Comparison of Mode Choice Behavior using Four Types of Artificial Neural Networks

Dongwoo Lee

Contents

Concept

- New approaches in choice modeling

  - Random utility theories for logit models (more than 50 years)

  - What is logit model (choice model)?

    → regression model where the dependent variable is categorical (classes)

    → we can regard this as **classification** problems

- **Machine learning (ML)**

  - Widely used in classification / clustering application

  - Logit model estimation = classification with ML

↓

Choice modeling is nothing but a classification problem based on probabilities

Concept

Machine Learning	
Unsupervised Learning = Clustering	Supervised Learning = Classification / Regression
▪ Looking for patterns	▪ We know what is 'right'
▪ Handles nominal / categorical, ordinal, discrete, and continuous variables	▪ Handles nominal / categorical, ordinal, discrete, and continuous variables
▪ Output is categorical (data point belongs to a cluster or not)	▪ Conceptually close to standard statistics (i.e, X and Y)
▪ Need to decide in advance the number of clusters	▪ Can be used for both classification and regression purposes
▪ Exist some validation metrics	▪ Exist validation processes

≈ Mode choice model
: lots of algorithm to solve it

Neural network is one of them

Source: Derrible (2017), Cities and Sustainable Infrastructure, Chapter 10

Concept

- Artificial Neural Networks (ANN)?

  - Machine learning techniques can be comparable with statistical modeling

    → Recently, it has been widely applied in many transportation fields (e.g. choice modeling, traffic control and operation, etc.)

    → Similarities and differences between statistical model and ANN

(McFadden, 2001, Sarles, 2009; Karlaftis, 2010)

Statistical model	Neural networks
Independent/estimated variables	Input/output
Dependent variables	Target values in training
Bias/residuals	Bias/errors
Estimation	Training, learning, adaptation, or self-organization
Estimation criteria	Cost function, Error function
Parameters	weights

! Most of literatures in transportation applies Backpropagation NN (BPNN)

Zhang and Xie (2008), Tillema et al. (2006), Cantarella and de Luca (2005), Mohammadian and Miller (2002), Nijkamp et al. (1996)

Purpose

- Check the possible ways for applying different ANN techniques to choice modeling instead of applying random utility theories such as logit models

- Compare prediction accuracy among 4 types of ANNs and CMAP mode choice model (Multinomial logit model)

- Contribute to the literature by demonstrating the use of ANN techniques

    - Methodological differences

    - Advantages and disadvantages

    - Model performances and future tasks in ANN

Data

- Travel Tracker Survey, CMAP (2007 ~ 2008)

  - 10,500 households: a complete travel diary for one or two randomly assigned dates.

  - We use the part of this dataset containing mode choice information in particular home-based shopping and others trips (reason: the lowest accuracy with logit modeling)

  - Detailed variables include:
    - trip-related variables (e.g., mode, purpose, departure time)
    - household and individual socio-demographic characteristics (e.g., age, income, employment status)
    - activity-related variables (e.g., type, duration).

  - Approximately 4,000 observations of home-based shopping trips and others were selected

