LPLS: A Selection Strategy Based on Pseudo-Labeling Status for Semi-Supervised Active Learning in Text Classification

Chun-Fang Chuang[1], Dongyuan Li[1], Satoshi Kosugi[1], Kotaro Funakoshi[1], Manabu Okumura[1]
[1]Tokyo Institute of Technology
{chunfang, kosugi, funakoshi, oku}@lr.pi.titech.ac.jp
lidy94805@gmail.com

Abstract

This paper proposes a new data selection strategy, the Least Pseudo-Labeling Status (LPLS) strategy, for semi-supervised active learning (SSAL). A selection strategy is used in the active learning phase in SSAL to ask for the gold labels for a limited amount of unlabeled data. The proposed LPLS strategy considers the pseudo-labeling status resulting from the semi-supervised learning phase in SSAL. Our SSAL method is based on JointMatch (Zou and Caragea, 2023), the state-of-the-art SSL method, which uses multiple models for automatic pseudo-labeling for unlabeled data. The proposed strategy utilizes these multiple models to measure the label uncertainty of a data point based on not only the intra-model uncertainty (entropy) but also the inter-model uncertainty (divergence). Our text classification experiments on three common benchmark datasets confirm that our proposed SSAL method using the LPLS strategy outperforms both JointMatch and AcTune (Yu et al., 2022), the state-of-the-art SSAL.

1 Introduction

The large amount of labeled data is a key to the successful results in text classification. However, not everyone can afford the high annotation costs required to train a model in real-life applications. Semi-supervised learning (SSL) and active learning (AL) mitigate this annotation cost issue. SSL is to automatically leverage large unlabeled data during the training process based on the initially provided small labeled data. On the other hand, AL is a human-in-the-loop approach, which iteratively queries the gold label for a data point in unlabeled data to the oracle (typically, a human annotator) during training. By selecting the most informative data points according to a particular *selection strategy*, AL tries to achieve the highest performance with a minimum annotation cost.

Semi-supervised active learning (SSAL) integrates SSL and AL (Wang et al., 2017; Gao et al., 2020; Yu et al., 2022). As AL and SSL only select data from one side in terms of confidence score, that is, SSL selects data with high confidence and AL selects data with low confidence, basically each of them can work independently of the other in a complementary way. However, we speculate that a synergy can be achieved by coupling them more tightly. Specifically, in previous studies, the performance of SSL methods is affected by pseudo-labeling. Therefore, we explore the possibility of helping SSL by gaining ground-truth data from AL with a novel selection strategy that considers SSL.

This paper proposes the least pseudo-labeling status (LPLS) selection strategy, an SSAL selection strategy considering the SSL status of pseudo-labeling, for a better integration of SSL and AL in SSAL. Pseudo-labeling (Xie et al., 2020; Sohn et al., 2020; Zhang et al., 2021; Zou and Caragea, 2023) generates the artificial labels for data whose predictions are confident. In other words, the data whose confidence scores pass the threshold will gain pseudo-labels. Our proposed overall SSAL method, which is illustrated in Figure 1, is based on the state-of-the-art (SOTA) SSL method, JointMatch (Zou and Caragea, 2023). In the AL part of our proposed SSAL method, priority is given to the class with the fewest pseudo-labeled instances from JointMatch. Then the LPLS strategy selects the most uncertain data point within the prioritized class. Additionally, for a better uncertainty estimation, multiple models used and tuned in JointMatch are also utilized, as a

