Similarity Learning and Visualization with Siamese Topological Networks

Pitoyo Hartono

School of Engineering, Chukyo University, 10-2 Yagotohonmachi, Showa-ku, Nagoya, 466-8666, Aichi, Japan.

Corresponding author(s). E-mail(s): hartono@ieee.org;

Abstract

In this study a model of Siamese topological neural networks, which consisted of a pair of hierarchical neural networks each with a low-dimensional internal layer, is proposed. Through similarity learning, the objective of the proposed siamese network is to learn low-dimensional topological representations of a given similarity between pairwise high-dimensional inputs. The low-dimensionality of the internal layer allows human to visualize the structure of the high-dimensional data in the context of their similarities, for example, label-similarity or rank-similarity. Different from many similarity learning techniques, dimensionality reduction is integrated in the proposed siamese networks allowing flexible context-oriented visualization analysis for high-dimensional data.

Keywords: Similarity Learning, Self-Organizing Maps, Visualization, Topological Representation

1 Introduction

In past few years, Similarity Learning Mathisen et al (2020); Kornblith et al (2019) have been extensively studied. The primary objective of for the Similarity Learning given a pair of inputs $(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ and their similarity measure $T(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \in \mathbb{R}$ is to find an approximation function $f(\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) \sim T(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$. The approximation function measures the similarity of the pair of inputs and can be expressed using bilinear matrix or non-linear functions

that are hand-designed or learned with various models of single neural networks Liu et al (2019), or more recently, Siamese NetworksChicco (2021); Koch and Zemel (2015); Jeon et al (2019); Berlemont et al (2015).

In this study, a new model of Siamese Network with the objective to learn a given similarity measure between a pair of inputs and at the same time to generate their low-dimensional representations for the visualization is proposed. The base network for the proposed Siamese Topological Networks (STN) is a hierarchical neural network containing a low-dimensional topological layer that maps high-dimensional input into a two-dimensional topological map that can be visualized. Most of the dimensional-reduction methods for visualization, for example, t-SNE van der Maaten and Hinton (2008), isomap Zhang et al (2018), Umap McInnes et al (2018) reduce the dimensionality while preserving a criterion of similarity between the high-dimensional data in their low dimensional representations. Different from these unsupervised dimensionality reduction algorithms that do not take the context of the data, for example, their labels, into account, STN forms a low-dimensional embedding that also reflects the context-similarity, and thus expands the visualization flexibility of high-dimnensional data. However, STN is not strictly a supervised algorithm, as it does not require the availability of data labels.

In the past there are many supervised dimensionality reduction methods Ghosh and Kirby (2022); Vogelstein et al (2021) that also take the data labels into account, Linear Discriminant Analysis (LDA) Fisher (1936); Tharwat et al (2017) being the most traditional example. While the idea for supervised dimensionality reduction methods has been known for a very long time, it is still actively studied, resulting in many interesting algorithms Goldberger et al (2004); Hartono (2018); Mika et al (1999); Hastie and Tibshirani (1996); Raducanu and Dornaika (2012). These methods formed low-dimensional representations in the context of the data label, and hence, consequently, the labels have to be provided. However, for real-world problems, data labels are not necessarily available and can be expensive to obtain. In the proposed study, pairwise similarity measure is required, but data labels are not. Here, the STN is trained to learn the similarity measure while at the same time executing dimensionality reduction.

The proposed study shares some resemblances with the past study of dimensionality reduction through kernel-based similarity learning for high-dimensional data Wang et al (2017), in that the networks execute similarity learning that that can be related to some given contexts, for example rank. However, this method needs to depend on separate dimensionality reduction method for visualization, while in the proposed STN, the dimensionality reduction is embedded.

The novelty of this study is two-fold: First showing that a topological representations of high-dimensional data can be learned from a set of paired data with arbitrarily designed similarity measure, resulting in a context-oriented representation that depends not only on the feature-similarity of the inputs

but also the similarity-context given to them. The second that the low-dimensionality of the internal representation allows the visual analysis on the high-dimensional data. This is a strength of the proposed model, because while many similarity learning methods require separate visualization algorithms, for example, t-SNE van der Maaten and Hinton (2008) being one of the most popular methods for visualizing their representations. In this study, the dimensionality reduction mechanism that allows visualization is an inherent part of the network. Further, as the topological representations here are context-oriented, in that different similarity criteria result in different topological structures, a flexible visual analysis is made possible.

The rest of the paper is structured as follows. Section 2 explains the structure of the STN, its learning dynamics, and the framework of the proposed STN as similarity learning. The subsequent section is for describing the experiments, while the conclusions will be elaborated in the final section.

2 Siamese Topological Networks

The structure of the proposed Siamese Topological Networks (STN) is outlined in Fig. 1. The base network for the STN is a hierarchical topological network, proposed in Hartono et al (2015); Hartono (2020). The structure of the base network is similar to the standard Multilayered Perceptron (MLP), except that the hidden layer is a two-dimensional topological map as in many cases of Self-Organizing Maps (SOM) Kohonen (1982, 2013). The low dimensionality allows the network to form topological representations of a given data set that can be visualized by human. The visualization of the internal representations, to some extent, enables human to intuitively understand how the network converts input into its output. In the past, it is utilized for visualizing high-dimensional data Hartono (2018). It can also be trained as an autoencoder, a classifier, or a mix of both Hartono (2020), and hence is able to visualize high-dimensional data constrained by various contexts.

