Accelerating Diversity Sampling for Deep Active Learning By Low-Dimensional Representations

Sandra Gilhuber[1], Max Berrendorf[1], Yunpu Ma[1], and Thomas Seidl[1]

Ludwig-Maximilians-Universität München, Munich, Germany
{gilhuber,berrendorf,ma,seidl}@dbs.ifi.lmu.de

Abstract. Selecting diverse instances for annotation is one of the key factors of successful active learning strategies. To this end, existing methods often operate on high-dimensional latent representations. In this work, we propose to use the low-dimensional vector of predicted probabilities instead, which can be seamlessly integrated into existing methods. We empirically demonstrate that this considerably decreases the query time, i.e., time to select an instance for annotation, while at the same time improving results. Low query times are relevant for active learning researchers, which use a (fast) oracle for simulated annotation and thus are often constrained by query time. It is also practically relevant when dealing with complex annotation tasks for which only a small pool of skilled domain experts is available for annotation with a limited time budget. Our code is available at: `https://github.com/sobermeier/low-dim-div-sampling`.

Keywords: Active Learning · Diversity Sampling

1 Introduction

Deep neural networks are the dominant choice for solving complex tasks, such as image classification. Their great success depends in large part on the availability of a sufficient amount of labeled data. Especially in domains with scarce publicly available data, such as medical or industrial applications, annotations can become prohibitively expensive due to the need for skilled domain experts. The field of active learning thus aims at reducing the number of required annotations by intelligently selecting instances for labeling. Since modern networks require a significant amount of training time, the traditional setting where instances are selected one after the other [13,15,20] has become infeasible [17], and a batch-setting is commonly applied, where a fixed number of instances is selected for annotation.

State-of-the-art approaches [3,9,18,19,16] follow two different paradigms (or a mixture thereof): In *uncertainty*-based methods [4,5,10], those instances are selected for which the model is the least certain about the prediction. In contrast, *diversity* methods [3,6,7,16,18,19,22] focus on selecting a representative subset of instances and avoid re-labeling similar instances. In this work, we focus on the second class.

Diversity-based methods often rely on high-dimensional representations extracted from the model's last layers [3,6,7,8,11,16,18,22,21]. In the presence of a large pool of unlabeled data, processing these representations can become a bottleneck of the approaches resulting in increased query times. While these can often be neglected when the annotation is delegated to a large pool of on-demand crowd workers, in settings where domain experts are required, there is often only a small number of available annotators with tight schedules. In these settings, it is desirable to reduce the query time in addition to only requesting useful instances for annotation. Similarly, in active learning research, where a simulated oracle is used for annotation, the computational bottleneck is often the instance selection.

2 Diversity Sampling on Low-Dimensional Representations

In this work, we present a simple yet effective approach to accelerate diversity-based methods, which replaces the high-dimensional latent features $\mathbf{x} \in \mathbb{R}^d$ by the vector of predicted class probabilities $\mathbf{p} \in \mathbb{R}^c$, where usually $c \ll d$. The approach can be applied to most diversity-based methods without large modifications and effectively reduces the instance selection times.

We empirically evaluate our approach with multiple different diversity-based active learning heuristics. Note that we do not consider uncertainty in this work and focus only on underlying diversity concepts. However, the selected diversity methods are key concepts of various popular active learning strategies, such as [1,3,16,18,22].

1. `KMeansCenter` selects the points closest to the centroids of k-means clustering [14] with $k = q$ clusters for annotation, where q denotes the query size. As a recent example, CLUE [16] uses k-means clustering as diversity concept enriched by uncertainty weighting.
2. `KCenterGreedy` iteratively selects the sample with the largest minimum distance to any already labeled instance. It is also known as CoreSet [18] and one of the first solely diversity-based active learning methods.
3. `KMeans++` [2] iteratively samples instances with probability proportional to the minimum distance to already selected points in the current acquisition round. BADGE [3] is a prominent example using `KMeans++` on high-dimensional vectors.

For the iterative `KCenterGreedy` and `KMeans++` algorithms, we keep an array of minimum distance to already labeled samples, and update it whenever we add another sample for labeling. The time complexity of selecting one batch of queries is given in Table 1. Notice that for all heuristics, the time complexity linearly depends on the vector dimension.

