Application of active learning algorithm in handwriting recognition numbers

Tiantian Pu*
Hunan University of Technology, Zhuzhou 412099, Hunan Province, China
vivian.wang@cas-harbour.org

Abstract. Active learning is very suitable for many problems in natural language processing, where unlabeled data may be abundant, but annotation is slow and expensive. This article aims to illustrate some active learning methods for handwritten digit recognition tasks, such as the least confidence and entropy methods. We investigated the previously used sequence model query selection strategies and used some selection strategies for sample labeling in handwritten digit recognition. We also conduct a large-scale empirical comparison of using multiple corpora, which shows that our proposed method improves the technical level.

1. Introduction
The emergence of Arabic numerals is a revolution in the field of human computing, which has connected the entire world, and human information activities have entered a new era. Arabic numerals, as the main medium for exchanges and information carriers for economic development in the world, have now become internationally used numbers. In the 21st century, people have entered the era of information and big data. The application of artificial intelligence (such as image recognition, face recognition, speech recognition, etc.) has gradually affected our daily lives. In the era of Internet big data, people need to deal with more data-related tasks every day, such as data statistics, invoice tax bills, bank checks, as well as cargo code sorting, express sorting, etc. How to use computers and other automated tools to identify characters and documents, and liberate manual work Labor, providing work efficiency, has become a problem that needs to be solved at present. Offline handwriting character recognition is the automatic transcription by computer, where only the image of that handwriting is available.[1]

Handwritten Numeral Recognition is a branch of Optical Character Recognition (OCR). Its research object is how to automatically recognize the Arabic numerals handwritten on paper by a computer. Digital recognition belongs to the branch of pattern recognition in the subject category, mainly involving digital signal processing, pattern recognition, artificial intelligence, image processing, fuzzy mathematics, natural language understanding, computer and other subjects.[2] Digital recognition systems can be divided into print recognition and handwritten recognition, and handwritten digital recognition can be divided into real-time handwritten digital recognition (online handwritten digital recognition) and non-real-time handwritten digital recognition (offline handwritten digital recognition).[3] Almuallim and Yamaguchi proposed a structural recognition technique for Arabic handwritten words.[4]

This paper chooses a pool-based active learning algorithm to select sample data through the least confidence method and entropy sampling method. The least confidence method selects the example with
the least confidence in the predicted value as the query instance, and entropy sampling selects the larger amount of information and annotates the data.

2. Active learning strategies

Active learning is a kind of machine learning framework. Its algorithm can relabel the samples with real labels by interacting with users (experts or authorities). The learning process is also called an optimal experimental design. Active learning to use fewer training samples to obtain a better-performing classifier, query the most useful unlabeled samples through a certain algorithm, and hand them over to experts for labeling, and then use the queried samples to train the classification model to improve the accuracy of the model degree.

In some cases, active learning performs better than random sampling. The following figure shows an example of linear classification, illustrating that active learning is more effective than random sampling. It should be noted that the entire data set (a red triangle and green circle) below is linearly inseparable.

![Fig.1 Sampling comparison of different methods](image)

Figure (a) shows a randomly generated two-dimensional data set, with green representing positive examples and red representing negative examples.

Figure (b) adopts a random sampling method, randomly selects 30 samples for training, and the final model prediction accuracy is 70%.

Figure (c) adopts the method of uncertainty sampling and randomly selects 30 samples for training, and the final prediction accuracy of the generated model is 90%.

It can be seen that the active learning strategy has achieved good results.

2.1. Pool-based sample selection algorithm

2.1.1 Method based on uncertainty reduction

This type of method selects those examples whose classification is least certain by the current benchmark classifier for labeling. This type of method uses information entropy as a measure of the amount of information contained in an example, and the example with the largest information entropy is the example that the current classifier is least able to determine its classification. From a geometric point of view, this method preferentially selects examples close to the classification boundary.

The intuition here is that the examples for which the model has the least certainty will likely be the most difficult examples — specifically the examples that lie near the class boundaries. The learning algorithm will gain the most information about the class boundaries by observing the difficult examples.
2.1.2 Method based on version reduction
This type of method selects examples that can minimize the version space after training for labeling. In the binary classification problem, the examples selected by this type of method always almost equally divide the version space.

2.2 Flow-based sample selection algorithm
Most of the pool-based algorithms can be adjusted to adapt to the flow-based situation. However, because the flow-based algorithm cannot compare the unlabeled examples one by one, it is necessary to set a threshold for the corresponding evaluation index of the example. When the evaluation index of the example submitted to the selection engine exceeds the threshold, it will be marked, but this method needs to be targeted at Different tasks are adjusted, so it is difficult to put into use as a mature method.

3. Realization of handwritten digit recognition based on mnist dataset

3.1 Dataset description
This experiment uses the MNIST data set. The MNIST data set (Mixed National Institute of Standards and Technology database) is a large-scale handwritten digit database collected by the National Institute of Standards and Technology. It contains a training set of 60,000 examples and a test of 10,000 examples. Set. This data set is a very classic data set in the field of machine learning. Each sample is a 28 * 28 pixels gray handwritten digital picture.

![Figure 2: Top 25 sample images of the MNIST dataset](image)

3.2 Data selection strategy

3.2.1 Least Confidence
For two-class or multi-class models, they are usually able to score each data to determine which category it is more like. For example, in a two-classification scenario, two data are predicted by a certain classifier, and the predicted probabilities for the two categories are (0.9, 0.1) and (0.51, 0.49). In this case, the probability that the first data is judged as the first type is 0.9, and the probability that the second data is judged as the first type is 0.51, so the second data is more "difficult" to be distinguished, so more There is a value that continues to be marked. The so-called Least Confident method is to select the samples with the largest probability and the smallest for labeling. The mathematical formula is:

$$x^*_{LC} = \arg \max_x (1 - P_\theta(\hat{y} | x)) = \arg \min_x P_\theta(y | x)$$ (1)
Least Confidence believes that the sample with a low probability value of the model predicting the sample is an "error-prone" sample, where is the prediction category, and $x$ is the most uncertain sample of the model. But this method does not consider samples with a relatively small probability.

