PATTERN CLASSIFICATION TOOLS FOR ROBUST LOSS FUNCTIONS

SRIDHAR REDDY SURUKANTI
Research Scholar
Dept of Computer Science & Engg
Arni University-Himachal Pradesh.

DR. PRASADU PEDDI
Research Supervisor
Dept of Computer Science & Engg
Arni University-Himachal Pradesh.

DR. BHASKAR MARAPELLI
Co-supervisor
Associate Professor
Dept of Computer Science & Engg
KL Deemed to be University-A.P

ABSTRACT

Image denoising using hierarchical structure facilitates extracting multiple features from noisy images when training a deep convolutional neural network (CNN). However, to further improve the denoising performance, there still has a tricky path to be treaded among increasing receptive fields, reducing information loss, raising feature utilization, and keeping computational complexity down. Besides, the choice of the loss function is mostly empirical without a convincing explanation. Our study proposes a novel method based on deep CNN for image denoising, which combined an adaptive hierarchical concatenated network with a robust loss function (ACRNet). The proposed method uses the hierarchical structure as backbone network architecture. Each level consists of residual learning and densely connective convolution layers with Mish activation function responsible for feature utilization and training process stabilization. Then, we integrate an adaptive sampling algorithm into the network to achieve a better trade-off among the large receptive fields, low computational complexity, and few information losses.

Keywords: *hierarchical structure, convolutional neural network (CNN), robust loss function, network architecture.*

INTRODUCTION

Classifying images in general, including document images, is one of the most popular tasks in computer vision. An image is classified based on special features included and not based on its structure. Following this definition, classification is indeed a sequential process that starts from preprocessing data, features extraction, features fusion, and finally assigns the input to one of the specified classes. Our model's input is the image that contains one or more document images, and the output of our model is the label of the document(s) included in the input image. Some examples of document labels are Email, handwriting, a report document, bank card, ID, etc. Various factors or worse artifacts in the input images may significantly reduce the classification confidence. Some related examples: artifacts in the picture such as rotation, blur, shadow, or spotlight. Nowadays, visual sensors can perform more complex computer vision tasks as they now have more processing power. Therefore, one can see more use-causes related to object classification (document classification being just a subset). Some well-known real-world examples of such usages are image recognition and detection, emotion-sensing from face images, e.g., in the context of driver status monitoring, machine vision involving search and rescue missions using drones, performing video-based traffic control and surveillance, etc.

The management of hardcopy documents within an enterprise (a company) traditionally requires too much time and knowledge for the secretarial and archival staff to fully understand the various documents' contents and classify them appropriately. This last-described task is not simple and generally consumes much time and is consequently proportionally expensive. Thus, an automated electronic classifier of (pre-scanned) document images (see "the digital office" concept) can reliably help to save both time and money for many organizations and companies. Jointly with the increasing popularity of mobile phones, the almost pervasive use of integrated cameras for capturing document images is also increasing. The captured images depend on environmental conditions such as photograph experience, light conditions, etc. can have different qualities. However, the presence of various artifacts/distortions such as noise, shadows, and blur can significantly decrease the performance of a classifier and eventually make it practically unusable (i.e., very low performing) in real-life conditions. Therefore, a robust (document) classification model for such hard/harsh scenarios is needed.

LITERATURE REVIEW

Beom-Soo Kang (2021) Unmanned aerial vehicles (UAVs) are being widely utilized for various missions: in both civilian and military sectors. Many of these missions demand UAVs to acquire artificial intelligence about the environments they are navigating in. This perception can be realized by training a computing machine to classify objects in the environment. One of the well known machine training approaches is supervised deep learning, which enables a machine to classify objects. However, supervised deep learning comes with huge sacrifice in terms of time and computational resources.

Vince Calhoun (2021) Recent critical commentaries unfavorably compare deep learning (DL) with standard machine learning (SML) approaches for brain imaging data analysis. However, their conclusions are often based on pre-engineered features depriving DL of its main advantage — representation learning. We conduct a large-scale systematic comparison profiled in multiple classification and regression tasks on structural MRI images and show the importance of representation learning for DL.

Kyandoghere Kyamakya (2021) This study core objective is to develop and validate a new neurocomputing model to classify document images in particularly demanding hard conditions such as image distortions, image size variance and scale, a huge number of classes, etc. Document classification is a special machine vision task in which document images are categorized according to their likelihood. Document classification is by itself an important topic for the digital office and it has several usages.

Chun-Hou Zheng (2020) As a machine learning method with high performance and excellent generalization ability, extreme learning machine (ELM) is gaining popularity in various studies. Various ELM-based methods for different fields have been proposed. However, the robustness to noise and outliers is always the main problem affecting the performance of ELM. In this study, an integrated method named correntropy induced loss based sparse robust graph regularized extreme learning machine (CSRGELM) is proposed.