We empirically evaluate the MNIST [12] dataset of handwritten digits with 10 classes and a simple 2-layer fully-connected network with embedding dimensionality 256 as in [3] for a proof-of-concept. The learning rate is set to 0.01,

Table 1. Time complexity of a single acquisition round of the different diversity-based heuristics. q denotes the query size, i.e., number of instances to select for labeling, n_l/n_u the number of labeled/unlabeled samples ($n_l \ll n_u$), d the vector dimensionality, and i the number of iterations until convergence.

Algorithm	Time Complexity
KMeansCenter	$\mathcal{O}(q \cdot n_u \cdot i \cdot d)$
KCenterGreedy	$\mathcal{O}(n_l \cdot n_u \cdot d + q \cdot n_u)$
KMeans++	$\mathcal{O}(q \cdot n_u \cdot d)$

and we train the network from scratch for 10 epochs in each iteration. The initial pool contains 100 randomly chosen samples, and we select additional 100 instances per active learning iteration until a budget of 2,500 samples is exhausted. We investigate three different input features $\mathbf{x}$ of the samples as input to the heuristics:

1. the full-dimensional latent features, i.e., $\mathbf{x} \in \mathbb{R}^d$,
2. the vector of predicted class probabilities, i.e., $\mathbf{x} \in \mathbb{R}^c$, where $c = 10$ denotes the number of classes,
3. PCA-reduced features, i.e., $\mathbf{x} \in \mathbb{R}^{d'}$, where $d' \ll d$ is the reduced dimension. For comparability, we use the same dimensionality $d' = c = 10$ for PCA.

Our results are shown in Fig. 1. The first column shows the accuracy vs. the number of acquired labels. We observe that using the vector of predicted probabilities not only maintains the performance of full-dimensional latent features but also surpasses it for all three investigated diversity-based heuristics. In contrast, PCA-reduced latent features result in comparable performance. The third column compares the number of acquired labels against the cumulative query time. Using the vector of predicted probabilities generally shows the lowest cumulative runtime. Compared to using the output vectors, PCA requires an extra step and is therefore somewhat weaker in terms of query times. However, using full-dimensional latent features can lead to more than four-fold increased cumulative query time depending on the heuristic, even in this relatively small toy setting. The second column then combines both plots and shows the accuracy vs. the cumulative query time, demonstrating that both label efficiency and query times benefit from our proposed method.

3 Conclusion

In this paper, we proposed to use the vector of predicted probabilities instead of the high-dimensional latent features as input to diversity-based active learning methods. As a proof-of-concept, we demonstrated on one dataset that for several diversity-based heuristics, we could strongly reduce the query time while at the same time improving the performance. Since the predicted probabilities of the unlabeled data are usually exploited anyway during the active learning process, no additional computations are required.

