Research on Disambiguation Attributes Based on Semantic Hierarchical Model

Linlin Yu, Zhenxia Wang, Li Su, Dan Wei
Dalian University of Science and Technology

Abstract: Feature extraction concepts and statistical classifiers began to be used in computer-aided detection and diagnosis, and many different feature learning techniques such as principal component analysis, image block clustering and dictionary methods are very popular. However, in the case of less labeled data, the application and effect of deep learning are restricted. The research content of this paper is based on items such as nodule multi-semantic and fine-grained classification and rating research of WGAN synthetic oversampling learning, multi-semantic assisted disambiguation research based on multi-level association transfer learning, and multi-label incremental learning research for small sample semantic hierarchical structure. Moreover, based on the synthesis of various semantic samples, multi-semantic association transfer, and multi-label incremental learning methods, a set of effective methods and application mechanisms for intelligent auxiliary semantic hierarchical structure disambiguation are formed by describing the problem definition, proposing research methods, and discussing the experimental effects of the proposed methods in detail.

1. Introduction
In recent years, the advantage convergence of labeled big data, deep learning methods, cloud storage technology, and significantly improved computing power has promoted the rapid development of artificial intelligence (AI) applications in various industries. Meanwhile, as deep convolutional neural networks have achieved a series of breakthrough research results in natural image fields such as target detection, image classification, semantic segmentation, and face recognition verification, the cross-fusion and in-depth application of artificial intelligence in the more difficult and more demanding medical imaging field has also aroused great interest and widespread attention of many researchers. The continuous growth of massive image data, the continuous iterative update of intelligent diagnosis model algorithms, the substantial increase in computing power, and the strong support of national policies have ushered in unprecedented new opportunities in the field of artificial intelligence medical imaging.

The cost function of the GAN discriminant model is a true or false binary loss function, and the cost function of the WGAN discriminant model is represented by the Wasserstein distance between the real sample and the synthetic sample data distribution. The Wasserstein distance is also often referred to as EarthMover (EM) Distance. Therefore, the training of WGAN discriminant model verifies the optimization effect of regression classification task. The EM distance is used to measure the distribution between the real sample and the synthetic sample to avoid the mode collapse problem caused by Kullback-Leibler divergence (KL) and the Jensen-Shannon divergence (JS) problem caused by the discontinuity of the loss function. Finally, it is proved that the EM distance can provide a reliable and practical gradient for the loss function, which makes the training process easier and the quality of the synthesized samples better.
2. Synthetic Oversampling Method Based on WGAN

2.1 Method Description and Definition

A typical GAN consists of a generative network model and a discriminative network model. The main function of the generative network model is to synthesize samples by estimating the potential distribution of the target domain, and the purpose of the discriminative network model is to distinguish these synthetic samples from real samples.

The optimization training process of GAN needs to promote the generation of network models to synthesize some samples that can deceive the discriminant model, and it also needs to continuously improve the discriminant model's ability to judge whether a given sample is true or false. Therefore, the effective training of GAN needs to optimize the two network models of generative model and discriminant model. The training optimization process of a typical GAN is shown in Figure 1.

![Figure 1. Typical GAN training framework process](image)

Since the iterative training of the two model networks GAN's generative model and discriminant model is needed to be alternated, the typical GAN optimization process is more difficult, and there are some shortcomings. The training of a typical GAN is quite unstable, which largely relies on the competition between the generative model and the discriminant model within the framework of the binary minimax game.

Wasserstein GAN (WGAN) is an improved GAN, which can not only greatly reduce the training difficulty of typical GANs, but also avoid potential mode collapse problems. The training optimization process of WGAN is shown in Figure 2.

![Figure 2. The training framework process of WGAN](image)
Compared with Figure 1, the objective cost function of the discriminant model of WGAN and GAN is different. The distance between the real sample distribution PV and the synthetic sample distribution Ps is defined as:

\[ W(P_x, P_y) = \text{sel}[f_{rh}, P_c], E_{(x,y)} \]  

(1)

Among them, x and y represent real samples and synthetic samples respectively, and \( \text{sel}[f_{rh}, P_c] \) represents the set of all possible joint distributions that \( f_{rh} \) and \( P_c \) combine \( e_r \). For each possible joint distribution \( E_{(x,y)} \), a real sample x and a synthetic sample y can be sampled from it. That is, there is \( (x,y) \rightarrow \phi \) and the EM distance of the \((x,y)\) sample is calculated, so the expected value of the distance between the sample pairs under the joint distribution \( d(\text{Or}, y) \) can be calculated \( E(\text{Or}, y) \). Moreover, in all possible joint distributions, the expected value can be taken to the infimum.

