ConvNet and Dempster-Shafer Theory for Object Recognition

Zheng Tong, Philippe Xu, Thierry Denœux

Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 7253)

https://orcid.org/0000-0001-6894-3521

Outline

Dempster-Shafer theory

Dempster-Shafer (DS) theory, also referred to as *Evidence Theory*, is

1. Representing independent pieces of evidence
   (a) Let X be a variable taking one and only one value in a finite set, called the frame of discernment
   (b) Evidence (uncertain information) about X can be represented by a mass function $m : 2^{\Omega} \to [0,1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1$$

   (c) Every subset A of such that $m(A) > 0$ is a focal set of m
   (d) if $m(\emptyset) = 0$, m is normalized
2. Aggregating mass functions using Dempster's rule

Dempster-Shafer theory

1. Representing independent pieces of evidence
2. Aggregating mass functions using Dempster's rule
   (a) Let m_1 and m_2 be two mass functions and

   $$\kappa = \sum_{B \cap C = \phi} m_1(B) m_2(C)$$

   their degree of conflict

   (b) If $\kappa < 1$, then m_1 and m_2 can be combined as

   $$(m_1 \oplus m_2)(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) m_2(C), \quad \forall A \neq \phi$$

   and $(m_1 \oplus m_2)(\emptyset) = 0$

Dempster-Shafer theory

Three main directions of DS theory for pattern recognition

1. Classifier fusion: classifier outputs are expressed as mass functions and combined by Dempster's rule (e.g., Liu et al., 2018)
2. Evidential calibration: the decisions of statistical classifiers are converted into mass functions (e.g., Xu et al., 2016)
3. Evidential classifier: the elements of each feature vector is considered as independent pieces of evidence and converted into mass functions. The mass functions are combined by Dempster's rule (e.g., Denœux, 2010)

Evidential classifier

1. Evidential classifiers can provide more informative outputs for
   - (a) Exploit for uncertainty quantification
   - (b) Make a decision allowing for ambiguous rejection
2. The performance of evidential classifiers depends on the training data set and reliability of object representation
3. Deep learning, especially convolutional neural network (ConvNet), has achieved remarkable success on object representation
   - (a) Robustness: strong tolerance to translation and distortion
   - (b) Automation: a data-driven method with no human assistance

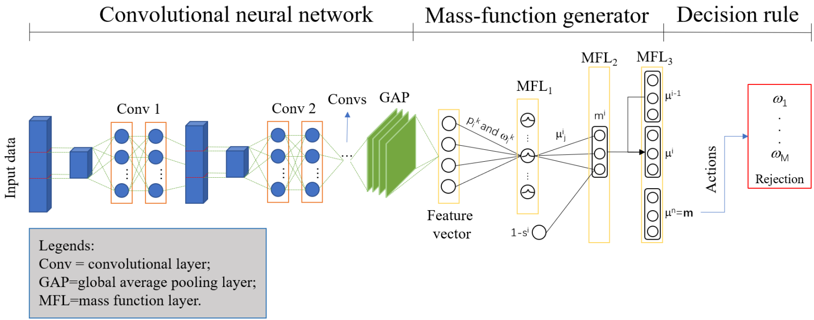
Objective

1. Build a novel classifier based on evidential classifier and ConvNet
2. Make a decision allowing for ambiguous rejection. Ambiguous rejection means the novel classifier cannot assign a pattern to one of the class membership because of conflict evidences from the input feature vector