For this study, two topological networks are paired to form the STN for similarity learning. The objective is to train the network with high-dimensional data in which a criterion of similarity between a pairwise data instances is given. The similarity measure may be defined qualitatively, quantitatively, or subjectively, and does not need to meet mathematically defined distance. The low dimensional topological map can be utilized for visualizing the structure of the data in the context of the defined similarity. Here, different similarity measure for the same data will generate a different map.

2.1 Learning Process

The learning process of STN is explained as follows.

For each of the pair of high-dimensional input to the network, $\mathbf{X}^{(k)}(k \in \{1, 2\})$, the best matching unit, $win^{(k)}$ is selected as in Eq. 1.

$$win^{(k)} = argmin_j \|\mathbf{W}_j - \mathbf{X}^{(k)}\| \tag{1}$$

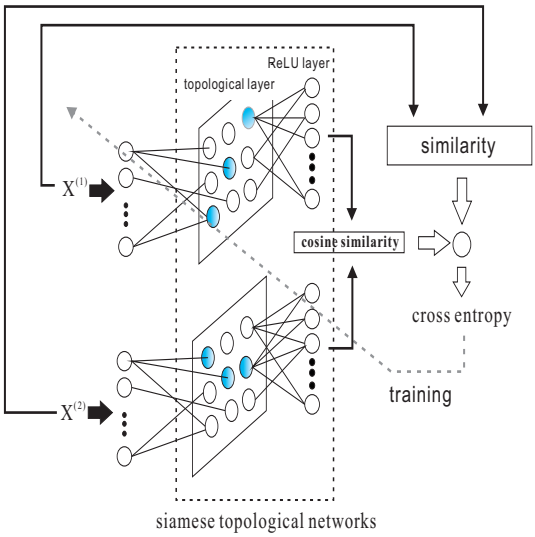


Fig. 1: Siamese Topological Networks

The activation of the j-th hidden neuron, $h_j^{(k)}(t)$, incurred by one of the input, $\mathbf{X}^{(k)}$, of the pair at epoch t is calculated as in Eq. 2. Here, $\mathbf{h}^{(k)}$ denotes the activation vector incurred by input $\mathbf{X}^{(k)}$.

$$\begin{aligned} h_j^{(k)}(t) &= \sigma(win^{(k)}, j, t) e^{-\|\mathbf{X}^{(k)}(t) - \mathbf{W}_j(t)\|^2} \\ \mathbf{h}^{(k)}(t) &= (h_1^{(k)}(t), h_2^{(k)}(t), \cdots h_{N_{hid}}^{(k)}(t))^{\mathrm{T}} \\ & (k \in \{1, 2\}) \end{aligned} \tag{2}$$

Here, N_{hid} is the number of the neurons in the topological layer, while the neighborhood function $\sigma(win^{(k)}, j, t)$ is defined as follows.

$$\begin{aligned} \sigma(win^{(k)}, j, t) &= exp(-\frac{dist(win, j)}{S(t)}) \\ S(t) &= \sigma_\infty + \frac{1}{2}(\sigma_0 - \sigma_\infty)(1 + cos\frac{\pi t}{t_\infty}) \end{aligned} \tag{3}$$

In Eq. 3, σ_0 and σ_∞ are the initial and terminal annealing constants, t_∞ is the termination epoch, while $dist(win^{(k)}, j)$ is the distance from the BMU to the j-th neuron on the topological layer.

The output of the ReLU layer, $\mathbf{O}^{(k)} = (O_1^{(k)}, O_2^{(k)} \cdot \cdot, O_{N_{out}}^{(k)})$, where N_{out} denotes the number of the neurons in ReLU layer, is as follows.

$$O_m^{(k)} = ReLU(\mathbf{V}_m \cdot \mathbf{h}^{(k)}) \tag{4}$$
$$ReLU(x) = \begin{cases} x & x \geqslant 0 \\ 0 & otherwise \end{cases}$$

In Eq. 4, $\mathbf{V}_m$ is the weight vector connecting the hidden layer with the m-th output neuron.

$$T(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \begin{cases} 1 & \mathbf{X}^{(1)}, \mathbf{X}^{(2)} : similar \\ 0 & (otherwise) \end{cases} \tag{5}$$

The cosine similarity between the activation vectors incurred by the input pair is defined in Eq. 6.

$$S(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \frac{\mathbf{O}^{(1)} \cdot \mathbf{0}^{(2)}}{\|\mathbf{O}^{(1)}\| \|\mathbf{O}^{(2)}\|} \tag{6}$$

Because $\forall j, k \in \{1, 2\}, 0 < h_j^{(k)} \leqslant 1$
in Eq. 6, $0 < S(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \leqslant 1$

The loss function is cross entropy defined as follows.

$$\mathcal{L}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = -(TlogS + (1-T)log(1-S)) \tag{7}$$