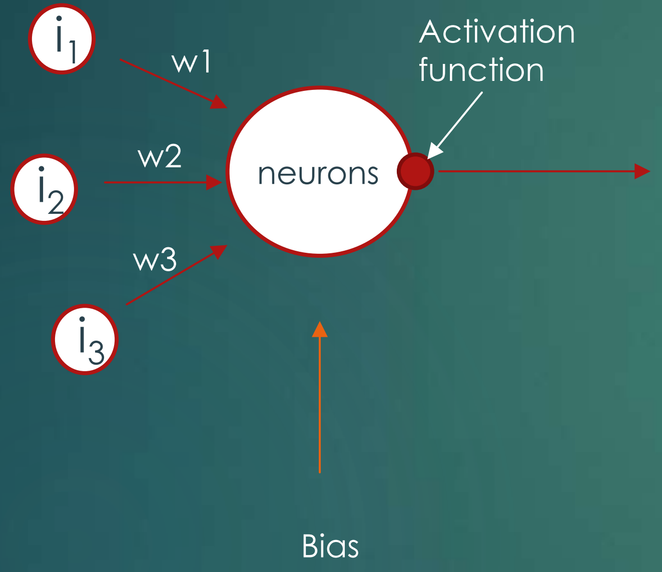
Data

- Descriptive statistics

Variable	Definition	Mean	St. dev.
Walk_TT	Travel time for walk mode (in hours)	2.45	2.89
Bike_TT	Travel time for bike mode (in hours)	0.55	0.64
Auto_TT	Travel time for auto drive mode (in hours)	0.34	0.39
Transit_TT	Travel time for transit mode (in hours)	0.34	0.33
Auto_cost	Travel cost for auto drive mode ($)	1.15	1.37
Transit_cost	Travel cost for transit mode ($)	1.80	1.59
Walk_accessible	1: if walking distance to the destination is less than 0.25 mile, 0: otherwise	0.08	0.27
Transit_egress	Egress distance to destination for transit mode (km)	1.50	3.23
Transit_access	Access distance from origin for transit mode (km)	2.38	4.20
Weekend	1: if the trip is made in weekend, 0: otherwise	0.11	0.31
HH_bikes	Number of bikes in the household	1.37	1.66
HH_size	Household size	2.70	1.36
HH_vehicle	Number of vehicles in the household	1.87	1.03
Part_work	1: if traveler works part time, 0: otherwise	0.14	0.35
Age –20	1: if traveler's age is less than 20, 0: otherwise	0.07	0.26
Age _40 – 65	1: if traveler's age is between 40 and 65, 0: otherwise	0.51	0.50
HH_car	1: if traveler's has no car	0.03	0.17
HH_bike	1: if traveler's has bike	0.58	0.49
EDU	Education level	3.68	1.87

Methodologies

- Artificial Neural Networks (ANNs)



Output = $f\,(i_{1\,w1} + i_{2\,w2} + i_{3\,w3} + \text{bias})$

- Function of the entire neural network is simply **the computation of the outputs of all the neurons**

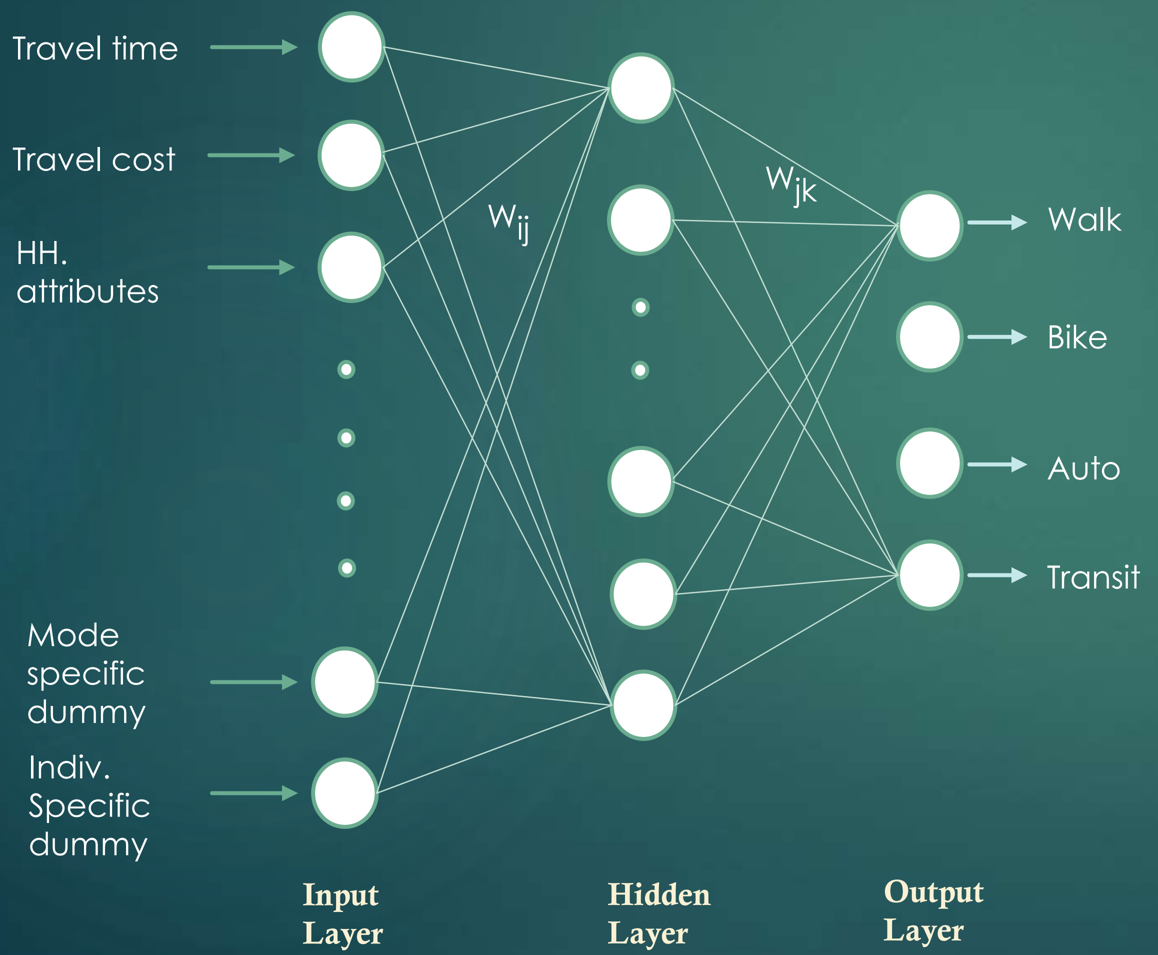
- Criteria for determining the type of neural network