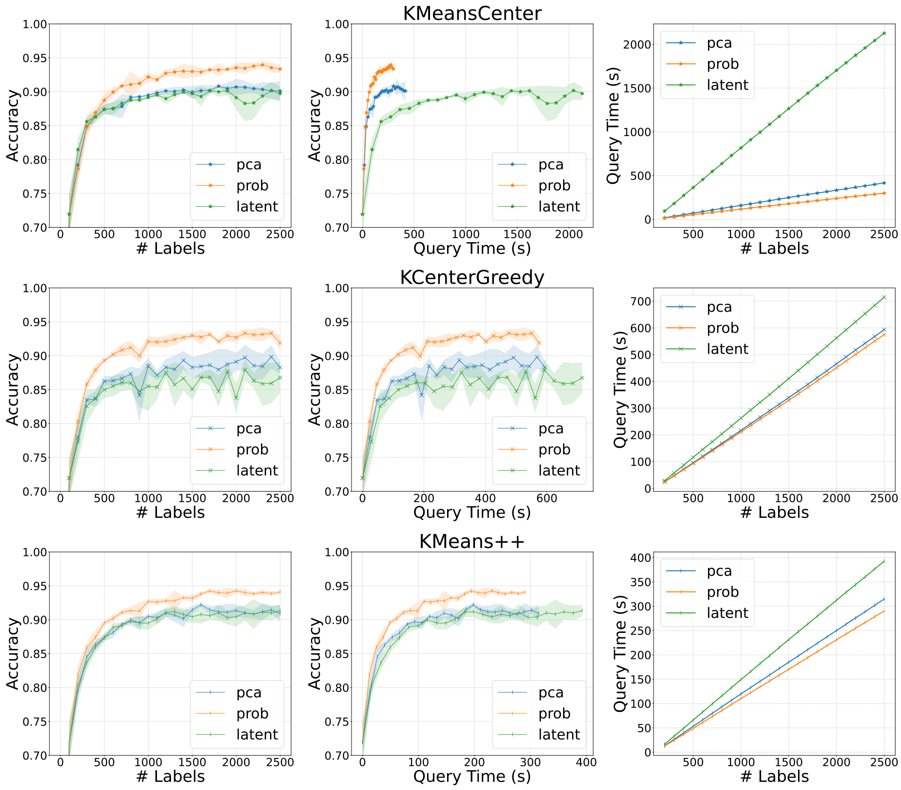

Fig. 1. Comparison of the different techniques for three different acquisition functions. The first column shows the accuracy w.r.t. the number of labels, the second column accuracy vs. cumulative query time, and the last column the cumulative query time vs. the number of acquired labels.

For future work, we would like to investigate this promising direction further, particularly how well the insights transfer to other datasets and how to best combine it with uncertainty-based methods. As an interesting observation, using samples with diverse predicted probabilities might also implicitly lead to selecting points of diverse uncertainty.

Acknowledgements

This work has been funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibilities for its content.

References

1. Abraham, A., Dreyfus-Schmidt, L.: Sample noise impact on active learning. arXiv preprint arXiv:2109.01372 (2021)
2. Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. Tech. rep., Stanford (2006)
3. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds. In: ICLR (2020)
4. Gal, Y., Islam, R., Ghahramani, Z.: Deep bayesian active learning with image data. In: ICML. pp. 1183–1192 (2017)
5. Gao, M., Zhang, Z., Yu, G., Arık, S.Ö., Davis, L.S., Pfister, T.: Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In: ECCV. pp. 510–526 (2020)
6. Geifman, Y., El-Yaniv, R.: Deep active learning over the long tail. arXiv preprint arXiv:1711.00941 (2017)
7. Gissin, D., Shalev-Shwartz, S.: Discriminative active learning. arXiv preprint arXiv:1907.06347 (2019)
8. Killamsetty, K., Sivasubramanian, D., Ramakrishnan, G., Iyer, R.: Glister: Generalization based data subset selection for efficient and robust learning. vol. 35, pp. 8110–8118 (2021)
9. Kim, K., Park, D., Kim, K.I., Chun, S.Y.: Task-aware variational adversarial active learning. In: CVPR. pp. 8166–8175 (2021)
10. Kirsch, A., Van Amersfoort, J., Gal, Y.: BatchBALD: Efficient and diverse batch acquisition for deep bayesian active learning. NeuRIPS pp. 7026–7037 (2019)
11. Kothawade, S., Beck, N., Killamsetty, K., Iyer, R.: Similar: Submodular information measures based active learning in realistic scenarios. NeuRIPS **34**, 18685–18697 (2021)
12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791, `https://doi.org/10.1109/5.726791`
13. Li, X., Guo, Y.: Multi-level adaptive active learning for scene classification. In: ECCV. pp. 234–249 (2014)
14. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2), 129–137 (1982)
15. McCallum, A.K., Nigam, K.: Employing EM and pool-based active learning for text classification. In: ICML. pp. 359–367 (1998)
16. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: ICCV. pp. 8505–8514 (2021)
17. Ren, P., Xiao, Y., Chang, X., Huang, P., Li, Z., Gupta, B.B., Chen, X., Wang, X.: A survey of deep active learning. ACM Comput. Surv. **54**(9), 180:1–180:40 (2022). https://doi.org/10.1145/3472291
18. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: ICLR (2018)
19. Sinha, S., Ebrahimi, S., Darrell, T.: Variational adversarial active learning. In: ICCV. pp. 5972–5981 (2019)
20. Tong, S., Koller, D.: Support vector machine active learning with applications to text classification. JMLR **2**(Nov), 45–66 (2001)
21. Wu, T.H., Liu, Y.C., Huang, Y.K., Lee, H.Y., Su, H.T., Huang, P.C., Hsu, W.H.: Redal: Region-based and diversity-aware active learning for point cloud semantic segmentation. In: ICCV. pp. 15510–15519 (2021)

22. Zhdanov, F.: Diverse mini-batch active learning. arXiv:1901.05954 (2019)