Mussarat Yasmin (2020) Documents are stored in digital form across several organizations. Printing this amount of data and placing it into folders instead of digitally is against the practical, economical, and ecological perspective. An efficient way of retrieving data from digitally stored documents is also required.

LOSS FUNCTION

In mathematical optimization and decision theory, a loss function or cost function (sometimes also called an error function) is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event. An optimization problem seeks to minimize a loss function. An objective function is either a loss function or its opposite (in specific domains, variously called a reward function, a profit function, a utility function, a fitness function, etc.), in which case it is to be maximized. The loss function could include terms from several levels of the hierarchy.

In statistics, typically a loss function is used for parameter estimation, and the event in question is some function of the difference between estimated and true values for an instance of data. The concept, as old as Laplace, was reintroduced in statistics by Abraham Wald in the middle of the 20th century. In the context of economics, for example, this is usually economic cost or regret. In classification, it is the penalty for an incorrect classification of an example. In actuarial science, it is used in an insurance context to model benefits paid over premiums, particularly since the works of Harald Cramér in the 1920s. In optimal control, the loss is the penalty for failing to achieve a desired value. In financial risk management, the function is mapped to a monetary loss.

Economic Choice Under Uncertainty

In economics, decision-making under uncertainty is often modelled using the von Neumann–Morgenstern utility function of the uncertain variable of interest, such as end-of-period wealth. Since the value of this variable is uncertain, so is the value of the utility function; it is the expected value of utility that is maximized.

Selecting a Loss Function

Sound statistical practice requires selecting an estimator consistent with the actual acceptable variation experienced in the context of a particular applied problem. Thus, in the applied use of loss functions, selecting which statistical method to use to model an applied problem depends on knowing the losses that will be experienced from being wrong under the problem's particular circumstances. A common example involves estimating "location". Under typical statistical assumptions, the mean or average is the statistic for estimating location that minimizes the expected loss experienced under the squared-error loss function, while the median is the estimator that minimizes expected loss experienced under the absolute-difference loss function. Still different estimators would be optimal under other, less common circumstances.

In economics, when an agent is risk neutral, the objective function is simply expressed as the expected value of a monetary quantity, such as profit, income, or end-of-period wealth. For risk-averse or risk-loving agents, loss is measured as the negative of a utility function, and the objective function to be optimized is the expected value of utility.

METHODOLOGY

Problem Formulation: Define the pattern classification problem and the desired classes. Ensure you have a labeled dataset with input features and corresponding class labels. Select a Classifier: Choose a suitable classifier algorithm for your specific problem, such as k-nearest neighbors, support vector machines, decision trees, random forests, or deep learning models like neural networks. Loss Function Selection: Instead of using traditional loss functions like mean squared error (MSE) or cross-entropy, opt for robust loss functions that can handle outliers and noisy data more effectively. Some popular robust loss functions include: Huber Loss: A combination of MSE for small errors and absolute loss for larger errors. It is less sensitive to outliers than MSE.

Log-Cosh Loss: An approximation of Huber loss that is easier to compute and still provides robustness to outliers. Quantile Loss: A loss function used for quantile regression, which is also robust to outliers. Tukey's Biweight Loss: Another robust loss function that is similar to Huber loss but assigns less weight to outliers.

RESULTS

We use the C-loss function for training single hidden layer perceptrons and RBF networks using back propagation. Our evaluations are divided into two parts.

Best results over Monte Carlo trials and the values in Tables 1–3 have shown the average generalization performance over 100 Monte Carlo runs with random initialization of network weights and random sampling of the training data. The averaged result reflects the likelihood outcome of a naive user using the proposed method. Out of these, we select the configuration that gives the best classification performance after testing on test dataset. This entire process is performed 100 times and an average of these best results is presented.

Table 1: Generalization performance (in percent) on the DIABETES dataset, at the end of training using the C-loss and the square loss functions, for different numbers of PEs and training epochs

Number of PEs	**10**			**15**			**20**		
Total training epochs, N	25	50	75	25	50	75	25	50	75
C-loss(σ=0.5)performance	75.66	75.23	74.86	75.56	74.95	74.82	**75:70**	75.08	74.14
C-loss (σ=1) performance	**75:04**	74.06	73.29	74.88	73.37	73.00	74.89	73.34	72.14
Square-loss performance	**74:04**	72.53	72.00	73.54	71.72	71.92	72.83	71.69	71.19

Table 3: Generalization performance (in percent) on the breast cancer dataset, at the end of training using the C-loss and the square loss functions, for different numbers of PEs and total training epochs

Numb	**1**			**1**			**2**		

er of PEs	**0**			**5**			**0**		
Total training epochs, N	25	50	75	25	50	75	25	50	75
C-loss(σ=0.5)performance C-loss (σ=1) performance Square-loss performance	**97:0796:8996:68**	96.7096.3296.19	96.5696.2496.07	96.7596.596.32	96.7596.4196.16	96.6296.2196.04	96.8996.6396.49	96.5196.2796.17	96.3196.5895.81

