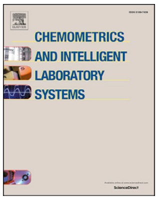
Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Itakura-Saito distance based autoencoder for dimensionality reduction of mass spectra

Yuji Nozaki, Takamichi Nakamoto*

Tokyo Institute of Technology, 4259 Nagatsuta, Midori, Yokohama, Kanagawa, 226-8503, Japan

ARTICLE INFO

Keywords:
Dimensionality reduction
Mass spectra
Autoencoder
Itakura-Saito distance

ABSTRACT

Small signals may contain important information. Mass spectra of chemical compounds are usually given in a format of sparse high-dimensional data of large dynamic range. As peaks at high *m/z* (mass to charge ratio) region of a mass spectrum contribute to sensory information, they should not be ignored during the dimensionality reduction process even if the peak is small. However, in most of dimensionality reduction techniques, large peaks in a dataset are typically more emphasized than tiny peaks when Euclidean space is assessed. Autoencoders are widely used nonlinear dimensionality reduction technique, which is known as one special form of artificial neural networks to gain a compressed, distributed representation after learning. In this paper, we present an autoencoder which uses IS (Itakura-Saito) distance as its cost function to achieve a high capability of approximation of small target inputs in dimensionality reduction. The result of comparative experiments showed that our new autoencoder achieved the higher performance in approximation of small targets than that of the autoencoders with conventional cost functions such as the mean squared error and the cross-entropy.

1. Introduction

The use of machine learning has spread widely over various fields in the last decade. Above all, the so-called ‘deep learning’ method emerging at the beginning of 2000's has shown great impacts in many cognitive computing applications [1], such as image recognition and speech recognition. In these successful applications, dimensionality reduction, which compresses higher dimensional data into manageable lower dimensional data, plays a fundamental role as redundancy in the input hurts the performance. An autoencoder, regarded as one special form of artificial neural networks, is known as one of the promising techniques of nonlinear dimensionality reduction [2]. By an autoencoder, the dimensionality of input data is reduced to a smaller dimensionality at the hidden layer as the number of the neurons in the hidden layer is smaller than that of input layer. Previous studies reported that autoencoder with nonlinear characteristic is more suitable than linear dimensionality reduction techniques such as PCA (Principal Component Analysis) when the data to be handled have nonlinear structure [3,4].

Learning process of neural networks can be converted into the problem of minimizing the cost function, the distance between the target signal and the output of the neural network. The cost function between the input and the target signals is minimized by optimizing a set of parameters in the network. For cost function, MSE (Mean Squared Error) and CE (Cross Entropy) are well known and have been used in a variety of machine learning systems so far [5]. However, it should be noted that these cost functions are mainly aimed to capture the large features in the target dataset. On the other hand, IS distance, a distance based on a logarithmic scale, is known as the distance which reflects perceptual similarity [6].

The Mass spectrum is one of the representative physicochemical properties of chemical substances. The dynamic range of mass-spectrum data is very large, however tiny peaks in the mass spectrum should not be ignored since it was argued that these peaks may affect our olfactory perception [7]. The Previous study showed a successful dimensionality reduction technique using IS distance based Non-negative Matrix Factorization [8,9]. In the IS distance, the meaningful tiny peaks can contribute to the distance, whereas they tend to be ignored in the Euclidean distance. Thus, IS distance is more sensitive to intensities near 0. This characteristic of IS distance is preferable to the other distance metrics for our application such as mass-spectrum approximation since it is suggested the small peaks in mass spectrum of a chemical at high *m/z* region contribute to its odor character [7].

As peaks in low *m/z* region mainly originated from typical molecules with low molecular weights, they could be much larger than those in high *m/z* region. However, these peaks are not important for our olfaction as they are molecules with relatively high human threshold. Although peaks

* Corresponding author.
E-mail address: nakamoto@mn.ee.titech.ac.jp (T. Nakamoto).

http://dx.doi.org/10.1016/j.chemolab.2017.05.002
Received 1 February 2017; Received in revised form 28 April 2017; Accepted 2 May 2017
Available online 20 May 2017

in high m/z region are relatively smaller than those in low m/z region, they could contain more important information for olfaction therefore they should be utilized. The molecules of large molecular weights which have relatively low human threshold tend to contribute to small peaks in high m/z region. Thus, an accurate approximation of small values in feature vectors would be quite useful to applications such as odor character prediction from its mass spectra [10].

Motivated by these backgrounds, we studied an autoencoder based upon IS distance. This paper presents an autoencoder with IS distance as cost function and compares the results of experiments on dimensionality reduction with those of the Euclidean and cross entropy distances to show the improvement in reproduction of small values in target dataset.

2. Cost functions

Backpropagation algorithm is one of the well known methods to train artificial neural networks including autoencoders [2]. In the algorithm, gradients of a cost function are calculated with respect to the weights and the biases in the network for the purpose of minimizing the cost function. Gradients should be iteratively calculated for each layer so that the gradient descent method can optimize the weights in the entire network.

Then, we firstly derived a gradient of IS distance with respect to weights and biases. Letting y be a target signal, an element of original spectrum, given to an autoencoder and $f(z)$ be an output of sigmoid function on input value z, an element of reproduced spectrum, the cost function based on IS distance E_{IS} is given by the following equation [6],

$$E_{IS} = \frac{y}{f(z)} - \ln\frac{y}{f(z)} - 1, \tag{1}$$

MSE (Mean squared error) and CE (Cross entropy) are quite common cost functions and are used in artificial neural networks in most cases. The cost functions of MSE and CE are given by the following equations [5].

$$E_{\mathrm{MSE}} = \frac{1}{2}(y - f(z))^2, \tag{2}$$

$$E_{CE} = y\ln f(z) + (1-y)\ln(1-f(z)). \tag{3}$$

Fig. 1 shows how each cost function evaluates the distance between the two values. $f(z)$ and y are changed from 0 to 1 with a step size of 0.01. As shown in Fig. 1c and f, IS distance changes drastically when values are near 0, compared with the distances given by other cost functions.

