Feature Learning by Least Genralization

Hien D. Nguyen, Chiaki Sakama

University of Information Technology, VNU-HCM, Vietnam

Wakayama University, Japan

1st International Joint Conference on Learning & Reasoning

25 - 27 October 2021
Virtual

IJCLR 2021

Contents

- Provide an empirical study for feature learning based on induction.
- Encode image data into first-order expressions and compute their least generalization. An interesting question is whether the least generalization can extract a common pattern of input data.
- There are three different methods for feature extraction based on symbolic manipulation were proposed.
- We perform experiments using the MNIST datasets and show that the proposed methods successfully capture features from training data and classify test data in around 90% accuracy.

Methodology

- An image (in black, white or grayscale) is presented by $28 \times 28 = 784$ pixels where each pixel is an integer value from 0 to 255.
- An image is then represented as a vector $\boldsymbol{v} \in \mathbb{R}^{784}$ that contains pixel values as elements.
- Each pixel x $(0 \leq x \leq 255)$ is transformed to the term $f^k(z)$ with a variable z where

$$k = \left\lfloor \frac{x}{64} \right\rfloor + 1,$$

$$f^1(z) = f(z) \text{ and } f^{k+1}(z) = f(f^k(z)) \ (1 \leq k \leq 3).$$

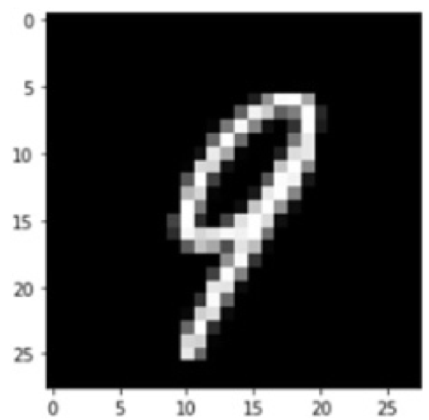
The function symbol f is used to represent "closeness" of pixels.

$$(\underbrace{f(z),\ldots,f(z),}_{28\times 6\,\text{values}} \qquad \%\text{1st to 6th rows}$$

$$\underbrace{f(z),\ldots,f(z)}_{16\,\text{values}},\ f^3(z),\ f^4(z),\ f^4(z),\ f^3(z),\ \underbrace{f(z),\ldots,f(z),}_{8\,\text{values}} \qquad \%\text{7th row}$$

$$\underbrace{f(z),\ldots,f(z)}_{14\,\text{values}},\ f^2(z),\ f^4(z),\ f^4(z),\ f^2(z),\ f^2(z),\ f^4(z),\ \underbrace{f(z),\ldots,f(z),}_{8\,\text{values}} \qquad \%\text{8th row}$$

$$\ldots \quad \underbrace{,f(z),\ldots,f(z)}_{28\,\text{values}}) \qquad \%\text{28th row}$$



Extracting Features

- Suppose a set of training data $C_l = \{A_1, \ldots, A_n\}$ (called a *class*) where l is a label and A_i $(1 \leq i \leq n)$ is a first-order atom (or a tuple) representing an image.
- Compute the least generalization of C_l using an algorithm in the literature (e.g. [Nguyen& Sakama, ILP 2019]).

The obtained vector $\boldsymbol{u} \in \mathbb{R}^{784}$ is viewed as features extracted by least generalization. We call $\boldsymbol{u}$ a *feature vector* by **GEN**.

Suppose a set of training data $D_l = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ where l is a label and $\boldsymbol{v}_k \in \mathbb{R}^{784}$ $(1 \leq k \leq n)$ is a vector representing an image. Put

$$\boldsymbol{v}_k = (x^k_{1.1}, \ldots, x^k_{1.28}, x^k_{2.1}, \ldots, x^k_{2.28}, \ldots, x^k_{28.1}, \ldots, x^k_{28.28}) \quad (1 \leq k \leq n)$$