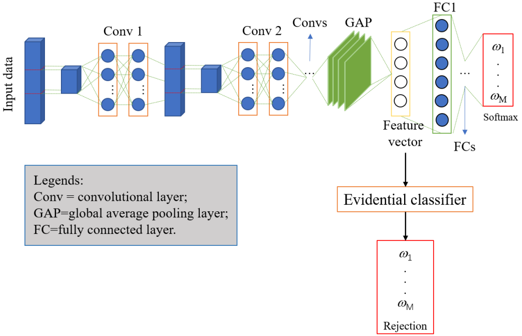
Outline

Outline

Architecture of a ConvNet-BF classifier



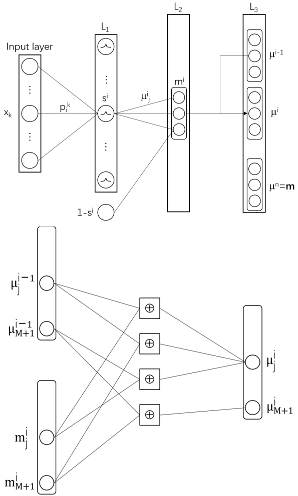
Object representation



- ConvNet: import into FCs and a softmax layer for classification (e.g., LeCun et al., 2015)
- ConvNet-BF classifier: consider elements of the feature vector as independent pieces of evidence for generating mass functions

Y. LeCun,Y. Bengio, and G. Hinton. Deep learning. Nature 521.7553 (2015): 436-444.

Generation of mass functions



1. Compute the distance between the feature vector and each prototype $\boldsymbol{p}^i$
2. Convert the activated distance into the mass m^i associated to prototype $\boldsymbol{p}^i$
3. Combine the n mass functions m^i, $i = 1, \ldots, n$ by Dempster's rule
4. Output $\boldsymbol{m} = (m(\omega_1), \ldots, m(\omega_M), m(\Omega))^T$

T. Denœux. A neural network classifier based on Dempster-Shafer theory. IEEE transactions on SMC A, 30(2):131-150, 2000.

Evidence-theoretic rules

1. Rejection
   (a) **Maximum credibility**: $\max_{j=1,\cdots,M} m(\{\omega_j\}) < 1 - \lambda_0$
   (b) **Maximum plausibility**: $\max_{j=1,\cdots,M} m(\{\omega_j\}) + m(\Omega) < 1 - \lambda_0$
   (c) **Maximum pignistic probability**: $\max_{j=1,\cdots,M} m(\{\omega_j\}) + \frac{m(\Omega)}{M} < 1 - \lambda_0$
2. Assignment to class ω_j

$$m(\{\omega_j\}) = \max_{j=1,\cdots,M} m(\{\omega_j\})$$

Outline

End-to-end learning I

1. Compute a normalized error $E_\nu(\boldsymbol{x})$ as

$$E_\nu(\boldsymbol{x}) = \frac{1}{2N}\sum_{i=1}^{I}\sum_{q=1}^{M}(Pre_{\nu,q,i} - Tar_{q,i})^2$$

$$Pre_{\nu,q,i} = m'_{q,i} + \nu m'_{M+1,i}$$

2. Compute the derivatives of $E_\nu(\boldsymbol{x})$ w.r.t the connection parameters between the mass-function generator and CovNet as

$$\frac{\partial E_\nu(\boldsymbol{x})}{\partial p_k^i} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial s^i}\frac{\partial s^i}{\partial p_k^i} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial s^i} \cdot 2(\eta^i)^2 s^i \cdot \sum_{k=1}^{P} w_k^i (x_k - p_k^i)$$

$$\frac{\partial E_\nu(\boldsymbol{x})}{\partial w_k^i} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial s^i}\frac{\partial s^i}{\partial w_k^i} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial s^i} \cdot \left(\eta^i\right)^2 s^i \cdot \left(x_k - p_k^i\right)^2$$

End-to-end learning II

3. Compute the derivatives of $E_\nu(\boldsymbol{x})$ w.r.t the parameters in the last convolutional layer of the ConvNet part as

$$\frac{\partial E_\nu(\boldsymbol{x})}{\partial w^m_{i,j,k}} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial f^m_{i,j,k}} \cdot \frac{\partial f^m_{i,j,k}}{\partial w^m_{i,j,k}} = w^m_{i,j,k} \cdot \frac{\partial E_\nu(\boldsymbol{x})}{\partial f^m_{i,j,k}} \quad k = 1, \cdots, P$$