Then, we calculated a gradient of each cost function with respect to the weights. With the chain rule of differential, the gradient of IS cost function, $\frac{\partial E_{IS}}{\partial w}$, can be described as,

$$\frac{\partial E_{IS}}{\partial \mathrm{w}} = \frac{\partial E_{IS}}{\partial f(z)}\frac{\partial f(z)}{\partial z}\frac{\partial z}{\partial w}\ , \tag{4}$$

Although the derivation here is just for a scalar value, it can be applied to every weight. When the sigmoid function, $f(z) = \frac{1}{(1+}\exp(-z))$, is used as the activate function of a network, each term in the right side of this equation can be described as,

$$\frac{\partial E_{IS}}{\partial f(z)} = -\frac{y}{f(z)^2} + \frac{1}{f(z)}, \tag{5}$$

$$\frac{\partial f(z)}{\partial x} = f(z)(1-f(z)), \tag{6}$$

$$\frac{\partial z}{\partial w} = x, \tag{7}$$

where $z = wx$. Thus, Equation (4) is now rewritten as

$$\frac{\partial E_{IS}}{\partial \mathrm{w}} = \left(1 - \frac{y}{f(z)}\right)(1-f(z))x. \tag{8}$$

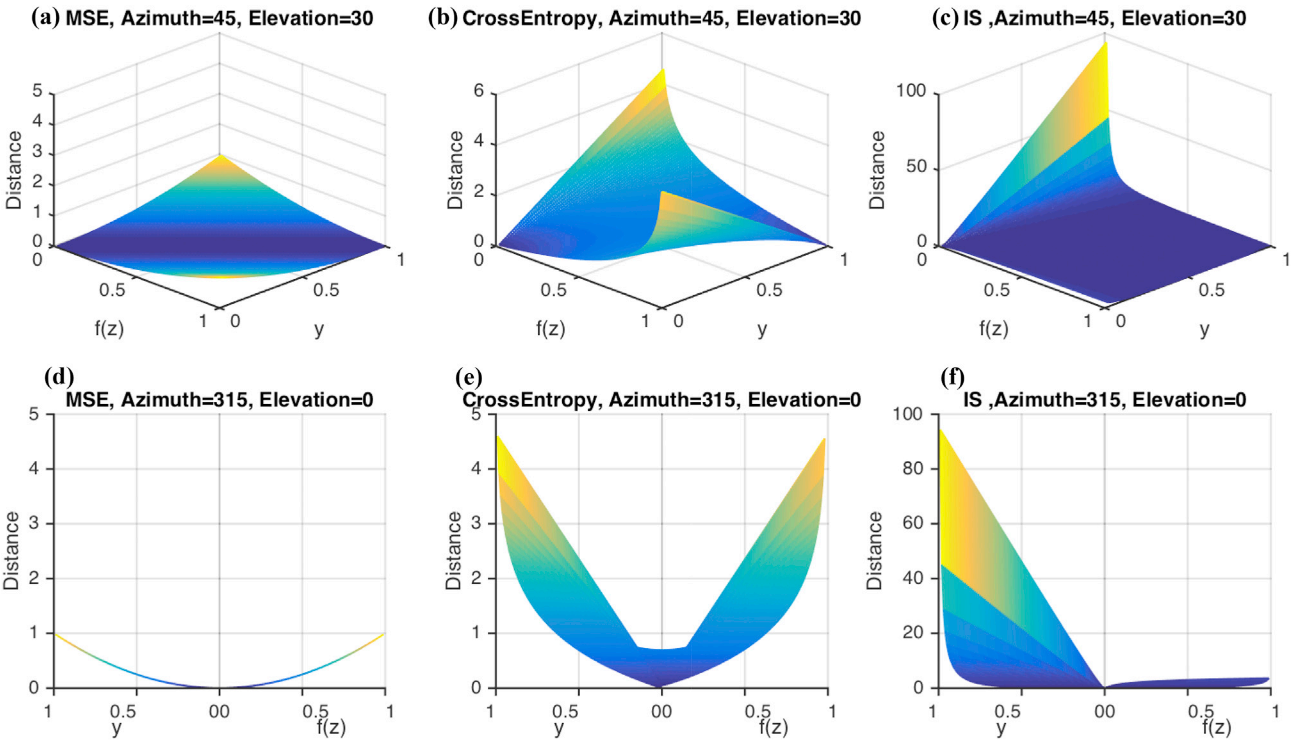


Fig. 1. Plots of distance between $f(z)$ and y calculated on three cost functions. Colors in each figure change depending on the magnitude of distance. (a) Mean squared error, Azimuth = 45°, Elevation = 30°, (b) Cross Entropy, Azimuth = 45°, Elevation = 30° (c) Itakura-Saito distance, Azimuth = 45°, Elevation = 30°, (d) Mean squared error, Azimuth = 315°, Elevation = 0°, (e) Cross Entropy, Azimuth = 315°, Elevation = 0° (f) Itakura-Saito distance, Azimuth = 315°, Elevation = 0°. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)