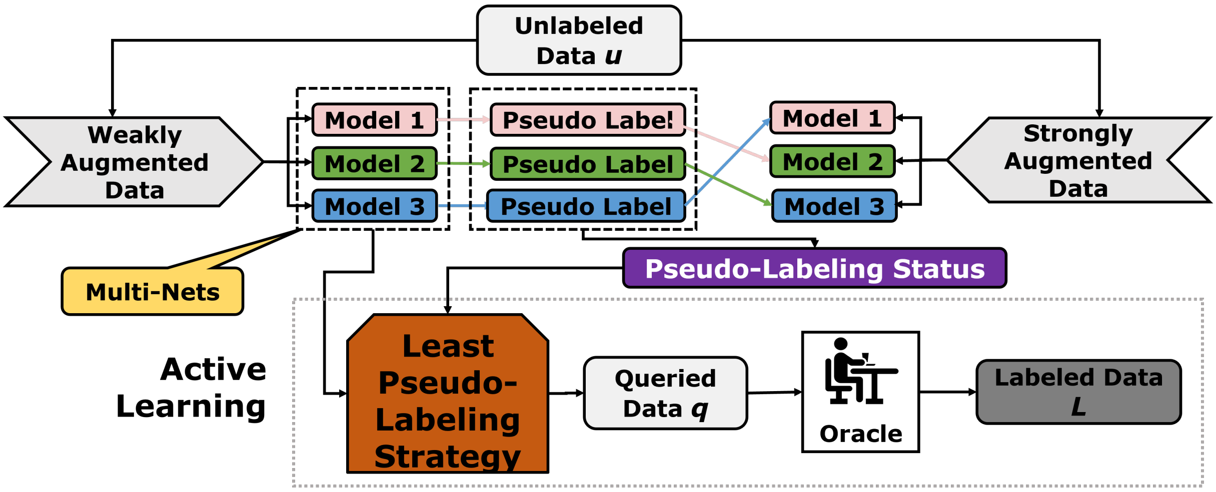

Figure 1: Overview of our SSAL method, which embeds the JointMatch SSL method (Zou and Caragea, 2023) in an AL framework. The upper half of the figure corresponds to JointMatch, which originally uses only two classification models (nets), while our version allows more than two nets (Multi-Nets). The lower half is the AL part, in which we use our LPLS selection strategy for querying the oracle. The LPLS strategy refers to Pseudo-Labeling Status and reuses the nets from JointMach SSL.

variant of query-by-committee (QBC) selection strategy (Seung et al., 1992), which leverages multiple models to select the data with the largest disagreement between the models. The final classification model is also obtained as an ensemble of the trained multiple models.

We evaluate our proposed LPLS-based SSAL method on three benchmark datasets, comparing it against JointMatch and the SOTA method AcTune (Yu et al., 2022). In our experiments, our proposed SSAL method outperfroms all baselines.

The contributions of this paper are: 1) the new effective AL selection strategy LPLS that considers the pseudo-labeling status of SSL in SSAL, 2) the new SOTA SSAL method that utilizes the proposed LPLS strategy, 3) additional experiments demonstrate that using more than two models does not lead to a better result while considering both intra-model uncertainty (entropy) and inter-model uncertainty (divergence) does.

2 Related Work

2.1 Semi-Supervised Learning

Semi-Supervised Learning (SSL) is a learning method to reduce the annotation cost by leveraging a large amount of unlabeled data. UDA (Xie et al., 2020) proposes the combination of data augmentation techniques such as backtranslation and a consistency regularization to reduce the distance of predicted results between different augmented data. MixText (Chen et al., 2020) proposes TMix, which interpolates labeled and unlabeled data to overcome the limitation of using them separately. FixMatch (Sohn et al., 2020) takes the prediction results of weakly augmented data as the pseudo-label of strongly augmented data. FlexMatch (Zhang et al., 2021) proposes curriculum pseudo-labeling which applies flexible thresholds adjusted by pseudo-labeling status (PLS). JointMatch (Zou and Caragea, 2023) trains two differently initialized models and uses them to teach each other in a cross-labeling manner to alleviate error accumulation. In SSL of our SSAL framework, we utilize pseudo-labeling and consistency regularization. Furthermore, based on JointMatch, we construct our SSAL framework as multi-nets framework and we utilize PLS not only in SSL but in AL to select the helpful data for SSL. We also take JointMatch as one of our baselines.