The gradients of the loss function are calculated as follows.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{V}_m} &= -\frac{(T-S)}{S(1-S)} \frac{\partial S}{\partial \mathbf{V}_m} \\ &= -\frac{(T-S)}{(1-S)} \sum_{k=1}^{2} sgn_m^{(k)} \delta_m^{(k)} \mathbf{h}^{(k)} \\ &\quad (m \in \{1, 2, \cdots, N_{hid}\}) \end{aligned} \tag{8}$$

In Eq.8,

$$\delta_m^{(k)} = \frac{O_m^{(\bar{k})}}{\mathbf{O^{(1)}} \cdot \mathbf{O}^{(2)}} - \frac{O_m^{(k)}}{\|\mathbf{O}^{(k)}\|^2},$$
$$\bar{k} = \begin{cases} 2 & (k = 1) \\ 1 & (k = 2), \end{cases}$$

and,

$$sgn_m^{(k)} = \begin{cases} 1 & (h_m^{(k)} > 0) \\ 0 & (otherwise), \end{cases}$$

Hence, the weight vector from the topological layer to the ReLU layer is modified as follows.

$$\mathbf{V}_m(t+1) = \mathbf{V}_m(t) + \eta \frac{(T-S)}{(1-S)} \sum_{k=1}^{2} sgn_m^{(k)} \delta_m^{(k)} \mathbf{h}^{(k)} \tag{9}$$

While the modification rule is executed as in standard Backpropagation, the interpretation for modification rule in Eq. 9 is interesting. When the teacher signal $T = 1$, indicating that the two inputs are similar, the direction of the modification only depends on $\delta_m^{(k)}$ that measures the difference between the influence of the m-th element from the $\bar{k}$-th input to the cosine similarity and that of the k-th input to the norm of $\mathbf{O}^{(k)}$. Here, for $\delta_m^{(k)} > 0$ the weight vector is corrected toward $\mathbf{h}^{(k)}$ that in turn will bring $\mathbf{0}^{(k)}$ closer to $\mathbf{0}^{(\bar{k})}$ and thus increasing their cosine similarity. For $\delta_m^{(k)} < 0$, the weight is modified in the opposite direction of $\mathbf{h}^{(k)}$, and hence will decrease $O_m^{(k)}$ and hence will bring $\mathbf{0}^{(k)}$ closer to $\mathbf{0}^{(\bar{k})}$ as well. Using the same rationale, it is obvious that when the teacher signal $T = 0$, the weight is modified so that the cosine similarity between $\mathbf{0}^{(k)}$ and $\mathbf{0}^{(\bar{k})}$ decreases.

The gradient of the loss function with regard to the reference to the reference vector that is associated with the m-th neuron in the topological layer is calculated as follows.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_n} &= -\frac{(T-S)}{S(1-S)} \frac{\partial S}{\partial \mathbf{W}_n} \\ &= -2 \frac{(T-S)}{(1-S)} \sum_{k=1}^{2} \epsilon_n^{(k)} h_m^{(k)} (\mathbf{X}^{(k)} - \mathbf{W}_n) \end{aligned} \tag{10}$$

In Eq. 10, the term $\epsilon_n^{(k)}$ is the weighted correction signal from the ReLU layer that is calculated as follows.

$$\epsilon_n^{(k)} = \sum_{m=1}^{N_{out}} sgn_m^{(k)} v_{mn} \delta_m^{(k)} (k \in \{1, 2\}) \tag{11}$$

In Eq. 11, v_{mn} is the weight from the m-th neuron in the topological layer to the n-th neuron in the ReLU layer, while N_{out} denotes the number of neurons in the ReLU layer.

Hence, the reference vector modification is as follows.

$$\mathbf{W}_n(t+1) = \mathbf{W}_n(t) + \eta \frac{(T-S)}{(1-S)} \sum_{k=1}^{2} \epsilon_n^{(k)} h_m^{(k)} (\mathbf{X}^{(k)} - \mathbf{W}_n) \tag{12}$$

Equation 12 significantly differs from that of conventional SOM, in that in SOM the reference vectors are always modified toward the input vector, while in STN the direction of modification is influenced by the sign of $(T-S)$, and the sign of $\epsilon_m^{(k)}$. The former indicates that during the learning process, the formation of the topological map is not only influenced by the topological

structure of the inputs, but also their context, in that the same set of inputs produce different topological maps if different similarity rules are assigned to them. The latter is the weighted sum of $\delta_m^{(k)}$ that can be considered as a result of votes between all the outputs neurons to decide the direction of the reference vector modification that is influenced by the similarity-contexts of the inputs as well.

2.2 A framework for similarity learning

In this study, the learning process can be formulated as similarity measure learning Cunningham (2009); Mathisen et al (2020). The framework for the similarity learning can be formulated as follows. Given a pair of high-dimensional inputs, $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ and their arbitrarily defined similarity $T(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$, the objective of the learning process is to generate the following approximations.

$$S(ReLU(\mathbf{V}^{\mathrm{T}}\mathbf{h}_{\mathbf{W}}(\mathbf{X}^{(1)}), ReLU(\mathbf{V}^{\mathrm{T}}\mathbf{h}_{\mathbf{W}}(\mathbf{X}^{(2)})) \sim T(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \tag{13}$$

In Eq. 13, $\mathbf{V} \in \mathbb{R}^{N_{hid} \times N_{out}}$ is the weight matrix leading from the internal topological map to the ReLU layer that should be learned and $\mathbf{h}_{\mathbf{W}}(\mathbf{X}) \in \mathbb{R}^{N_{hid}}$ is a parameterized mapping function that projects the d-dimensional input $\mathbf{X} \in \mathbb{R}^d$ into the output of the topological layer. Here, while $S()$ is a fixed cosine similarity function, and $ReLU()$ is a rectified linear function that is fixed as well, $\mathbf{V}$ is a bilinear matrix and $\mathbf{W}$ is the reference matrix for the topological layer that should be learned, and hence they formed a non-linear similarity measure that should be learned from the pairwise training data.