However, it is very difficult to directly obtain the infimum value in Equation 1, and it can be transformed into a solution formula:

\[ \max R_{yy} \]

\[ p_y^y - p(x) - E_{y-x} \]

\[ [D(y)]_{x \rightarrow y} \]

(2)

Among them, D represents the neural network of the discriminant model. By optimizing the maximum value obtained by D, the EM distance can be approximated. What is more, the goal of the generative model network G is to make \( P(x) \) continuously approach the real sample distribution IPV and synthesize samples that obey the 3 distribution. In the training process, the generative model network will map a random vector to obey the normal distribution \( y \). If the sample synthesis of the generative model is considered in Equation 2, Equation 2 can be written as the formula:

\[ \min_{D} \max_{G} \max_{x \rightarrow y} [D(x)] = E_{x \rightarrow p(c(z))} \]

(3)

The maximum value in Equation 3 also represents the EM distance between \( P_x \) and \( P_y \). \( \max_{x \rightarrow y} \) is used to adjust and update the parameters in the neural network of the generated model. Therefore, the entire training process of WGAN can be expressed as:

\[ \min_{D} \max_{G} [D(x)] - B_{x \rightarrow 0} = \sum_{i=1}^{n} P_{y}^{y} [D(G(g(z)))] \]

(4)

In Equation 4, WGAN training can be expressed as a binary minimax game between D and G, which can be decomposed into iterative optimization of the discriminator through Equation 5:

\[ \max_{x \rightarrow e} [D(x)] - E_{x \rightarrow 0} (P) = \sum_{i=1}^{n} [D(G(G(z)))] \]

(5)

2.2 Frame Design and Parameter Setting

Both WGAN discriminant model and generative model network are designed and implemented with the help of the convolutional neural networks. The frame structure of the discriminant model and the generative model is shown in Figure 3.
The input layer size of the discriminant model is 64x64, including real samples and synthetic samples. The next series of convolutional layers are processed by Batch Normalization (BN) and Leaky Rectified Linear Unit (LReLU) to improve the training stability of the network and reduce the difficulty of training, which can greatly improve the speed of network training. In addition, in order to avoid unstable synthetic samples, batch normalization is not recommended for the input layer of the discriminant model and the output layer of the generative model.

The output size of the generated model is consistent with the input size of the discriminant model. The input of the generative model is a 100-dimensional random noise vector initialized with a normal distribution. On the basis of reconstructing the random noise vector into 100×1×1, the transposed convolution filtering is used to obtain the 4×4 feature map of 512 channels. Here, the transposed convolution operation is denoted as Convolution Transpose (CONV T). The next four layers of the generative model are also composed of CONVT. Similarly, batch normalization is also used to stabilize the training process of the generative model, and ReLU is applied to generate the model. In the four-layer CONVT, after each layer is processed, the dimension of the feature map is doubled, and the number of channels is halved. Finally, the output size of the generated model is reached 64×64).

Besides, the two convolutional neural networks of WGAN are trained and optimized by RMSProp algorithm. The initialization of WGAN's generative model and discriminant model adopts normal distribution N(0,0.02), the slope of LReLU in discriminant model is set to 0.2, and the learning rate of discriminant model and generative model is set to 5e-5.

3. Experiment and Analysis
Based on the ontology data, one’s own annotations can be modified by checking and comparing the annotations of the other three data in the first stage, so that the final annotation results can be given. Moreover, each semantic attribute is used to describe the nodule’s semantic feature performance, which is essential for the comprehensive evaluation of the ontology. Additionally, each semantic attribute has 5-6 ratings to distinguish the degree of expression of semantic features. The image data in the LIDC-IDRI dataset are all stored in the DICOM standard format, which adopts windowing technology [-600HU, 1600HU] to extract and normalize image values. The window level is -600HU, and the
window width is 1600HU. Then, according to the nodule size report and the nodule coordinate position, a Region of Interest (ROI) of 64x64 size will be cut out.

In the horizontal section of the image sequence, the diameter of the nodules in the LIDC-IDRI dataset does not exceed 64x64 pixels. Therefore, the ROI size setting can well cover all the nodules. What is more, by parsing the XML markup file, a total of 2632 nodules are extracted from the LIDC-IDRI data set, and 7 attribute ratings given by each nodule are obtained as well.

3.1 Experimental Program
The 7 different semantic attributes of the nodules in the LIDC data set are classified in fine-grained, and the five technical schemes of ORI, AUG, GAN, DCGAN and WGAN are compared experimentally. The first scheme, ORI, is the original scheme without oversampling and adding training data. The second plan, AUG (Augmentation), implements a standard data augmentation plan, which uses a data augmentation method to perform random rotation of the image from 0 to 356 degrees, and then flip it horizontally or vertically. The other three schemes use typical GAN, DCGAN and WGAN to synthesize sample data and oversampling. All five schemes use the same CNN model architecture and parameters shown in Table 1 for semantic fine-grained classification.