2.2 Active Learning

Active Learning (AL) is a learning method that achieves high performance with minimal labeling cost by querying data with the oracle. AL can selectively query the most informative data from a large pool of unlabeled data and send the selected data to be annotated by the oracle. There are vari-

Algorithm 1 The Least Pseudo-Labeling Status selection strategy

Input: s, the pseudo-labeling status
Input: N_C, the number of classes
Input: $U = \{u_i \mid i \in (1, 2, \ldots, N_U)\}$, a set of unlabeled data
Input: $M = \{m_i \mid i \in (1, 2, \ldots, N_M)\}$, a set of differently initialized models
Output: the data point to query

// Set the target query class with the least pseudo-labeling status value
$query\_class \leftarrow \text{argmin}(s)$

// Calculate the uncertainty of each data point
$D \leftarrow \{\}$ // a set to hold the data with content, uncertainty, and prediction results
for u **in** U **do**
  $P \leftarrow \{P_j = \text{predict}(m_j, u) \mid j \in (1, \ldots, N_M)\}$ // all models predict for the unlabeled data
  $entropy \leftarrow \frac{-1}{N_M} \sum_{j=1}^{N_M} \sum_{c=1}^{N_C} P_j^c \log(P_j^c)$ // calculate the mean entropy
  $divergence \leftarrow \frac{1}{N_M(N_M-1)} \sum_{i=1}^{N_M} \sum_{j=1, j\neq i}^{N_M} \text{KLD}(P_i \| P_j)$ // calculate the divergence
  $uncertainty \leftarrow entropy \cdot divergence$ // calculate the uncertainty
  $prediction \leftarrow \text{argmax}(\frac{1}{N_M} \sum_{i=1}^{N_M} P_i)$ // obtain the prediction result (class index number)
  $d \leftarrow (u, uncertainty, predicition)$; $D \leftarrow D \cup \{d\}$
end for

// Select the data point to query
$L \leftarrow \text{sortByUncertainty}(D)$ // sort D in descending order based on the uncertainty of each data point
for d_i in L **do**
  if $d_i.prediction = query\_class$ **then**
    return d_i // d_i is the i-th element of sorted list L
  end if
end for
return d_1 // return the data point with the highest uncertainty because we did not find a data point which matches the query class

ous query strategies in AL. In our proposed method, we apply uncertainty-based sampling and the disagreement-based strategy. Uncertainty sampling prefers the most uncertain instances and disagreement-based strategies utilize multiple models to select the data which has the most disagreement among the models. The most well-known disagreement-based method is QBC (Seung et al., 1992) which trains a distinct group of models to select the data with the greatest disagreement. In our work, we merge both QBC and uncertainty-based methods to measure uncertainty in AL for semi-supervised active learning (SSAL).

2.3 Semi-Supervised Active Learning

Both SSL and AL aim to reduce annotation costs while achieving high performance. Therefore, recent studies have started to explore whether these two methods can be used simultaneously. Gao et al. (2020) proposes a query selection strategy for SSAL. The selection strategy in the paper is to select the data based on the difference in predictions between augmentations and the original data. AcTune (Yu et al., 2022) proposes a region-aware querying strategy and a momentum-based method to enforce both the informativeness and the diversity of queried samples during AL and alleviate the label noise in self-training. We select the SOTA SSAL method, AcTune, as one of our baseline models.

3 Proposed SSAL Method

In this section, we introduce our proposed selection strategy, the Least Pseudo-Labeling Status (LPLS) strategy, for semi-supervised active learning (SSAL). This section will describe the key concepts in LPLS, (1) pseudo-labeling, (2) measuring uncertainty, and (3) selection strategy.

Figure 1 shows the pipeline of our SSAL method based on JointMatch SSL (Zou and Caragea, 2023). The upper half of the figure corresponds to JointMatch SSL. The lower half is the AL part, in which we use our proposed LPLS selection strategy (illustrated in Figure 2) for querying the oracle. The LPLS strategy refers to pseudo-labeling status and reuses the nets from JointMatch. At the last

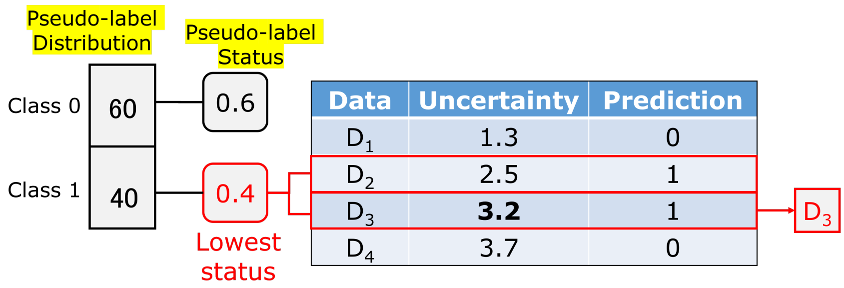

Figure 2: Illustration of the LPLS strategy for AL. The most uncertain data point in the prediction class of the least pseudo-label status value at the moment is queried for the gold label.

of this section, the overall SSAL training procedure on the presented pipeline will be described.

