The Evidence-Theoretic *k*-NN Rule for Rank-Ordered Data: Application to Predict an Individual's Source of Loan

Supanika Leurcharusmee[1], Peerapat Jatukannyaprateep[1], Songsak Sriboonchitta[1], and Thierry Denoeux[2]

[1] Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
[2] Université de Technologie de Compiègne CNRS, UNR 7253 Heudiasyc, France

Abstract. We adapted the nonparametric evidence-theoretic k-Nearest Neighbor (k-NN) rule, which was originally designed for multinomial choice data, to rank-ordered choice data. The contribution of this model is its ability to extract information from all the observed rankings to improve the prediction power for each individual's primary choice. The evidence-theoretic k-NN rule for heterogeneous rank-ordered data method can be consistently applied to complete and partial rank-ordered choice data. This model was used to predict an individual's source of loan given his or her characteristics and also identify individual characteristics that help the prediction. The results show that the prediction from the rank-ordered choice model outperforms that of the traditional multinomial choice model with only one observed choice.

Keywords: Rank-ordered Choice Data, k-Nearest Neighbor, Belief Functions, Classifier, Household Debt.

1 Introduction

For the purpose of understanding the objective and the contribution of this study, let us first clarify the distinction between the traditional *multinomial choice data* and the *rank-ordered choice data* that is of the concern here. Suppose there are M available objects. In traditional multinomial choice data, there is only one choice for each individual. In contrast, the rank-ordered choice data contains more information regarding each individual's preference as they capture each individual's ranking of the objects. If the ranks of all M objects can be observed, the data are said to be *completely* rank-ordered. If only $L < M$ ranks are observed, the data are *partially* rank-ordered. Moreover, if the number L_i of observed ranks for each individual i is different across i, then the data are called *heterogenous* rank-ordered. From these definitions, the rank-ordered choice data is reduced to the multinomial choice data when $L_i = 1$, for all i.

The main purpose of the model is to predict each individual's most preferred choice out of M available alternatives using heterogenous rank-ordered data. In

F. Cuzzolin (Ed.): BELIEF 2014, LNAI 8764, pp. 58–67, 2014.

particular, we modify the Evidence-theoretic k-Nearest Neighbor (k-NN) Rule, which was originally designed for traditional multinomial choice data [14] to take advantage of the additional information provided by rank-ordered choice data. The main idea is that the secondary or other choices also provide valuable information for the primary choice prediction.

Two main problems can be considered in relation with rank-ordered data. The first problem is to predict an individual's choices given information on choice attributes. Suppose we have a new alternative with a set of attributes, this problem is to predict the chance that this alternative will be chosen. The traditional methods to tackle this problem in economics are the rank-ordered logit model and the rank-ordered probit model. These models were later extended into the rank-ordered mixed logit model [3] [13]. The second problem is to predict an individual's choices given information on individual characteristics. Suppose we have an individual with a set of characteristics, we may wish to predict the alternative that he or she is most likely to choose. There is no logit/probit-based model designed to solve this problem. The closest methods are those developed to explain how each individual chooses a bundle of products. As discussed in Bhat, Srinivasan and Sen (2006) [5], commonly used models are the traditional multinomial probit/logit models with composite alternatives and the multivariate probit/logit models [2] [4]. Although both models allow each individual to choose more than one alternative, all the choices are equally weighted. Moreover, none of these models is appropriate for problems with a large choice set. Since there is no standard methodology for the second problem, a contribution of this study is to develop a methodology to fill this gap.

Traditional methods to analyze multiple choice problems in economics are mostly of the logit/probit types and based on maximum likelihood (ML) method. In contrast, the k-NN method is intuitively simple and requires fewer assumptions. Formally, k-NN is a classification method that can be used to predict an individual's choice based on information from the observed choices of the k neighbors with the closest characteristics. An advantage of the k-NN model being nonparametric is that it does not require distributional assumptions like the ML method. Moreover, since the method only uses the k nearest neighbors for prediction, it is robust to outliers. It is also flexible in the sense that it can be applied consistently for complete, partially and heterogeneous rank-ordered data. With a set of restrictions, the method boils down to the traditional evidence-theoretic k-NN rule.

The Application of the evidence-theoretic k-NN rule model relies on the availability of ranking data. The most obvious applications concern consumer choice models, in which each customer buys more than one product or one brand. For the empirical application considered in this study, the model was used to analyze each individual's choices of loan sources. The main objective of this exercise is to predict where each individual borrows from, given his or her characteristics.