  - Layers between input and output layers (e.g. hidden layers, pattern layers)
  - Learning techniques (e.g. feedforward, backpropagate , recurrent)
  - Decision criteria (e.g. Gaussian, Bayesian, min. squared error)

Methodologies

- Backpropagation NN (BPNN)

  - Adjust weights (w) by comparing and minimizing actual targets and outputs of neural networks
    : error = (target – output)^2



$$I_{j(k)} = \sum_{i=1}^{M} w_{i(k-1),j(k)} O_{i(k-1)} + \gamma$$

$$O_{j(k)} = f\left(I_{j(k)}\right)$$

□ Activation Function

Sigmoid: $f(I) = \frac{1}{1+exp(-I)}$

Hyperbolic tangent: $f(I) = tanh\, I = \frac{sinh\, I}{cosh\, I}$

Rectified linear unit: $f(I) = \begin{cases} I & for\ I \geq 0 \\ 0 & for\ I < 0 \end{cases}$

Methodologies

- Radial Basis Function NN (RBF)

  - Simplified Gaussian function when calculating the output of hidden nodes

  $$\varphi(x) = e^{-\beta\|x-\mu\|^2}$$

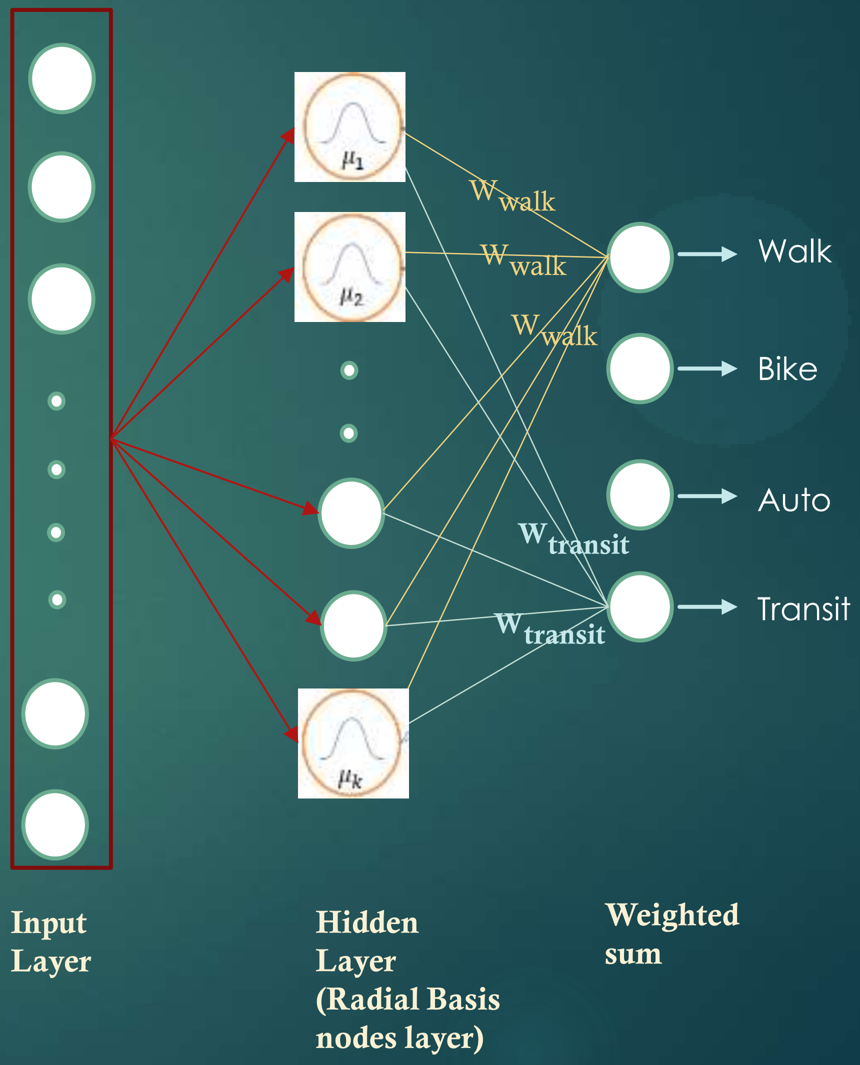
  - Beta controls the width of bell curve