3.1 Pseudo-Labeling

Pseudo-labeling is a technique used in SSL where a model trained on a small labeled dataset is used to predict labels for an unlabeled dataset. There are different SSL works applying pseudo-labeling. UDA relied on the principle of consistency regularization where the model is encouraged to produce consistent predictions for augmented and original versions of the same data. FlexMatch proposed the concept of flexible thresholds which adjusts the value of thresholds on each class for pseudo-labeling during the training based on pseudo-labeling status (PLS). Inspired by FlexMatch and UDA, our proposed method also applies pseudo-labeling with flexible thresholds. Furthermore, we apply cross-labeling which means the pseudo-labels from one model are used in another model to filter out more noise in pseudo-labeling.

We adopt the same data augmentation techniques with JointMatch for fair comparisons. In detail, we use backtranslation for strong augmentation and synonym replacement for weak augmentation.

To enhance the synergy between SSL and AL, our SSAL utilizes PLS in AL. Pseudo-labeling status is the representation of the distribution of pseudo-labels at each time step t across each class. At first, s_0 is $[1/C,\ 1/C,\ 1/C...,\ 1/C]$ where C is the number of classes. Then at each time step t during training, s_t will be $[s_t(1), s_t(2), s_t(3), ..., s_t(C)]$ where $s_t(c)$ is the pseudo-labeling status of class c, that is,

$$s_t(c) = \frac{\textit{the number of pseudo-labels in } c}{\textit{the number of all pseudo-labels}}.$$

By observing the status of pseudo-labels in each class, we can identify which class is not learning well and prioritize AL to obtain data that likely belongs to that class.

We extend the cross-labeling of JointMatch to enable more than two nets. Our extension simply matches i-th model to $((i+1) \mod M)$-th model, where M is the number of models, as shown in Figure 1.

3.2 Measuring Uncertainty

In LPLS, the purpose of this strategy is to select the uncertain data whose prediction is the same as the needy class based on the status of pseudo-labeling. To achieve this goal, we need to measure the uncertainty. Our framework is multi-nets. To make full use of it, we utilize the multiple nets as query by committee (QBC) and consider the uncertainty from each model. QBC is an active learning algorithm where a committee of models, each trained on the current labeled dataset, is used to select the most informative samples from a pool of unlabeled data. The main idea is to identify samples on which the committee members disagree the most, as these samples are considered the most informative for improving the model. But in LPLS, we consider not only the disagreement of models' prediction for unlabeled data but also the total uncertainty of models. The uncertainty of a data sample is calculated as follows based on mean entropy and mean divergence:

$$\textit{Uncertainty} = \textit{Entropy} \cdot \textit{Divergence}.$$

Entropy and *Divergence* are defined as below:

$$Entrropy = \frac{1}{N_M} \sum_{i=1}^{N_M} Ent_i,$$

$$Divergence = \frac{\sum_{f=1}^{N_M} \sum_{g=1, g \neq f}^{N_M} \mathrm{KLD}(P_f || P_g)}{N_M(N_M - 1)},$$

where Ent_i means the prediction entropy in i-th model for the given data point and $KLD(P_f||P_g)$ means KL-Divergence of predictions between two models f and g.

3.3 Least Pseudo-Labeling Status Selection Strategy

In this paper, we proposed a new query strategy, LPLS, for SSAL. In SSL with pseudo-labeling, the performance is largely affected by the quality of pseudo-labeling. However, in the past, SSL only leverages unlabeled data whose prediction confidence is above the fixed threshold and there are some classes which the model has difficulty learning. Therefore, FlexMatch proposed Curriculum Pseudo-Labeling (CPL) which adjusts thresholds for each class actively based on PLS. However, in SSAL, the model has the opportunity to obtain ground-truth data during training. Based on this characteristic, we proposed LPLS, a query strategy considering the PLS.