where x^k_{ij} is a pixel value. Then, define

$$S_{ij} = \{x^k_{ij} \mid 1 \leq k \leq n\} \quad (1 \leq i, j \leq 28).$$

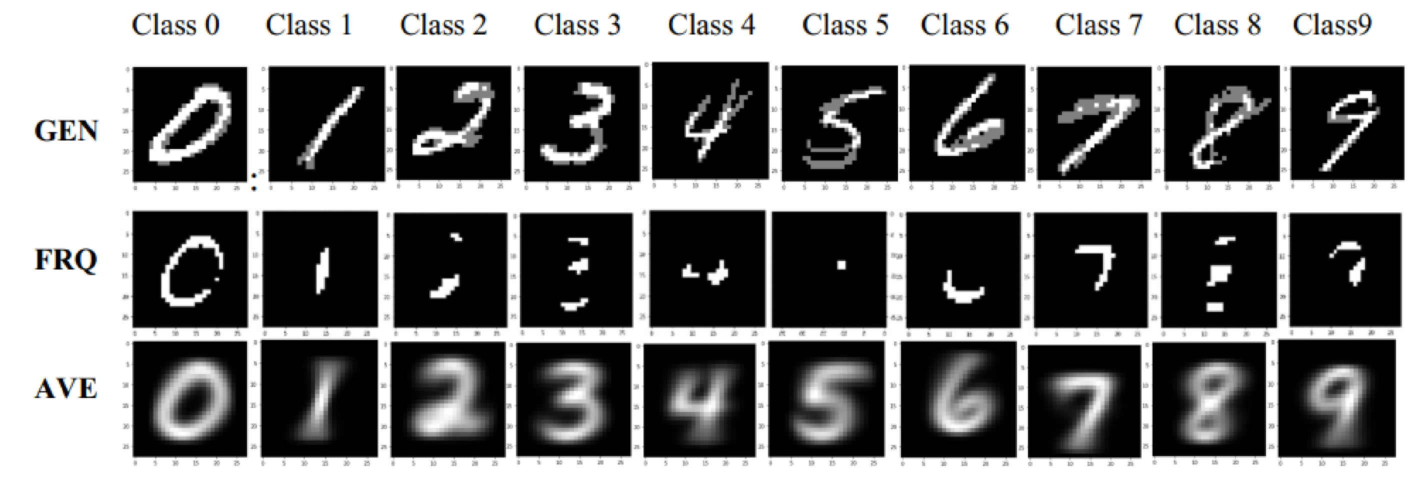
S_{ij} is a collection of pixel values at the location (i, j) from training data. Then, **FRQ** and **AVE** are defined as follows.

FRQ: Select the integer value u_{ij} that appears most frequently in S_{ij}.

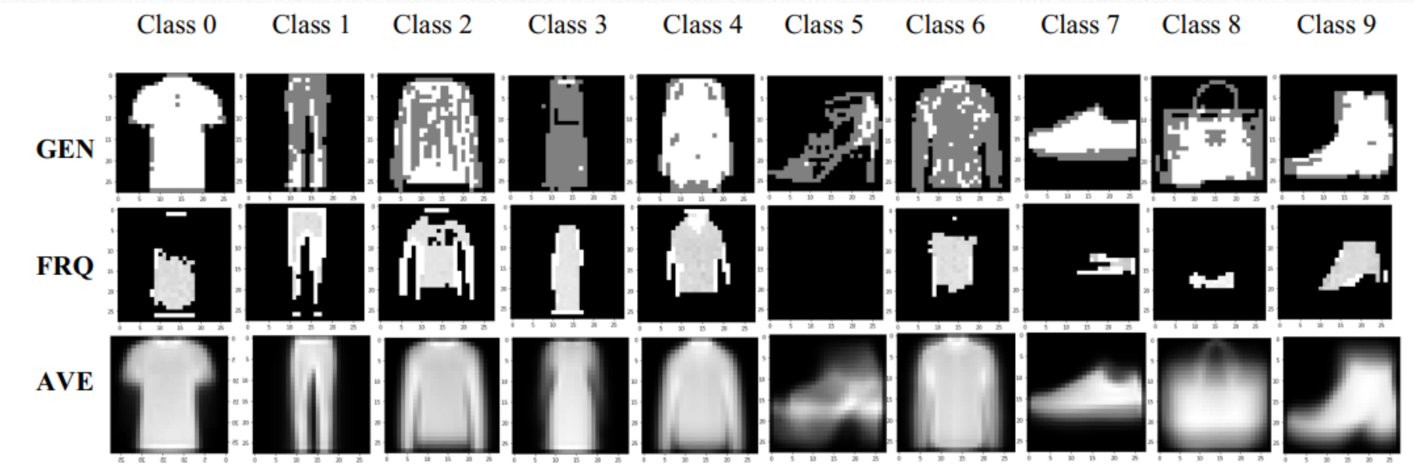
AVE: Compute $w_{ij} = \lfloor v_{ij} \rfloor$ where v_{ij} is the average value of elements in S_{ij}.