Further, the winner-takes-all process in the topological layer enables mapping of the d-dimensional input $\mathbf{X}$ into a coordinate $cor(\mathbf{X}) \in \mathbb{R}^2$, as follows.

$$cor(\mathbf{X}) = pos(BMU(\mathbf{X})) \tag{14}$$

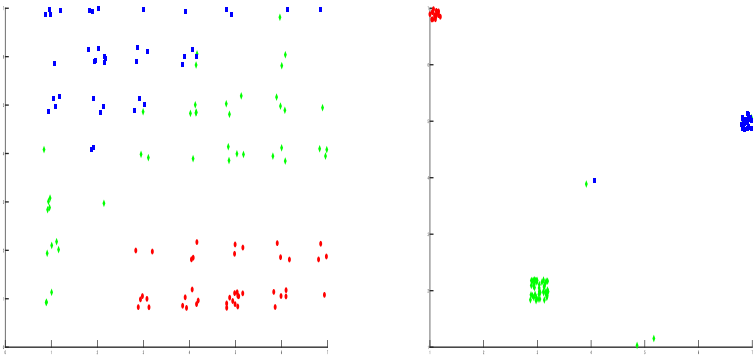
Here, $pos(BMU(\mathbf{X}))$ is the two-dimensional coordinate of the BMU for $\mathbf{X}$ on the internal topological layer.

Hence, in this study the objective of STN is twofold, one is for learning the similarity measure that approximate the given similarity criterion, and the other for reducing the dimensionality of the input for the purpose of visualization.

3 Experiment

3.1 Preliminary Experiments

The proposed STN is first tested against some simple labeled data from UCI machine learning repository UCI (2022). Here, although the data are labeled, the labels were not explicitly utilized during the learning process. Here, when



(a) SOM (b) TSN

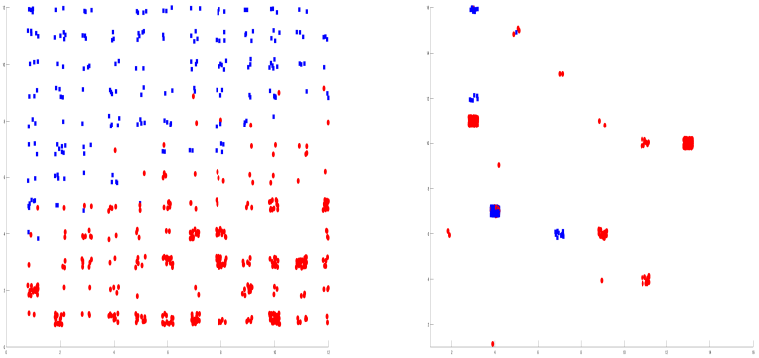
Fig. 2: Topological Representation for Iris Problem

an arbitrarily chosen pair shares a same label they are defined as similar, and they are defined to be dissimilar otherwise. The resulting topological map is then visually compared with that of standard SOM.

The first example is the well-known Iris problem, a four-dimensional three-classed data set. Here, during the learning process, the class labels of the data are not explicitly utilized, but used to define the similarity between two samples. In Fig. 2a, the topological structure of the data is visualized using the standard SOM. Here, for visualization clarity, the samples are colored according to their class labels, although the class labels did not have any role in the self-organizing process. From this figure, a well known profile of this problem, that one of the classes (illustrated by ●s) is linearly separable with the other two classes (illustrated by ■s and ♦s that are not linearly separable, can be observed.

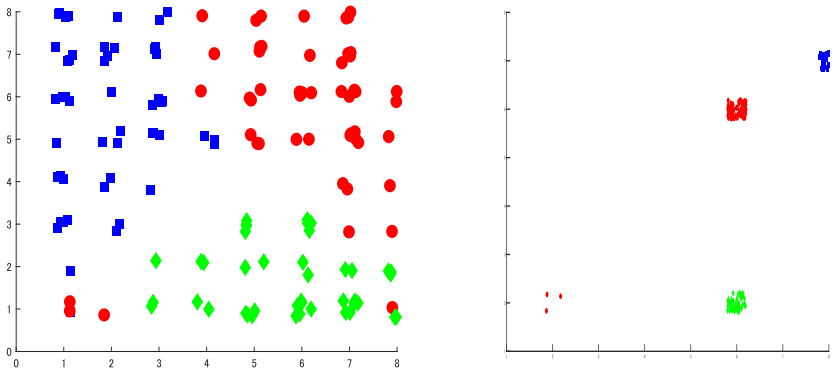
Figure 2b shows the topological representations of STN. It can be observed from this figure that STN generates sparser representations of the high dimensional input, in which the three classes are equally separable except for some anomalies. The better separability is due to the pairwise-similarity information that is not available to SOM. It is also interesting to notice that although the class labels are not explicitly provided, STN is able to form distinctive clusters according to the data labels.