The pseudo-code of our strategy is presented in Algorithm 1. First, for each unlabeled data, we use all models to make predictions and calculate the sum of uncertainty from each model's prediction on this data. Then, we multiply this sum by the total difference between the models' predictions to get the total uncertainty. This uncertainty score serves as the ranking order for each data to be compared. Finally, we consider the pseudo-labeling status s. The class with the lowest value is the queried class we are looking for. We start to compare data from the highest uncertainty. If the prediction result of the compared data matches the queried class, we select this data and send it to the oracle. On the other hand, if we can not find the data in the queried class, the strategy will send the data with the highest uncertainty to the oracle.

3.4 SSAL Training Procedure

Algorithm 2 shows our SSAL training procedure. The procedure interleaves the SSL part and the AL part up to N_E times. First, the SSL part conducts pseudo-labeling-based semi-supervised learning on unlabeled data using differently initialized N_M models. Then, the LPLS strategy selects a data point q, which is queried to the oracle for its gold label[1]. The AL part is skipped once the amount of labeled data reaches the annotation cost limit $L_F \cdot C$.

Algorithm 2 Our SSAL training procedure

Input: N_C, the number of classes
Input: N_L, the number of initial data per class
Input: $\alpha(L) = \{\alpha(l_i) \mid i \in (1, 2, \ldots, N_L \cdot N_C)\}$,
  a set of labeled data
Input: $U = \{u_i \mid i \in (1, 2, \ldots, N_U)\}$,
  a set of unlabeled data // $N_L << N_U$
Input: $M = \{m_i \mid i \in (1, 2, \ldots, N_M)\}$,
  a set of differently initialized models
Input: N_F, the final annotation amount limit
Input: N_E, the max number of SSL epochs
  // $N_L < N_F$ and $N_F \cdot N_C < N_E$
for *epoch* $= 1$ to N_E **do**
  // SSL
  for $m \in M$ **do**
    m.supervised-learning(L)
  end for
  $(s, M) \leftarrow$ MultiNetsSSL(N_C, M, U) // s: PLS
  // AL
  if $|L| < N_F \cdot N_C$ **then**
    $q \leftarrow$ LPLS(s, N_C, U, M) // Algorithm 1
    $l \leftarrow$ oracle(q) // obtain the label of q
    $U \leftarrow U \setminus \{q\}$ // remove q from U
    $L \leftarrow L \cup \{(q, l)\}$ // add the new instance
  end if
end for
return ensemble(M) // return the final model

	AL-QBC	**AcTune**	**JointMatch**	**Ours**
Type	AL	SSAL	SSL	SSAL
Multi-Nets	✓	✗	✓*	✓
LPLS	✗	✗	–	✓

* The original JointMatch uses only two nets, while our extension enables more than two.

Table 1: Qualitative comparisons to baselines. Our proposed method is an SSAL method based on JointMatch SSL. Although our selection strategy is a variant of QBC, our method takes the pseudo-labeling status into consideration. Moreover, our framework utilizes multiple nets in SSL and AL, while AcTune, the SOTA SSAL, does not.

4 Experiments

4.1 Baselines

We consider three baselines for comparison, that is, AL-QBC, AcTune, and JointMatch.

[1] Our experiments emulate unlabeled data by using labeled datasets, where the gold labels are accessible for models only through the oracle, except for the initial small portion of labeled data.

Dataset	Label Type	N_C	#Training	#Validation	#Test	N_L	N_F
AG News	Topic	4	5000	2000	1900	25	35
Yahoo! Answer	Topic	10	5000	2000	6000	27	35
IMDB	Sentiment	2	5000	1000	12500	25	35

Table 2: Dataset statistics and splits. The numbers of training, validation and test data mean the number of data points per class. N_C, N_L, N_F are defined in Algorithm 2. N_L data points are randomly sampled from the training data per class to be included in labeled data L. The remaining training data are used as unlabeled data U.