Table 4: Generalization performance (in percent) on the connectionist bench dataset, at the end training of using the C-loss and the square loss functions, for different numbers of PEs and total training epochs

Number of PEs	**10**			**15**			**20**		
Total training epochs, N	25	50	75	25	50	75	25	50	75
C-loss(σ=0.5)performance C-loss (σ=1) performance Square-loss performance	**89:1788:9388:02**	88.988.0587.09	88.4387.5087.07	88.9888.8387.95	88.8187.9787.64	88.9488.2487.64	88.7488.4587.72	88.9387.7887.66	88.0787.1786.71

Table 5: Variance of generalization performance of the C-loss function (σ=0.5) and (σ=1) and square loss function, across the 9 different combinations of training parameters

Dataset	**Diabetes**	**Breast Cancer**	**Conn. Bench**
C-loss(σ=0.5) variance C-loss (σ=1) variance Square-loss variance	**0:2477** 1.0010 0.8762	**0:0488** 0.0498 0.0655	**0:1227** 0.1807 0.1936

the generalization performance nonetheless tends to suffer on excessive training. The use of the C-loss function improves the generalization. Note that the performance of the C-loss in this experiment was found to be roughly the same for both (σ=1) and (σ=0.5). We do not show the performance variation across system parameters like number of RBF centers, but we have observed similar trends.

CONCLUSION

We have presented a loss function for classification that is inspired by a robust similarity measure, Correntropy. Correntropy induces a loss function that is smooth and non-convex and effectively approximates the L_0 loss for samples with high errors, and the L_2 loss for small errors. We have tested the proposed loss function on some noisy, real world datasets. Instead of simply showing the 'best classification results' of the proposed loss function on a larger number of datasets, we chose to study deeper the behavior of the loss function as the system parameters are varied. Many classification algorithms often underperform due to improper choice of such parameters, or require additional techniques like cross validation to scan for the best set of parameters. Using the proposed loss function, this problem is alleviated. A potential and important class of applications is for big data, where the online nature of gradient descent can be used by stochastically sampling the data, i.e. without requiring the availability of all the data in memory, nor the use of huge matrices as in SVMs.

Reference

1. *Beom-Soo Kang (2021),"Robust Approach to Supervised Deep Neural Network Training for Real-Time Object Classification in Cluttered Indoor Environment",Appl. Sci.,ISSNno:2076-3417,Vol.11(15),Pages.7148.https://doi.org/10.3390/app11157148*
2. *Chun-Hou Zheng (2020),"Correntropy induced loss based sparse robust graph regularized extreme learning machine for cancer classification",BMC Bio-informatics,ISSNno:1471-2105,Vol.21,Pages.445.doi:10.1186/s12859-020-03790-1*
3. *Huangang Wang (2017),"Ramp Loss based robust one-class SVM",Pattern Recognition Letters,ISSNno:0167-8655,Vol.85,Pages.15-20.https://doi.org/10. 1016/j.patrec.2016.11.016*
4. *Kai Zhang (2015),"Outlier-robust extreme learning machine for regression problems",Neurocomputing,ISSNno:1872-8286,Vol.151,Pages.1519-1527.https://doi.org/10.1016/j.neucom.2014.09.022*
5. *Kyandoghere Kyamakya (2021),"A Deep-Learning Based Visual Sensing Concept for a Robust Classification of Document Images under Real-World Hard Conditions",Sensors (Basel),ISSNno:1424-8220,Vol.21(20),Pages.6763. doi:10.3390/s21206763*
6. *Mussarat Yasmin (2020),"Pearson Correlation-Based Feature Selection for Document Classification Using Balanced Training",Sensors,ISSNno:1558-1748,Vol.20(23),Pages.6793.DOI:10.3390/s20236793*
7. *Ping Zhong (2020),"Robust regularized extreme learning machine for regression with non-convex loss function via DC program",Journal of the Franklin Institute,ISSNno:0016-0032,Vol.357,Issue.11,Pages.7069-7091.https ://doi.org/10.1016/j.jfranklin.2020.05.027*
8. *Sadaf Manzoor (2022),"Robust Extreme Learning Machine Using New Activation and Loss Functions Based on M-Estimation for Regression and Classification",Scientific Programming,ISSNno:1875919X,Vol.2022,https://doi.org/10.1155/2022/6446080*
9. *Vince Calhoun (2021),"Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning", Nature Communications,ISSNno:2041-1723,Vol.12,https://doi.org/10.1038/s4 1467-020-20655-6*
10. *Yong Dou (2017),"Robust regularized extreme learning machine for regression using iteratively reweighted least squares",Neurocomputing,ISSN no:1872-8286,Vol.230,Pages.345-358.https://doi.org/10.1016/j.neucom.2016. 12.029*