The second example is Breast Cancer problem, a 9-dimensional 3-classes problem. The SOM representation in Fig. 3a indicates that while there are some overlapping samples, most of the instances can be easily classified. As shown in Fig. 3b, the STN generates sparser representations but also nicely visualizes the instances that are potentially hard to classify due to their similarity with other samples belonging to contrasting classes.



(a) SOM (b) TSN

Fig. 3: Topological Representation for Breast Cancer Problem



(a) SOM (b) TSN

Fig. 4: Topological Representation for Wine Problem

The third problem is Wine problem, a 13-dimensional 3-classes problem. Figure 4a shows the SOM representations. It can be observed that ●s form two distinct clusters that are also well represented by the STN as in Fig. 4b.

The three examples illustrated the ability of STN to form not only topologically correct representations but also reflect given context of similarity, which in these three examples are their class-similarities.

3.2 Ranked-data

In the previous preliminary experiments, STN is trained with data that are originally labeled and hence whose similarities can be naturally defined from their original labels. Here, the STN is trained using data that have no natural labels but their similarities can be arbitrarily designated. For example, countries are often ranked based on various criteria, for example, their economic power, the educational level of their populations, their richness in natural resources, and so on. These kinds of ranks usually only take the criteria that are used to define the ranks while ignoring other features. However, in visualizing the structure of the countries' relative similarities, rank can be considered as a context for the countries' profiles, and hence, for a clearer understanding of the similarity structure, both the rank and the features that characterized the countries should be taken into account. For example, it is intuitive to think that countries that have some similarities in their features and also similar ranks should be visualized close to each other, while at the other extreme, countries that have dissimilar profiles and have a wide rank gaps are visualized far from each other. Here, it is interesting to observe, for example, how countries that share similar profiles but have wide rank gap, or countries that have dissimilar profiles but have similar ranks, are aligned. This kind of context-oriented visualization may be interesting in understanding deeper inherent structures of their relative similarities.

For the subsequent experiments, 32 Asia and Pacific countries are chosen. Each country is characterized by 11 economic and demographic profiles, which are: 1. Life expectancy at birth, 2. Fertility rate (general population), 3. Under-5 mortality rate, 4. Immunization (measles) rate, 5. CO_2 emissions per capita, 6. Start-up procedure to register a business, 7. Foreign direct investment, 8. Age dependency ratio, 9. Access to electricity, 10. Access to clean fuels and technology for cooking, 11. GDP per capita. These data are obtained from World Bank's DataBank DataBank (2016) and are standardized. While a country can be characterized using many more profiles, the chosen variables are sufficient to show at least the economic prowess of those countries that may correlate with many factors as their level of education, the state of population welfare, military strength, and so on.

Figure 5 is the SOM'srepresentations of the 32 countries, in which the positions on the map reflect the profiles' similarities. It can be observed that economic powers in these region, as Japan, NZ, Singapore, Australia and Rep. of Korea are aligned close each other, while PNG, Pakistan and Afghanistan are aligned far from those countries.

These data are then ranked according to a criterion that is not included in their original profiles. The objective here is to observe how STN generates different topological maps for different rank-criteria, and how it generalizes to out-of-sample data.

The first rank-criterion is the Health Expenditure per capita, in which the top 6 countries are Australia, Japan, NZ, Singapore, Rep. Korea, Maldives,

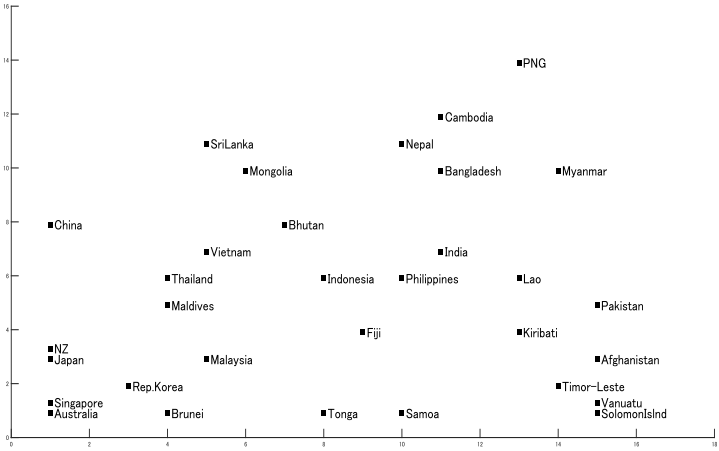


Fig. 5: Country Map: SOM

while the bottom 6 countries are, Bangladesh, Pakistan, Nepal, PNG, Lao, and Afghanistan.