	AG News			Yahoo!			IMDB		
Methods	Accuracy	Macro-F1	p	Accuracy	Macro-F1	p	Accuracy	Macro-F1	p
AL-QBC	0.821	0.819	**	0.607	0.598	**	0.735	0.736	**
AcTune	0.877	0.877	*	0.666	0.661	**	0.791	0.790	+
JointMatch	0.881	0.880	*	0.675	0.667	*	0.753	0.752	**
Ours	**0.885**	**0.885**	–	**0.681**	**0.673**	–	**0.796**	**0.792**	–

Table 3: Performance results. The best in each column is marked in bold. * and ** indicate a difference to our method using the proposed LPLS strategy with statistical significance of $p < 0.05$ and $p < 0.01$, respectively. $^+$ indicates a significant tendency of $p < 0.1$. Multiple testing correction is not applied.

AL-QBC is a pure AL framework utilizing a QBC strategy. The implemented QBC strategy is equivalent to our LPLS (Algorithm 1) except that it always returns d_1, the most uncertain data point in unlabeled data U. AcTune (Yu et al., 2022) is the SOTA SSAL text classification method. JointMatch (Zou and Caragea, 2023) is the SOTA SSL text classification method, on which our method is based. Table 1 clarifies the relationship between each method and our method. Our SSAL method is the only one to utilize multi-nets and consider PLS in AL.

4.2 Datasets

We evaluate LPLS on common text classification datasets: IMDB (Pal et al., 2020), AG News (Zhang et al., 2015) and Yahoo! Answers (Chang et al., 2008). Following JointMatch (Zou and Caragea, 2023), we use the original test set and randomly sample from the training set to construct our training labeled set, and training unlabeled set. Table 2 shows the dataset statistics and split information.

4.3 Experimental Setups

Following JointMatch, we used the BERT-based-uncased model[2] as our backbone model and the HuggingFace Transformers library for the implementation.

The training procedure of our method followed Algorithm 2. However, after passing $(N_F - N_L) \cdot N_C$ steps, the training could be early-stopped before reaching the N_E-th step based on performance check on validation data. We set N_E to 100. AcTune and LPLS were trained in accordance with this procedure. The training of AL-QBC was stopped at the N_E-th step if there was no early-stopping.

To verify the feasibility of our approach, as shown in Table 2, we set the total annotation cost of AL as $N_F = 35$ multiplied by the number of classes N_C and start with $N_L = 25$ annotated samples per class for AG News, IMDB, and start with $N_L = 27$ annotated samples per class for Yahoo!.[3] To make fair comparisons, we provide JointMatch with $N_F \cdot N_C$ samples as the initial small training data L, while other AL methods receive only $N_L \cdot N_C$ at first.

4.4 Comparisons with Baselines

We summarize the comparison with baselines on different text classification datasets in Table 3. We reproduced the baseline results. All results in table 3 are the average of five runs. We conducted McNemar's test (McNemar, 1947) between each baseline method and our proposed method on the three datasets separately.

[2] https://huggingface.co/google-bert/bert-base-uncased

[3] We slightly boosted the start-up with extra samples as the Yahoo! dataset has more classes.

	AG News		Yahoo!		IMDB	
# of nets	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
2	0.885	0.885	0.681	0.673	0.796	0.792
3	0.883	0.883	0.680	0.674	0.788	0.786

Table 4: Comparison between different numbers of nets in the proposed Multi-Nets SSAL method.

Selection Strategies		AG News		Yahoo!		IMDB	
PLS	Uncertainty	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
✗	Random	0.876	0.875	0.663	0.660	0.759	0.756
✗	Entropy	0.880	0.878	0.673	0.672	0.776	0.772
✗	Divergence	0.881	0.881	0.670	0.664	0.780	0.778
✗	Entropy · Divergence	0.882	0.882	0.670	0.664	0.784	0.782
✓	Random	0.879	0.879	0.665	0.660	0.768	0.766
✓	Entropy	0.882	0.881	0.675	0.670	0.795	0.794
✓	Divergence	0.880	0.880	0.674	0.667	0.786	0.787
✓	Entropy · Divergence	**0.885**	**0.885**	**0.681**	**0.673**	**0.796**	**0.792**

Table 5: Comparison between different selection strategies. The bottom row corresponds to our proposed LPLS selection strategy. The column labeled as PLS indicates if it considers PLS.