- Differences between BPNN

  - Single-pass learning (no backpropagation)

  - Higher accuracy (Gaussian activation)

  - No local minima issues



Methodologies

- Probabilistic NN (PNN)

  - Input – hidden - output

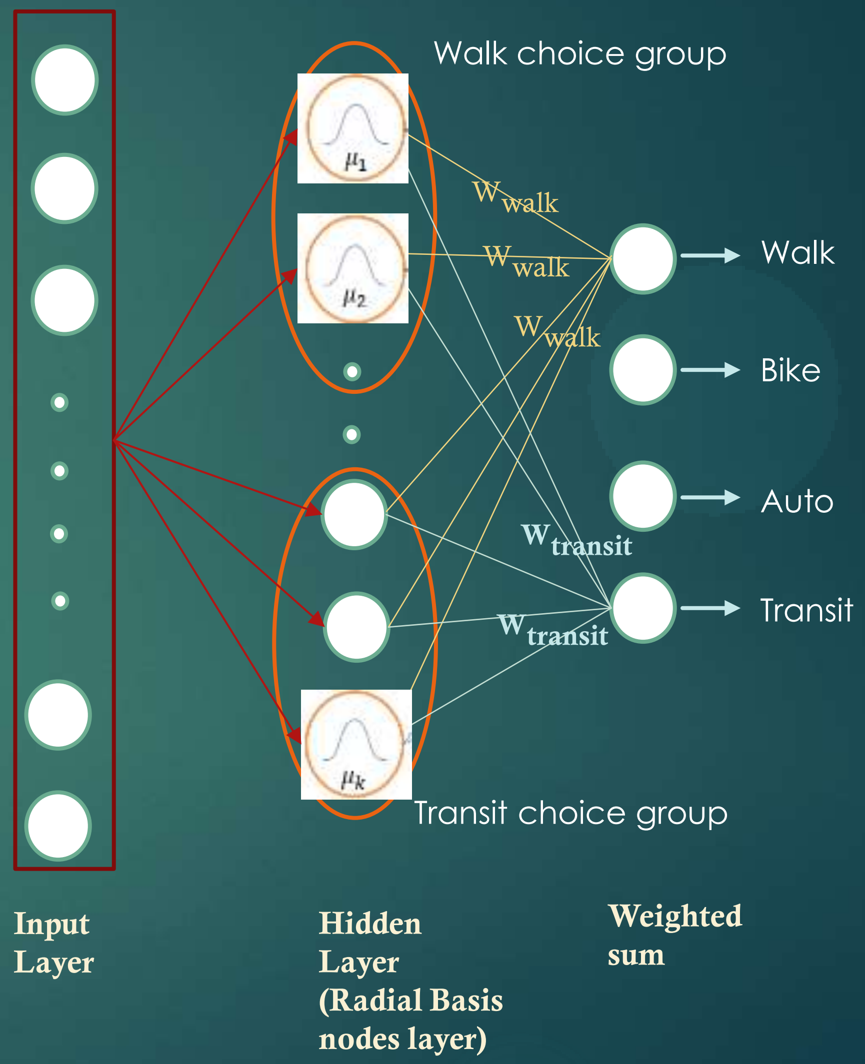
  - Hidden nodes are collected into each choice group