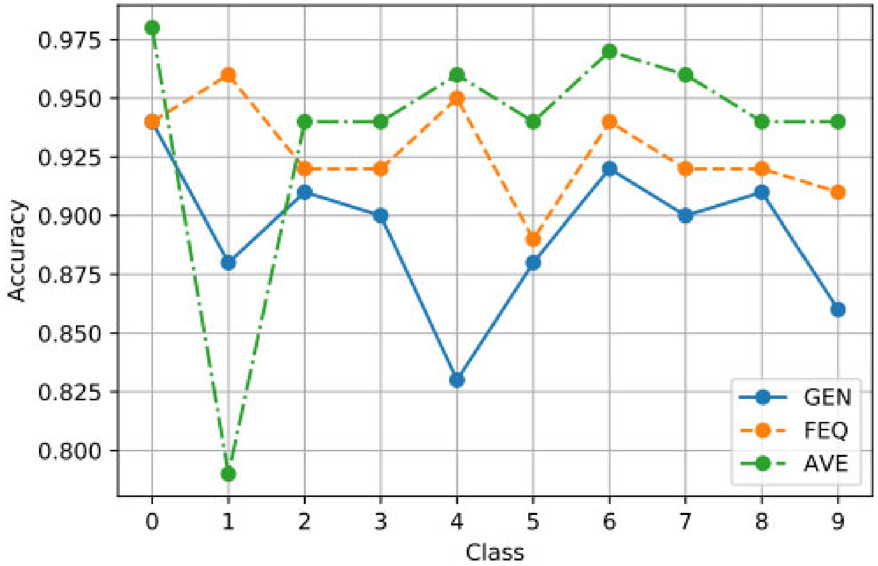
We call $\boldsymbol{u} = (u_{ij})$ (resp. $\boldsymbol{w} = (w_{ij})$) *a feature vector* by **FRQ** (resp. **AVE**).

MNIST dataset

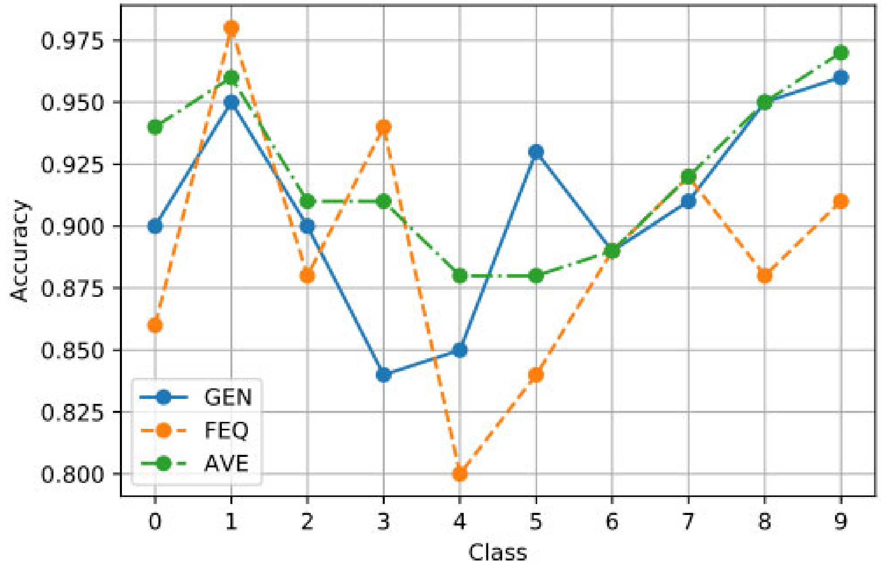


Fashion-mnist dataset





MNIST dataset



Fashion-mnist dataset

Conclusions

- This paper introduced new methods that learn features of labelled images by symbolic reasoning.
- This approach is purely symbolic and does not use NN for learning from image data.
- This is the first attempt that realizes feature learning using symbolic reasoning without relying on neural network.
- We continue experiments to verify the effect of noise in images.

Thank you !

Contact: **Hien D. Nguyen**
hiennd@uit.edu.vn

Contact: **Chiaki Sakama**
sakama@wakayama-u.ac.jp