**Fig. 3 Least confidence select data code**

### 3.2.2 Entropy Sampling

Entropy sampling is to select the most uncertain sample according to the entropy of the class probability of each sample through the base learner. The larger the entropy, the smaller the uncertainty is.

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i)$$  \hspace{2cm} (2)

For two-class model classification tasks, after training a base learner, the base learner performs model update iterations on uncertain samples in the remaining sample pool. If there are 2 unstandardized samples in the pool, after being input to the learner, the probability that the 2 samples belong to 2 categories will be output.

**Fig. 4 Entropy sampling select data code**

### 3.2.3 Randomly select data:

**Fig. 5 Random sampling code**
3.3 Discussion of learning curves

![Accuracy curve on the MNIST data set](image)

Fig.7 Accuracy curve on the MNIST data set
It can be seen from figure 3 that on the MNIST data set, compared to random sampling, Least Confidence predicts based on the maximum probability value output by each sample. When encountering the most uncertain sample of the model, manually adjust the sample standard; reduce the model's misidentification of such samples. When Entropy Sampling encounters an uncertain sample, it will determine the category of the sample according to the probability entropy of the category to which the sample belongs. Therefore, the recognition rates of Least Confidence and Entropy Sampling are higher than those of random sampling. The highest recognition rate can reach about 98.5%, while the highest recognition rate of random sampling is about 97.3%.

![Fig. 8 Correction rate curve in MNIST after improved Least Confidence and Entropy Sampling](image)

Dropout means that during the training of the deep learning network, neural network units are temporarily dropped from the network according to a certain probability. Add dropout training to the Entropy Sampling and Least Confidence methods, and figure 4 shows that the accuracy curve after adding dropout is smoother, and compared with the method without dropout, the accuracy rate is still maintained at a relatively high level, without dropout the previous accuracy curve fluctuates too much, indicating that the ability to fit the data is not very good.
Random Sampling: Select a batch of $k$ samples from the unlabeled set at random.
Uncertainty-based Sampling (including Entropy Sampling and Least Confidence method): For each unlabeled sample, compute the classification entropy as $e = - \sum p_i \log(p_i)$, where $i$ runs from 1 to the number of classes and $p_i$ is the probability that the sample belongs to class $i$. Select the $k$ samples producing the highest entropy. Therefore, setting the same training conditions on the MNIST data set, the training effects obtained by Entropy Sampling and Least Confidence are significantly better than Random Sampling. After adding dropout, the model can better fit the data and prevent over-fitting as the epoch increases.

4. Conclusion
Through this research, it is understood that the essence of active learning is to maximize the utilization of training samples. Active learning methods try to solve the difficulties of labeling a large number of training data set. The most valuable unlabeled samples are actively selected for labeling, with as few as possible and the annotated samples achieve the expected performance of the model. The different methods used in this experiment to select data samples, one is the least confidence method, and the other is entropy sampling. By comparing with the random sampling results, the recognition accuracy after training with these two methods is higher. Subsequently, the drop-out technique was added to avoid poor performance of the training set in response to new situations and overfitting. In general, active learning can effectively discover high-information samples in the training data set and train the model.

Fig. 9 Comparison of accuracy curves of different methods

![Comparison of accuracy curves of different methods](image_url)
efficiently. Compared with traditional supervision methods, active learning can handle larger training data sets well, select sample points with distinguishing ability from them, reduce the amount of training data, and reduce the cost of manual labeling.

References
[1] Somaya Al-Ma’adeed, Dave Elliman, and Colin Higgins(2003) A Data Base for Arabic Handwritten Text Recognition Research. The International Arab Journal of Information Technology, vol. 1, no. 1, pp.117-121,2004
[2] J.J. Hull, "A Database for Handwritten Text Recognition Research", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 5, pp. 550-554, May 1994.
[3] Le Cun, Y., Jackel, L., Boser, B., Denker, J., Graf, H., Guyon, I., Henderson, D., Howard, R., Hubbard, W.: Handwritten digit recognition: Applications of neural network chips and automatic learning. IEEE Communications Magazine 27(11), 41–46 (1989)
[4] Almuallim H. and Yamaguchi S., "A Method of Recognition of Arabic Cursive Handwriting,” IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI-9), pp. 715-722, 1987.
[5] Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang and Hong-Jiang Zhang,” Two-Dimensional Active Learning for image classification,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1–8.
[6] Tong S , Koller D . Support vector machine active learning with applications to text classification[C]// JMLR.org, 2002:999-1006.
[7] Krogh A , Vedelsby J . Neural Network Ensembles, Cross Validation, and Active Learning[C]// International Conference on Neural Information Processing Systems. MIT Press, 1995.
[8] Li X , Guo Y . Active learning with multi-label SVM classification[C]// International Joint Conference on Artificial Intelligence. AAAI Press, 2013.
[9] L.M. Tiwari, S. Agrawal, S. Kapoor and A. Chauhan, "Entropy as a measure of uncertainty in queueing system,” 2011 National Postgraduate Conference,Kuala Lumpur, 2011, pp. 1–4.
[10] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.
[11] MLAHara, Kazuyuki, D. Saitoh, and H. Shouno. "Analysis of Dropout Learning Regarded as Ensemble Learning." International Conference on Artificial Neural Networks Springer, Cham, 2016:72-79.