    → K-mean clustering (Euclidean distance)

- Differences between RBF

  - Gaussian function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Methodologies

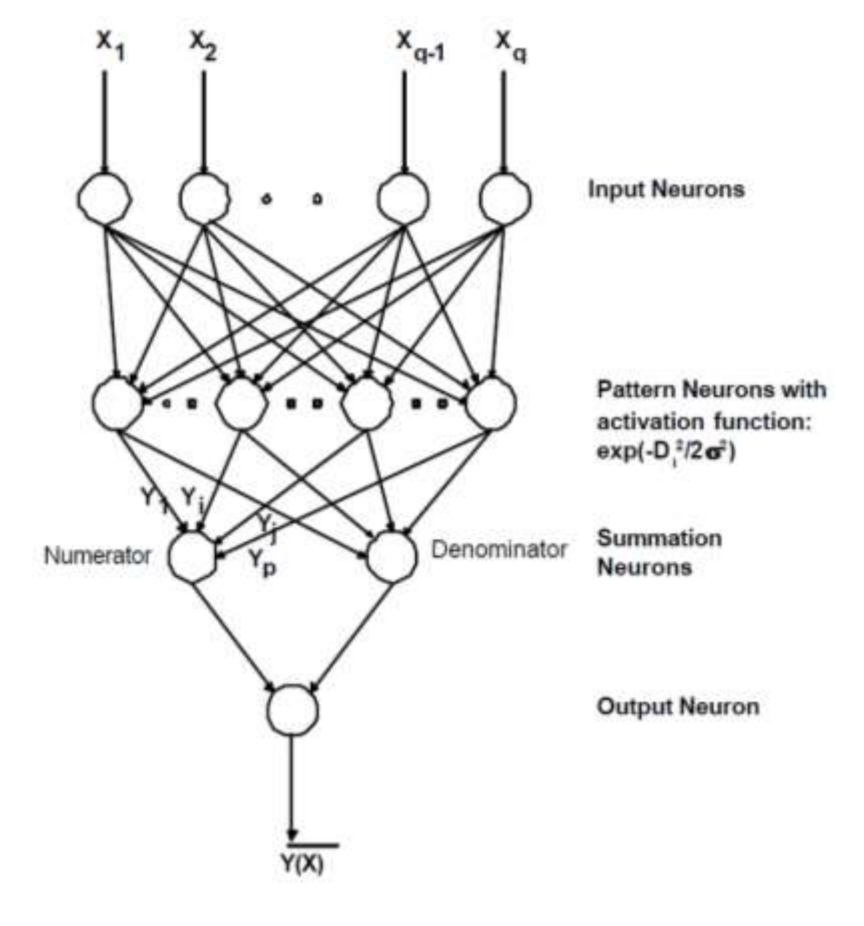
- Generalized Regression NN (GRNN)

  - Input – pattern – summation - output

  - Specific version of RBFNN for non-parametric regression and classification

  - Measures the distance among a given training case is in n-dimensional space (for n inputs)

$$Y(x) = \frac{\sum_{k=1}^{N} y_k e^{d_k/2\sigma}}{\sum_{k=1}^{N} e^{d_k/2\sigma}}, \qquad d_k = (x - x_i)^T (x - x_i)$$