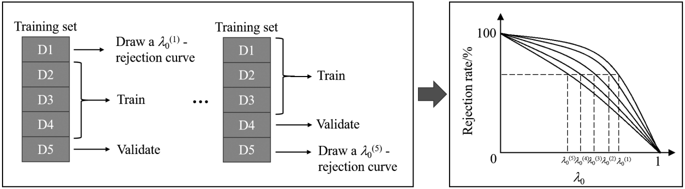
and

$$\frac{\partial E_\nu(\boldsymbol{x})}{\partial b^m_k} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial f^m_{i,j,k}} \cdot \frac{\partial f^m_{i,j,k}}{\partial b^m_k} = \frac{\partial E_\nu(\boldsymbol{x})}{\partial f^m_{i,j,k}} \quad k = 1, \cdots, P$$

4. The derivatives of $E_\nu(\boldsymbol{x})$ w.r.t the parameters in the mass-function generator can be found in the work of Denœux

Determination of λ_0

A data-driven method for a complete learning set

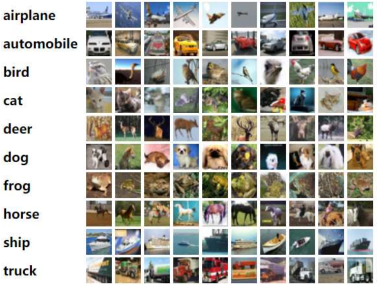


$$\lambda_0 = \sum_{i=1}^{n} \lambda_0^{(i)}$$

Outline

Outline

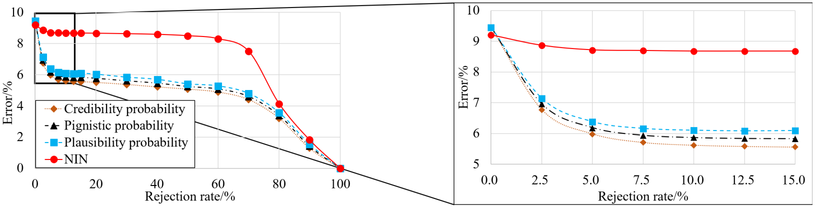
The CIFAR-10 dataset



- Consist 60,000 RGB images of size 32×32 in 10 classes
- There are 50,000 training images, and we randomly selected 10,000 images as validation data for the ConvNet-BF classifier
- There are 10,000 testing images

A. Krizhevsky and G. Hinton, Learning multiple layers of features from tiny images. Tech. report, University of Toronto, 2019.

Test results of the CIFAR-10 dataset



- NIN=network in network, a type of ConvNet
- Curves of credibility, plausibility, and Pignistic probability presents the results of ConvNet-BF classifiers with different decision rules
- Rejection is not considered as an error

Test results of the CIFAR-10 dataset

$n = 10,000$		Labels									
		Airplane	Automobile	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
Actions	Airplane	-	0.03	0.03	0.01	0.02	0.05	0.04	0.01	0.04	0.05
	Automobile	0	-	0.04	0.04	0.08	0.08	0.04	0.06	0.03	0.07
	Bird	0.02	0.04	-	0.05	0.04	0.07	0.03	0.08	0	0.04
	Cat	0.02	0.03	0.13	-	0.06	0.44	0.11	0.04	0.05	0.06
	Deer	0.01	0.04	0.07	0.12	-	0.03	0.12	0.34	0.04	0.08
	Dog	0.02	0.03	0.05	0.49	0.11	-	0.06	0.09	0.01	0.04
	Frog	0.02	0.04	0.08	0.06	0.12	0.06	-	0.06	0.06	0.05
	Horse	0.01	0.02	0.04	0.06	0.31	0.1	0.05	-	0.04	0.04
	Ship	0.04	0.05	0.02	0.04	0.12	0.05	0.04	0.18	-	0.02
	Truck	0.02	0	0.06	0.09	0.03	0.06	0.07	0.06	0.04	-
	Rejection	0.2	0.13	0.14	1.05	0.84	1.07	0.14	1.14	0.18	0.11

* The table reports the errors and rejection rates of a ConvNet-BF classifier in maximum credibility rule