For in addition to SOM that executes unsupervised self-organization, Sammon Maps Sammon (1969); Ghojogh et al (2020) , in which the pairwise distance between two countries, i and j defined in Eq. 15 is utilized. For visualization clarity, six top countries are illustrated in ★s, six bottom countries in •s, while other countries in ■s. It can be observed that Sammon Map nicely assigned rank wise-similar countries close to each other while distancing them with dissimilar countries.

$$dist(\mathbf{X}^{(i)}, \mathbf{X}^{(j)}) = \frac{\alpha}{11}\|\mathbf{X}^{(i)} - \mathbf{X}^{(j)}\| + (1-\alpha)|r(i) - r(j)| \tag{15}$$

In Eq.15, the distance between two countries is defined as weighted difference of their features as well as their ranks. Here, $\mathbf{X}^{(i)}$ and $r(i)$ are the profile vector and the standardized rank for country i, while $0 \leqslant \alpha \leqslant 1$ is the weighting coefficient that is set to 0.5.

The topological representations of STN to the same data set are shown in Fig. 7. For this experiment, in training the STN, two countries whose rank-different is less than 3 are designated as similar and they are designated as dissimilar, otherwise. In this figure, it is visually clear that the STN generated a more clustered-topological map. For example, the top four countries are concentrated in one cluster, while the number 5 and six countries (Korea and Maldives) form a different cluster, due to their profile differences from the top four countries. It is also interesting to observe that the bottom countries are more spread out than the top countries, which indicates that their profiles are more diverse. Furthermore, while it is not possible for Sammon Map to map

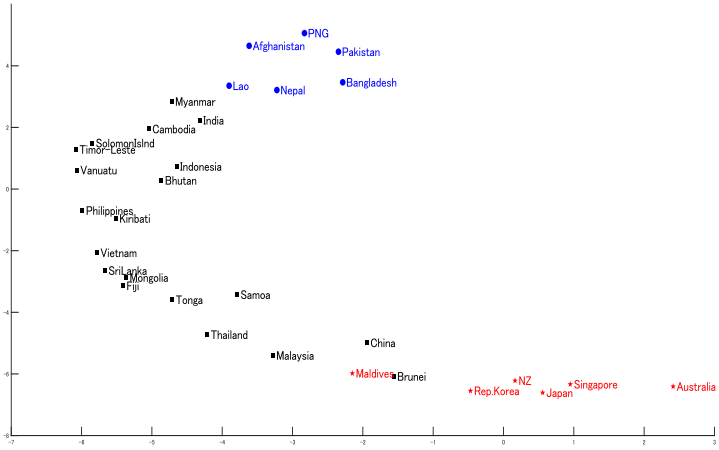


Fig. 6: Health Expenditure-ranked Sammon Map

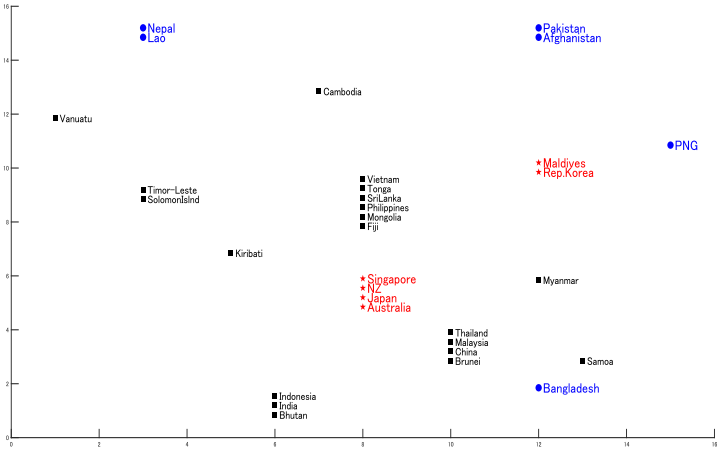


Fig. 7: Health Expenditure-ranked Map

out-of-sample data into the generated map, in STN it can be easily done by calculating the BMU for an out-of-sample country without retraining the STN.

For testing the out-of-sample data projection of STN, some European and Central Asia countries are projected as shown in Fig.8. From this figure, it can be learned that if Austria were in Asia, it can be predicted that Austria has a similar rank to Japan in health expenditure due to its profile similarity.

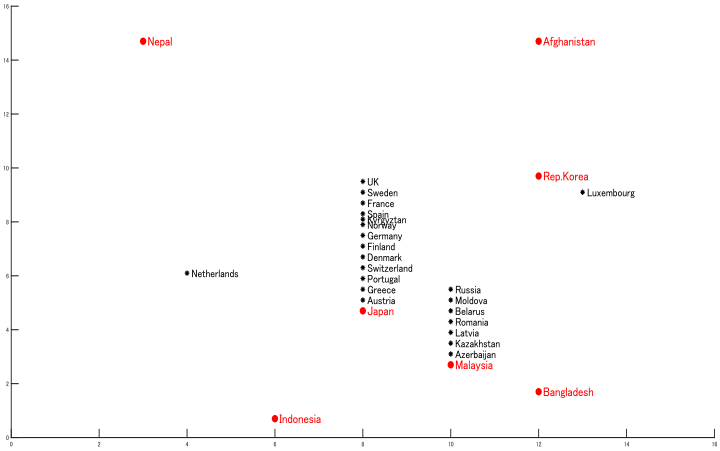


Fig. 8: Health Expenditure-ranked Out of samples

Similarly, Luxembourg with Korea, Azerbaijan with Malaysia, while no European and Central Asia country share similarities with Afghanistan and Nepal. Here for visualization clarity, only a few in-sample Asian countries are plotted. To differentiate from the out-of-sample countries, the in-sample countries are plotted in red.