Source: MM Bauer, Generalized Regression Neural Networks

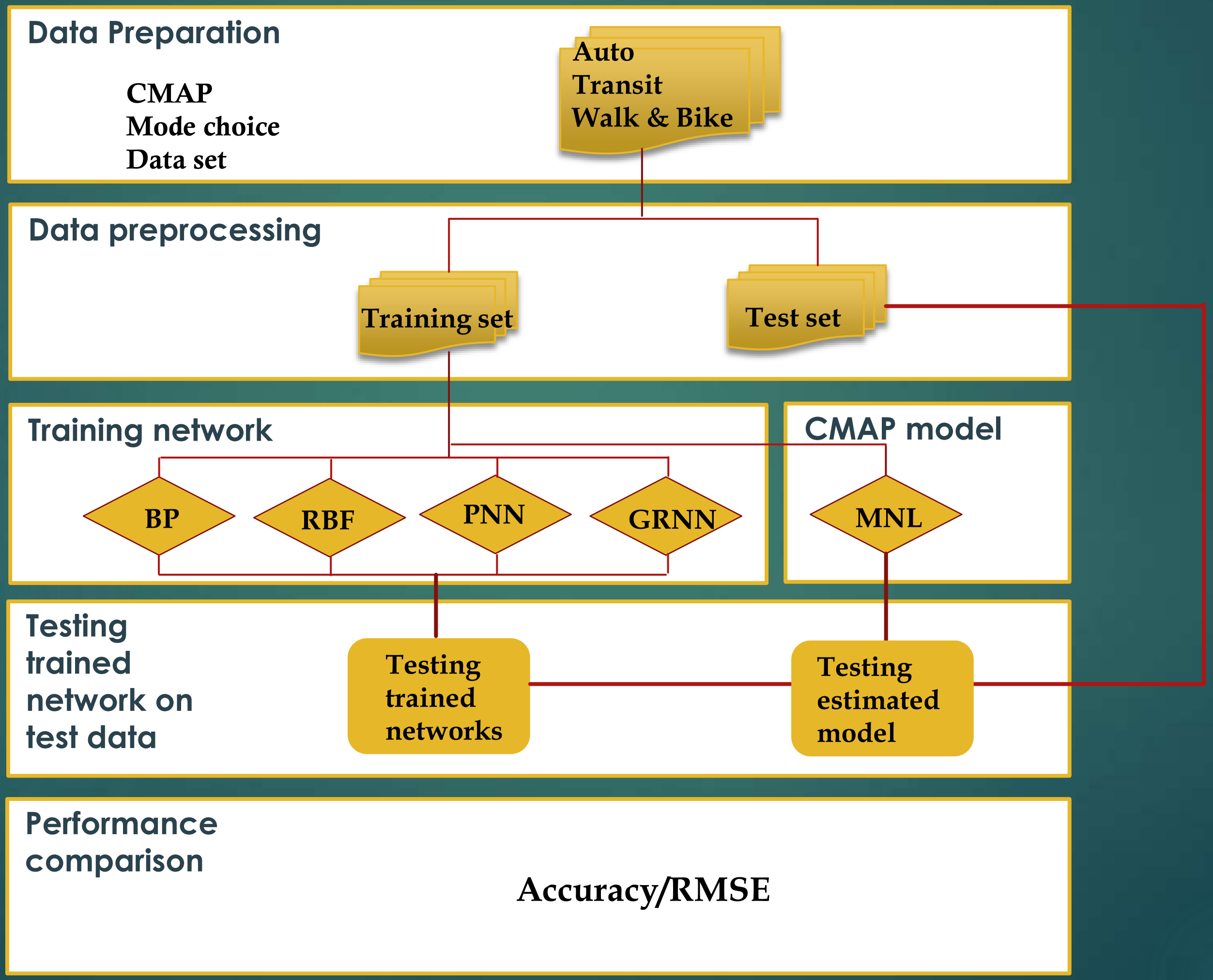
Comparison of ANNs

- Advantages and disadvantages

	BPNN	PNN	RBFNN	GRNN
Advantages	- Simple application - Does not require any statistical features in the learning process - Easy to identify the magnitude of attributes based on weights - A variety of applications are available → easy to implement	- Simpler architecture (no backpropagation) - More way to manage the algorithm by determining the shape of bell curve (specified than RBFNN) - Relatively good accuracy in classification problem	- Simpler format of Gaussian function enables to faster learning process than other Gaussian models - Radial basis function nodes can be substituted with different functional forms - Relatively performs well in both smaller and larger dataset	- Similar to RBFNN - High accuracy in the function estimation than classification
Disadvantages	- Easily get stuck in local minima resulting in suboptimal solution - Blackbox (not sure how to estimate the model) - Need sufficient observations - Overfitting problems	- Computational expensive - Limited to small and mid-sized dataset. - Saturated Gaussian function can lead some misclassification	- Difficult to determine the sigma values - Constructing network architecture is complicated.	- Long training time - No ways to improve the performance of the networks