** The total error rate is 5.99%, while the rejection rate is 5%

*** A rejection action is not considered as a error

**** The unit in the table is %

Exploiting in the view of DS theory

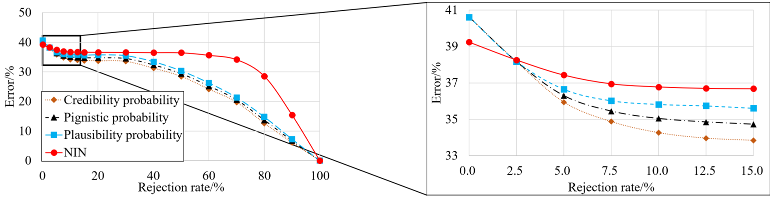
1. Conflicting evidences from the ConvNet part
   (a) Confusing features from convolutional and pooling layers when there are two or more similar patterns
   (b) The maximally conflicting evidence corresponds that $m(\omega_i) = m(\omega_j) = 0.5$
2. The additional $m(\Omega)$ provides the possibility to verify whether a ConvNet-BF classifier is well trained
   (a) $m(\Omega)$ equals 1 when the ConvNet part cannot provide any useful evidence
   (b) $m(\Omega)$ decreases during the training

Outline

The CIFAR-100 and MNIST data set

1. The CIFAR-100 data set is just like the CIFAR-10, except it has 100 classes containing 600 images each (e.g., Krizhevsky and Hinton, 2009)
2. The MNIST data set of handwritten digits consists of a training set of 60,000 examples and a test set of 10,000 examples (e.g., Li, 2012)

Test results of the CIFAR-100 data set



- NIN=network in network, a type of ConvNet
- Curves of credibility, plausibility, and Pignistic probability presents the results of ConvNet-BF classifiers with different decision rules
- Rejection is not considered as an error

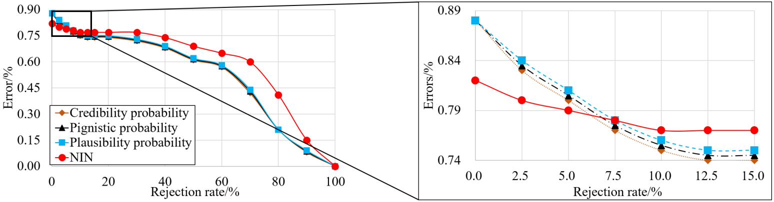
Test results of the CIFAR-100 data set

Table: Confusion matrix for the Cifar100 data set (unit:%)

n=500		Labels				
		Orchids	Poppies	Roses	Sunflowers	Tulips
Actions	Orchids	-	0.24	0.23	0.28	0.15
	Poppies	0.14	-	0.43	0.10	0.90
	Roses	0.27	0.12	-	0.16	0.13
	Sunflowers	0.18	0.15	0.12	-	0.22
	Tulips	0.08	1.07	0.76	0.17	-
	Rejection	0.09	0.37	0.63	0.12	0.34

* The table reports the errors and rejection rates of a ConvNet-BF classifier in maximum credibility rule
** The total error rate is 36.30%, while the rejection rate is 5%
*** A rejection action is not considered as a error
**** The unit in the table is %

Test results of the MNIST data set



- NIN=network in network, a type of ConvNet
- Curves of credibility, plausibility, and Pignistic probability presents the results of ConvNet-BF classifiers with different decision rules
- Rejection is not considered as an error

Conclusions and Perspective

1. Conclusions
   (a) The proposed classifiers can reduce the errors by rejecting a part of the incorrect classification
   (b) The proposed classifiers are prone to assign a rejection action when there are conflicting features
   (c) The method opens a way to explain the relationship between the extracted features and class membership of each pattern
2. Perspective
   (a) Other evidence-theoretic rules for set-valued classification
   (b) Pixel-wise recognition using the proposed model

References

T. Denœux.
A neural network classifier based on Dempster-Shafer theory.
IEEE transactions on SMC A, 30(2):131-150, 2000.

M. Lin, Q. Chen, S. Yan.
Network in network
2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, 2014.

Thank you!