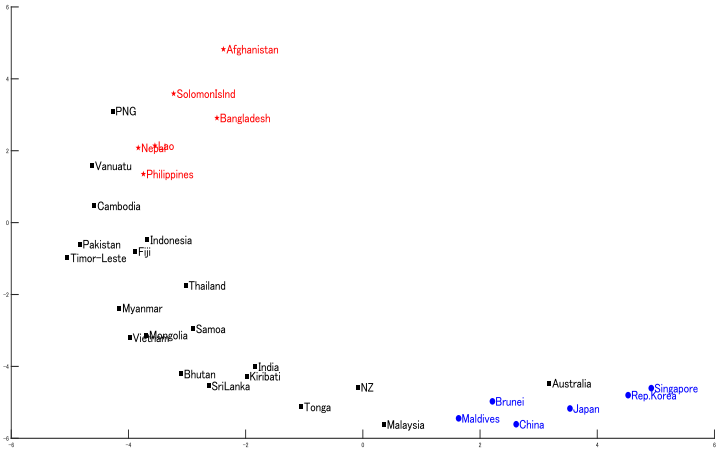


Fig. 9: Adolescence Fertility-ranked Sammon Map

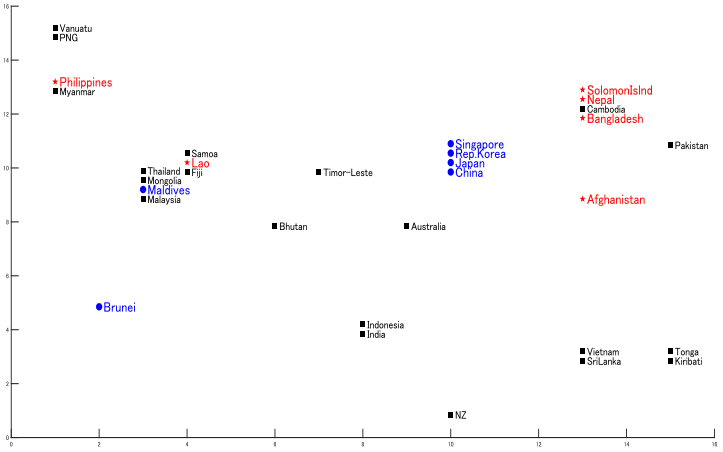


Fig. 10: Adolescence Fertility-ranked Map

In the next experiment, the Asian countries in the previous experiment are ranked using a different criterion that is Adolescence Fertility (births per 1,000 women ages 15-19). Here, the top six countries are Bangladesh, Afghanistan, Solomon Island, Nepal, Lao and the Philippines, while the bottom six countries are Korea, Singapore, Japan, China, Maldives, and Brunei.

The Sammon Map is given in Fig. 9, while the rank-oriented topological map generated by the STN is shown in Fig. 10. From this figure, among the top six countries, Solomon Island, Nepal and Bangladesh form a cluster that also includes Cambodia (ranked 9), due to their profiles similarities, while there is no country that shares large similarity with Afghanistan. For the bottom six countries, Singapore, Korea, Japan, and China form a cluster, while Brunei is solitary, and Maldives is close to Malaysia (ranked 25).

The out-of-sample projection is shown in Fig.11. Here in can be predicted that Austria will have similar rank with Japan, Luxembourg with Australia, Greece with Malaysia, Sweden and Finland with NZ and Kyrgyzstan with Tonga. Within the context of this ranking, no countries share similarity with Nepal.

4 Conclusions

In this study, a model of Siamese Neural Networks, is proposed for a flexible similarity visualization of high dimensional data. Here, it is shown that topological internal layer is able to capture the representation of flexible similarity-context given to high-dimensional data. Aside for learning good similarity measure and topological representations, STN also inherently embeds

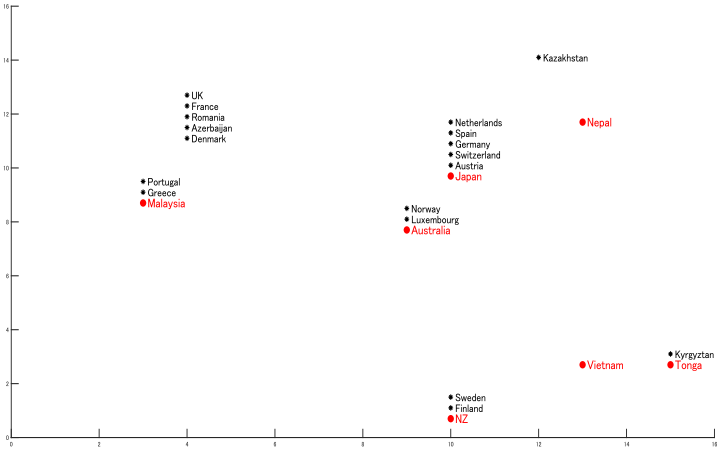


Fig. 11: Adolescence Fertility-ranked Out-sample

dimensionality reduction algorithm by producing context-oriented topological map that can be used for visual analysis.

In this short paper, the mathematical characteristics of STN was elaborated and its fundamental behaviour was clarified through simple experiments.

The experiments show that STN can flexibly learn the topological internal representations that are constrained by various similarity definitions.