Analysis process

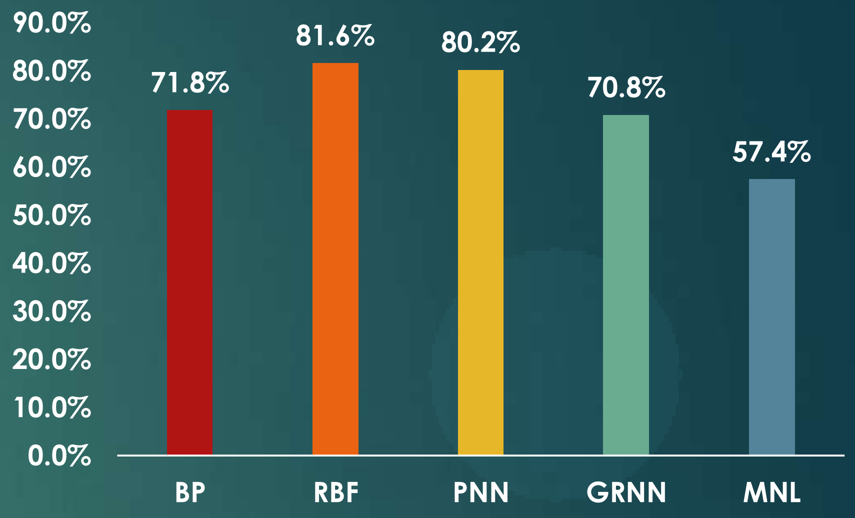
- Flow chart



Result

- Overall model accuracy

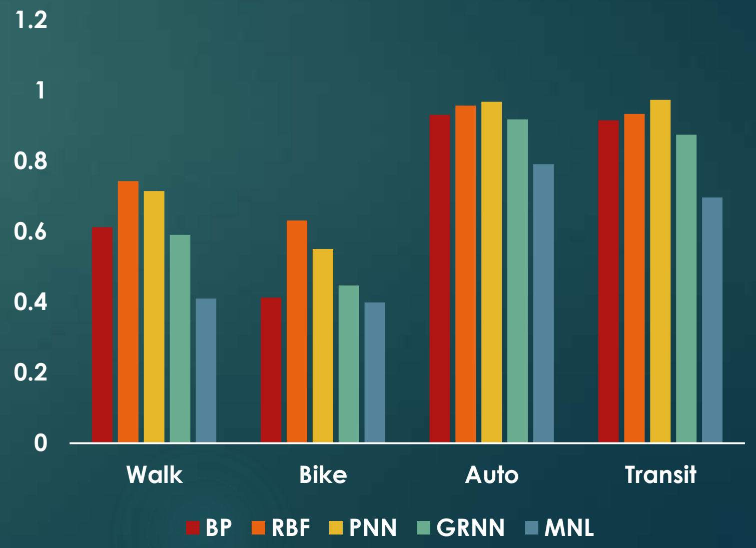
Accuracy					
	Walk	Bike	Auto	Transit	Overall
BP	0.612	0.413	0.931	0.916	71.8%
RBF	0.743	0.631	0.957	0.934	81.6%
PNN	0.715	0.551	0.968	0.974	80.2%
GRNN	0.591	0.447	0.918	0.875	70.8%
MNL	0.410	0.399	0.791	0.697	57.4%



- Accuracy: RBF > PNN > BP > GRNN> MNL (CMAP model)
  : even NN networks has higher accuracy than Copula-based model (Golshani, 2016)

- Computational cost (time): GRNN > PNN > RBF > BP
  → Gaussian function enhances computational complexity (GRNN, PNN, RBF)

- Software and packages: Neupy, Theano, Scikit-learn built in Python (ANN)



Result

- Test trained networks with the test dataset

BP

Test set		Predicted Walk	Bike	Auto	Transit	
Observed	Walk (372)	228				61.2%
	Bike (98)		41			41.3%
	Auto (1724)			1605		93.1%
	Transit (432)				395	91.6%

RBF

Test set		Predicted Walk	Bike	Auto	Transit	
Observed	Walk (372)	276				74.3%
	Bike (98)		62			63.1%
	Auto (1724)			1650		95.7%
	Transit (432)				403	93.4%

PNN

Test set		Predicted Walk	Bike	Auto	Transit	
Observed	Walk (372)	266				71.5%
	Bike (98)		54			55.1%
	Auto (1724)			1669		96.8%
	Transit (432)				420	97.4%

GRNN

Test set		Predicted Walk	Bike	Auto	Transit	
Observed	Walk (372)	220				59.1%
	Bike (98)		44			44.7%
	Auto (1724)			1583		91.8%
	Transit (432)				378	87.5%