We compared our experimental results with AcTune, JointMatch, and AL-QBC. LPLS outperforms all baselines on all benchmark datasets. Our method has better results than JointMatch by about 0.5% points on AG News, Yahoo!. On IMDB, our method surpasses JointMatch by about 4% points but AcTune only by 0.5%.

Although our method presents higher scores than others in terms of accuracy and F1 score, it does not have a statistical significance over AcTune in IMDB. We consider that the reason is because the trait of our LPLS and IMDB is a two-class classification. LPLS is trying to raise the opportunity of producing the labeled data in the least class in PLS. Our method comes from the imbalance of pseudo-labeling. However, IMDB is a two-class classification task. If one class is classified as particularly effective, it can also directly improve the learning performance of the other class. Therefore, it is more challenging to become more effective in our method during training.

4.5 Ablation Studies

4.5.1 Multi-Nets SSAL with different numbers of models

Because our proposed method is a framework of multiple networks, we explore if the accuracy will be improved when the number of models increases. The results have been shown in Table 4. As shown in Table 4, using more models will not yield better results.

4.5.2 Multi-Nets SSAL with Different Settings

Our selection strategy in AL leverages PLS in SSL. To evaluate the effectiveness of our proposed strategy, there is a comparison between cases where PLS is considered and those where it is not, as well as the comparison of different uncertainty measuring methods. The results are shown in Table 5. In the situation without considering PLS, although our proposed measuring method for uncertainty has the best result, it just improves a little by the other uncertainty methods.

5 Conclusion

We proposed a query selection strategy based on pseudo-labeling status for semi-supervised active learning (SSAL) and empirically confirmed the effectiveness of the proposed selection strategy on text classification. Our method is inspired by the observed impact that pseudo-labeling status (PLS) affects a lot in semi-supervised learning (SSL) with pseudo-labeling. Furthermore, in SSAL, the model has the opportunity to obtain correctly labeled data which helps improve SSL performance. Therefore, we proposed a data selection strategy based on PLS.

We demonstrated that our proposed method can outperform or compete with AL-QBC, AcTune, and JointMatch across all benchmark datasets. We also explored the performance of using three models but the results show that adding more models can not improve the accuracy. Our selection strategy is better than the entropy-based selection method which shows our framework is effective. We hope that our research can raise the importance of PLS in SSAL.

References

Ming-Wei Chang, Lev Ratinov, Dan Roth, and Vivek Srikumar. 2008. Importance of semantic representation: dataless classification. In *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 2*, AAAI'08, page 830–835. AAAI Press.

Jiaao Chen, Zichao Yang, and Diyi Yang. 2020. MixText: Linguistically-informed interpolation of hidden space for semi-supervised text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2147–2157, Online. Association for Computational Linguistics.

Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan O. Arik, Larry S. Davis, and Tomas Pfister. 2020. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. *Preprint*, arXiv:1910.07153.

Quinn McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.

Aditya Pal, Abhilash Barigidad, and Abhijit Mustafi. 2020. Imdb movie reviews dataset.

H. S. Seung, M. Opper, and H. Sompolinsky. 1992. Query by committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, page 287–294, New York, NY, USA. Association for Computing Machinery.

Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, volume 33, pages 596–608. Curran Associates, Inc.

Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. 2017. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(12):2591–2600.

Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. 2020. Unsupervised data augmentation for consistency training. In *Advances in Neural Information Processing Systems*, volume 33, pages 6256–6268. Curran Associates, Inc.

Yue Yu, Lingkai Kong, Jieyu Zhang, Rongzhi Zhang, and Chao Zhang. 2022. AcTune: Uncertainty-based active self-training for active fine-tuning of pretrained language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1422–1436, Seattle, United States. Association for Computational Linguistics.

Bowen Zhang, Yidong Wang, Wenxin Hou, HAO WU, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 18408–18419. Curran Associates, Inc.

Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Henry Zou and Cornelia Caragea. 2023. JointMatch: A unified approach for diverse and collaborative pseudo-labeling to semi-supervised text classification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7290–7301.