As the strength of the proposed STN is in its ability to flexibly visualize context-oriented data, the immediate future works is to utilize STN as a tool for context discoveries, for example for drug response through similarity learning.

References

Berlemont S, Lefebvre G, Duffner S, et al (2015) Siamese neural network based similarity metric for inertial gesture classification and rejection. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp 1–6, https://doi.org/10.1109/FG.2015.7163112

Chicco D (2021) Siamese neural networks: An overview. In: Cartwright H (ed) Artificial Neural Networks. Methods in Molecular Biology, New York, https://doi.org/10.1007/978-1-0716-0826-5_3

Cunningham P (2009) A taxonomy of similarity mechanisms for case-based reasoning. IEEE Transactions on Knowledge and Data Engineering 21(11):1532–1543. https://doi.org/10.1109/TKDE.2008.227

DataBank (2016) World bank, world development indicators. URL https://databank.worldbank.org/indicator/NY.GDP.PCAP.CD/1ff4a498/Popular-Indicators

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7(2):179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Ghojogh B, Ghodsi A, Karray F, et al (2020) Multidimensional scaling, sammon mapping, and isomap: Tutorial and survey. https://doi.org/10.48550/ARXIV.2009.08136

Ghosh T, Kirby M (2022) Supervised dimensionality reduction and visualization using centroid-encoder. Journal of Machine Learning Research 23(20):1–34. URL http://jmlr.org/papers/v23/20-188.html

Goldberger J, Hinton GE, Roweis S, et al (2004) Neighbourhood components analysis. In: Saul L, Weiss Y, Bottou L (eds) Advances in Neural Information Processing Systems, vol 17. MIT Press

Hartono P (2018) Classification and dimensional reduction using restricted radial basis function networks. Neural Computing and Applications 30:905–915. https://doi.org/10.1007/s00521-016-2726-5

Hartono P (2020) Mixing autoencoder with classifier: Conceptual data visualization. IEEE Access 8:105,301–105,310. https://doi.org/10.1109/ACCESS.2020.2999155

Hartono P, Hollensen P, Trappenberg T (2015) Learning-regulated context relevant topographical map. IEEE Transactions on Neural Networks and Learning Systems 26(10):2323–2335. https://doi.org/10.1109/TNNLS.2014.2379275

Hastie T, Tibshirani R (1996) Discriminant analysis by gaussian mixtures. Journal of the Royal Statistical Society: Series B (Methodological) 58(1):155–176. https://doi.org/10.1111/j.2517-6161.1996.tb02073.x

Jeon M, Park D, Lee J, et al (2019) ReSimNet: drug response similarity prediction using Siamese neural networks. Bioinformatics 35(24):5249–5256. https://doi.org/10.1093/bioinformatics/btz411

Koch G, Zemel SRR. (2015) Siamese neural networks for one-shot image recognition. In: Proceedings of the 32nd International Conference on Machine Learning

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics 43:59–69. https://doi.org/10.1007/

BF00337288

Kohonen T (2013) Essential of self-organizing map. Neural Networks 37:52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Kornblith S, Norouzi M, Lee H, et al (2019) Similarity of neural network representations revisited. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 97. PMLR, pp 3519–3529

Liu W, Liu Z, Rehg JM, et al (2019) Neural similarity learning. ArXiv abs/1910.13003

van der Maaten L, Hinton G (2008) Visualizing data using t-sne. Journal of Machine Learning Research 9(86):2579–2605. URL http://jmlr.org/papers/v9/vandermaaten08a.html

Mathisen BM, Aamodt A, Bach K, et al (2020) Learning similarity measures from data. Progress in Artificial Intelligence 9:129–143. https://doi.org/10.1007/s13748-019-00201-2

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. 1802.03426

Mika S, Ratsch G, Weston J, et al (1999) Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468), pp 41–48, https://doi.org/10.1109/NNSP.1999.788121

Raducanu B, Dornaika F (2012) A supervised non-linear dimensionality reduction approach for manifold learning. Pattern Recognition 45(6):2432–2444. https://doi.org/10.1016/j.patcog.2011.12.006

Sammon J (1969) A nonlinear mapping for data structure analysis. IEEE Transactions on Computers C-18(5):401–409. https://doi.org/10.1109/T-C.1969.222678

Tharwat A, Gaber T, Ibrahim A, et al (2017) Linear discriminant analysis: A detailed tutorial. AI Communications 30(2):169–190. https://doi.org/10.3233/AIC-170729

UCI (2022) Machine learning repository. URL https://archive.ics.uci.edu/ml/index.php

Vogelstein JT, Bridgeford EW, Tang M, et al (2021) Supervised dimensionality reduction for big data. Nature Communications 12(2872). https://doi.org/10.1038/s41467-021-23102-2

Wang B, Zhu J, Pierson E, et al (2017) Visualization and analysis of single-cell rna-seq data by kernel-based similarity learning. Nature Methods 14(4). https://doi.org/10.1038/nmeth.4207

Zhang Y, Zhang Z, Qin J, et al (2018) Semi-supervised local multi-manifold isomap by linear embedding for feature extraction. Pattern Recognition 76:662 – 678. https://doi.org/10.1016/j.patcog.2017.09.043