Conclusion

- Summary

  - Applied ANN to mode choice problem (CMAP dataset)
  - BP, RBF, PNN, GRNN, and MNL are applied to address this choice problem
  - Mode choice prediction accuracy in NN is relatively higher than MNL.
  - RBF and PNN has good prediction performances than other ANNs.
  - BP is the simplest way to train the network

- Future works

  - Try different scenarios to check the performances of ANNs

    (1) Observation size (2) parameters (μ, σ) (3) different NN packages (Tensorflow, Matlab)
  - Different mode choice dataset
  - Sensitivity analysis (to test marginal changes in input factors)
  - Instead of using GRNN, try other neural networks such as convolution and recurrent NNs

Acknowledgement

This study is supported by the National Science Foundation (NSF) CAREER Award #1551731

References

Cantarella, G.E., de Luca, S., 2005. Multilayer feedforward networks for transportation mode choice analysis: an analysis and a comparison with random utility models. Transportation Research Part C 13, 121–155.

Celikoglu, H.B., 2006. Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling. Math. Comput. Model. 44, 640–658.

CMAP, Travel Model Documentation: Final Report (2010)

Dougherty, M., 1995. A review of neural networks applied to transport. Transp. Res. Part C Emerg. Technol. 3, 247–260.

Frost, F., Karri, V., 1999. Performance comparison of BP and GRNN models of the neural network paradigm using a practical industrial application. IEEE, pp. 1069–1074. doi:10.1109/ICONIP.1999.844684

Gao, W., Ma, L., Jia, Z., Ning, Y., 2010. Comparison of the GRNN and BP neural network for the prediction of populus (P.×euramericana cv.“74/76”) seedlings’ water consumption. IEEE, pp. V2-389-V2-392. doi:10.1109/ICACTE.2010.5579296

Lin, C.-T., Lee, C.S.G., 1991. Neural-network-based fuzzy logic control and decision system. IEEE Trans. Comput. 40, 1320–1336.

McFadden, J., Yang, W.T., Durrans, S., 2001. Application of artificial neural networks to predict speeds on two-lane rural highways. Transportation Research Record 1751, 9–17.

Mohammadian, A., Miller, E., 2002. Nested Logit Models and Artificial Neural Networks for Predicting Household Automobile Choices: Comparison of Performance. Transp. Res. Rec. J. Transp. Res. Board 1807, 92–100. doi:10.3141/1807-12

Rasouli, M., Nikaz, H., 2013. Trip Distribution Modelling Using Neural Network. Austalasian Transp. Res. Forum 2013.

Specht, D.F., 1990. Probabilistic neural networks. Neural Netw. 3, 109–118.

Specht, D.F., 1988. Probabilistic neural networks for classification, mapping, or associative memory. Presented at the IEEE international conference on neural networks, pp. 525–532.

Sarle, W.S., 1994. Neural networks and statistical models. In: Proceedings of the Nineteenth Annual SAS Users Group International Conference (April 1–13).

Tillema, F., van Zuilekom, K.M., van Maarseveen, M.F.A.M., 2006. Comparison of Neural Networks and Gravity Models in Trip Distribution. Comput.-Aided Civ. Infrastruct. Eng. 21, 104–119. doi:10.1111/j.1467-8667.2005.00421.x

Wasserman, P.D., 1993. Advanced methods in neural computing. John Wiley & Sons, Inc.

Xiong, Y., Schneider, J.B., 1992. Transportation network design using a cumulative genetic algorithm and neural network. Transp. Res. Rec.

Yang, H., Kitamura, R., Jovanis, P.P., Vaughn, K.M., Abdel-Aty, M.A., 1993. Exploration of route choice behavior with advanced traveler information using neural network concepts. Transportation 20, 199–223.

Yang, H.H., Murata, N., Amari, S., 1998. Statistical inference. Learning in artificial neural networks. Trends in Cognitive Sciences 2 (1), 4–10.

Zhang, Y., Xie, Y., 2006. Application of genetic neural networks to real-time intersection accident detection using acoustic signal. Transportation Research

Record 1968, 75–82.

Zhang, Y., Xie, Y., 2008. Travel mode choice modeling with support vector machines. Transportation Research Record 2076, 141–150.