Table 1 Architecture parameter settings of semantic fine-grained classification model

| Layer | Feature maps | Kernel size | Stride |
|-------|--------------|-------------|--------|
| Input | 64×64×32     | -           | -      |
| C₁    | 32×32×16     | 4 1         |        |
| M₁    | 24×24×16     | 6 2         |        |
| C₂    | 16×16×64     | 6 1         |        |
| M₂    | 16×16×32     | 8 2         |        |
| FC₁   | 500          | -           | -      |
| FC₂   | j            | -           | -      |

3.2 Experimental Data Analysis
The experiment uses the five-fold cross-validation method to perform sample synthesis oversampling processing on the minority classes of the four schemes of AUG, GAN, DCGAN and WGAN. For the training data set of each fold in the five-fold crossover, an oversampling operation needs to be performed, but the verification data must remain unchanged. Meanwhile, the oversampling process is only for the training data, not the verification data, to reflect the impartiality of the experimental test. For a fair comparison of experiments, the samples of the five-fold cross data division of ORI, AUG, GAN, DCGAN and WGAN schemes are the same, and the number of synthesized samples of each subcategory in each semantic of each fold in the AUG, GAN, DCGAN, and WGAN schemes is also the same. The test results are shown in Figures 4 and 5.
The number of synthetic samples of the minority class is determined by the difference between the minority class and the majority class. After synthetic oversampling, the number of samples in each minority class is equal to the number of samples in the majority class. When the number of training
iterations of WGAN is set to 2000, it is enough to obtain better quality synthetic samples, and the training process is stable and the convergence effect is also very good. What is more, the training process of GAN and DCGAN is very unstable, while the sample quality is not as good as the samples synthesized by WGAN.

4. Conclusion
A priori method of feature regularization, namely the multi-label square gradient level (MLSGM) feature regularization method is proposed in the paper, which uses small samples and multiple labels for incremental learning, quantifies the association relationship between different semantic tasks through transfer learning, and combines these semantic tasks for multi-level association transfer learning to construct a multi-level semantic task migration diagram under different supervision budgets, so that different target tasks can find the best migration source, and a set of target tasks can obtain the maximum performance gain to quantitatively analyze the relationship between the amount of supervised learning of labeled data and the performance of the target task. In addition, the effectiveness of the method is verified through experiments.

Acknowledgement
This paper is supported by the project of Research on Key Technologies of media data depth analysis system under the background of big data, which number is 2019-ZD-0349, and the project of Research on Key Technologies of indoor positioning based on multi-source data fusion, which number is 2019-ZD-0348.

References
[1] Wang S,Huang M,Deng Z. Densely connected CNN with multi-scale feature attention for text classification. Proceedings of International Joint Conferences on Artificial Intelligence . 2018
[2] Improved semantic representations from treestructured long short-term memory networks. Tai K S,Socher R,Manning C D. . 2015
[3] Improving Language Understanding with Unsupervised Learning. Radford A,Narasimhan K,Salimans T, et al. . 2018
[4] BERT:Pre-training of deep bidirectional transformers for language understanding. Devlin J,Chang M W, Lee K, et al. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:Human Language Technologies . 2019
[5] Heterogeneous Multimedia Cooperative Annotation based on Multimodal Correlation Learning[J] . Feng Tian,Quge Wang,Xin Li,Ning Sun. Journal of Visual Communication and Image Represe . 2018
[6] Binary Dragonfly Optimization for Feature Selection using Time-Varying Transfer functions[J] . Majdi Mafarja,Ilbrahim Aljarah,Ali Asghar Heidari,Hossam Faris,Philippe Fournier-Viger,Xiaodong Li,Seyedali Mirjalili. Knowledge-Based Systems . 2018
[7] SVM based multi-label learning with missing labels for image annotation[J] . Yang Liu,Kaiwen Wen,Quanxue Gao,Xinbo Gao,Feiping Nie. Pattern Recognition . 2018
[8] Feature Selection[J] . Jundong Li,Kewei Cheng,Suhang Wang,Fred Morstatter,Robert P. Trevino,Jiliang Tang,Huan Liu. ACM Computing Surveys (CSUR) . 2017 (6)
[9] Socializing the Semantic Gap[J] . Xirong Li,Tiberio Uricchio,Lamberto Ballan,Marco Bertini,Cees G. M. Snoek,Alberto Del Bimbo. ACM Computing Surveys (CSUR) . 2016 (1)
[10] Attentionxml:Extreme multi-label text classification with multi-label attention based recurrent neural networks. You R,Dai S,Zhang Z,Mamitsuka H,Zhu S. . 2018