamir Nadav, Noy Asaf, Friedman Itamar, Protter Matan, and Zelnik-Manor Lihi. Asymmetric loss for multi-label classification. In **Proceedings of the IEEE/CVF International Conference on Computer Vision**, pp. 82–91, October 2021.
- [11] Yi Huang, Buse Giledereli, Abdullatif Köksal, Arzucan Özgür, and Elif Ozkirimli. Balancing methods for multi-label text classification with long-tailed class distribution. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 8153–8161, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [12] Lin Tsung-Yi, Goyal Priya, Girshick Ross, He Kaiming, and Dollár Piotr. Focal loss for dense object detection. In **Proceedings of the IEEE International Conference on Computer Vision**, pp. 2999–3007, 2017.
- [13] Szegedy Christian, Vanhoucke Vincent, Ioffe Sergey, Shlens Jon, and Wojna Zbigniew. Rethinking the inception architecture for computer vision. In **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**, pp. 2818–2826, 2016.
- [14] Yang Pengcheng, Sun Xu, Li Wei, Ma Shuming, Wu Wei, and Wang Houfeng. SGM: Sequence generation model for multi-label classification. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 3915–3926, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.
- [15] Ken Lang. Newsweeder: Learning to filter netnews. In **Proceedings of the Twelfth International Conference on Machine Learning**, pp. 331–339, 1995.
- [16] Chidanand Apté, Fred Damerau, and Sholom M. Weiss. Automated learning of decision rules for text categorization. **ACM Trans. Inf. Syst.**, Vol. 12, No. 3, pp. 233–251, July 1994.
- [17] Zhuang Liu, Wayne Lin, Ya Shi, and Jun Zhao. A robustly optimized BERT pre-training approach with post-training. In **Proceedings of the 20th Chinese National Conference on Computational Linguistics**, pp. 1218–1227, Huhhot, China, August 2021. Chinese Information Processing Society of China.
- [18] Paszke Adam, Gross Sam, Massa Francisco, Lerer Adam, Bradbury James, Chanan Gregory, Killeen Trevor, Lin Zeming, Gimelshein Natalia, Antiga Luca, Desmaison Alban, Kopf Andreas, Yang Edward, DeVito Zachary, Raison Martin, Tejani Alykhan, Chilamkurthy Sasank, Steiner Benoit, Fang Lu, Bai Junjie, and Chintala Soumith. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 32. Curran Associates, Inc., 2019.
- [19] Wolf Thomas, Debut Lysandre, Sanh Victor, Chaumond Julien, Delangue Clement, Moi Anthony, Cistac Pierric, Rault Tim, Louf Remi, Funtowicz Morgan, Davison Joe, Shleifer Sam, von Platen Patrick, Ma Clara, Jernite Yacine, Plu Julien, Xu Canwen, Le Scao Teven, Gugger Sylvain, Drame Mariama, Lhoest Quentin, and Rush Alexander. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 38–45, Online, October 2020. Association for Computational Linguistics.
- [20] Loshchilov Ilya and Hutter Frank. Decoupled weight decay regularization. In **International Conference on Learning Representations**, 2019.
- [21] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In **International Conference on Learning Representations**, 2020.

## A ラベル頻度グループ別の結果

実験に使用したデータの頻度グループ別の結果を表3、表4、表5にそれぞれ示す。データセット中のラベルを出現頻度順に並べたときの上位を High、中位を Middle、下位を Low に分割し、性能を比較した。なお、各グループのラベル数は均等になるように分割し、Reuters-21578 と AAPD、20 Newsgroups の各グループはそれぞれ 30 件、18 件、11 件のラベルから構成されている。

表3 頻度グループごとのラベルにおける実験結果 (Reuters-21578)

Reuters-21578						
Methods/Metrics	macro-f1 (high)	micro-f1 (high)	macro-f1 (middle)	micro-f1 (middle)	macro-f1 (low)	micro-f1 (low)
BCE	0.8249	0.9220	0.7077	0.7465	0.2686	0.4209
FOC	0.8263	0.9215	0.6465	0.7088	0.1732	0.2810
CBL	0.8100	0.9175	0.6633	0.7137	0.1957	0.3269
CBL-NTR	0.8205	0.9191	0.7283	0.7556	0.2827	0.4065
DBL	0.8243	0.9180	0.7269	0.7522	0.3173	0.4448
ASL	0.8349	0.9213	0.7533	0.7838	0.3446	0.4869
ASLWC	<b>0.8446</b>	<b>0.9256</b>	<b>0.7793</b>	<b>0.8063</b>	<b>0.3833</b>	<b>0.5473</b>

表4 頻度グループごとのラベルにおける実験結果 (AAPD)

AAPD						
Methods/Metrics	macro-f1 (high)	micro-f1 (high)	macro-f1 (middle)	micro-f1 (middle)	macro-f1 (low)	micro-f1 (low)
BCE	0.6756	0.7586	0.5438	0.5859	0.2770	0.3564
FOC	0.6687	0.7584	0.5526	0.5988	0.2513	0.3370
CBL	0.6703	0.7597	0.5530	0.5960	0.2762	0.3399
CBL-ntr	0.7088	0.7753	0.6006	0.6282	0.3517	0.4261
DBL	0.7017	0.7675	0.6034	0.6301	0.3626	0.4232
ASL	0.7274	0.7857	<b>0.6286</b>	<b>0.6464</b>	0.4328	0.4650
ASLWC	<b>0.7299</b>	<b>0.7869</b>	0.6227	0.6429	<b>0.4487</b>	<b>0.4826</b>

表5 頻度グループごとのラベルにおける実験結果 (20 Newsgroups)

20 Newsgroups						
Methods/Metrics	macro-f1 (high)	micro-f1 (high)	macro-f1 (middle)	micro-f1 (middle)	macro-f1 (low)	micro-f1 (low)
BCE	0.9167	<b>0.9189</b>	<b>0.9158</b>	<b>0.9160</b>	<b>0.8213</b>	<b>0.8219</b>
FOC	0.9142	0.9153	0.9065	0.9074	0.8156	0.8160
CBL	0.9160	0.9178	0.9114	0.9119	0.8180	0.8195
CBL-NTR	0.9172	0.9181	0.9100	0.9105	0.8171	0.8175
DBL	0.9135	0.9153	0.9110	0.9111	0.8126	0.8129
ASL	<b>0.9169</b>	0.9178	0.9114	0.9114	0.8212	0.8219
ASLWC	0.9159	0.9179	0.9109	0.9109	0.8183	0.8189