This paper is organized as follows. Section 2 recalls the original evidence-theoretic k-NN rule for multinomial choice data. Section 3 introduces the evidence-theoretic k-NN rule for heterogeneous rank-ordered data and discusses how the method can be applied to completely and partially rank-ordered data. Section 4 provides an empirical example by applying the method to predict an individual's primary source of borrowing. Finally, Section 5 presents our conclusions and remarks.

2 The Evidence-Theoretic *k*-NN Rule

The original Evidence-theoretic k-NN Rule is a method to classify each individual into M classes based on his attributes [14] [7]. The model can thus be applied for the multiple choice problem using multinomial choice data. Let the set of alternatives be $\Omega = \{\omega_1, \omega_2, ..., \omega_M\}$. For each individual i, we observe information $(x^{(i)}, \omega^{(i)})$, where $x^{(i)}$ is the vector for individual i's attributes and $\omega^{(i)}$ is the alternative that individual i has chosen. That is, $(x^{(i)}, \omega^{(i)})$ constitutes an evidence for the class membership of x. The mass function for each individual i is

$$\begin{aligned} m^{(i)}(\omega^{(i)}) &= \alpha\phi(d^{(i)}) \\ m^{(i)}(\Omega) &= 1 - \alpha\phi(d^{(i)}), \end{aligned} \tag{1}$$

where $d^{(i)}$ is the distance between x and $x^{(i)}$, ϕ is the inverse distance-normalization function that maps the distance $d^{(i)}$ from $[0, +\infty)$ to $[0, 1]$ and α is a parameter in $[0, 1]$.

Information from each individual is considered as evidence. For independent and identically distributed (iid) data, we can combine all the pieces of evidence from k nearest neighbors using Dempster's rule. The combined mass function for each choice $\{\omega_q\}$ is

$$\begin{aligned} m(\{\omega_q\}) &= \frac{1}{K}(1 - \prod_{i \in I_{k,q}} (1 - \alpha\phi(d^{(i)}))) \prod_{r \neq q} \prod_{i \in I_{k,r}} (1 - \alpha\phi(d^{(i)}) \\ m(\Omega) &= \frac{1}{K} \prod_{r=1}^{M} \prod_{i \in I_{k,r}} (1 - \alpha\phi(d^{(i)})), \end{aligned} \tag{2}$$

where $I_{k,q}$ is the set of the k nearest neighbors that chose alternative q and K is the normalizing factor.

3 The Evidence-Theoretic *k*-NN Rule for Heterogeneous Rank-Ordered Data

Consider a general model for heterogenous rank-ordered choice data with M available alternatives, $L_i \leq M$ of which are ranked for each individual i. The objective of this model is to predict the choice of an individual given his or her T characteristics x. Therefore, we construct a model using the data $(\omega^{(i)}, x^{(i)})$ from k individual i whose characteristics $x^{(i)}$ are the closest to x. Each of the

k individuals ranks L_i objects, providing L_i pieces of evidence for his or her preferences through the mass functions. For each individual i, we can observe the L_i most preferred choices $\omega^{(i)} = \{\omega^{(i1)}, ..., \omega^{(iL_i)}\}$, where $\omega^{(i1)}$ is the most preferred choice $\omega^{(iL_i)}$ is the L_i^{th} choice. The mass function for individual i can be defined as

$$\begin{aligned} m^{(i)}(\{\omega^{(i1)}\}) &= \alpha_1 \phi(d^{(i)}) \\ m^{(i)}(\{\omega^{(ij)}\}) &= \begin{cases} \alpha_j \phi(d^{(i)}), \text{ if } j \leq L_i \\ 0, \qquad\qquad \text{otherwise} \end{cases} \\ m^{(i)}(\Omega) &= 1 - \sum_{j=1}^{L_i} \alpha_j \phi(d^{(i)}). \end{aligned} \tag{3}$$

where $d^{(i)} = (x - x^{(i)})' \Sigma (x - x^{(i)})$ is the weighted squared Euclidean distance between x and $x^{(i)}$ with a $T \times T$ diagonal matrix $\Sigma = diag(\sigma_1, \ldots, \sigma_Y)$ with and $\phi(d^{(i)}) = \exp(-\gamma {d^{(i)}}^2)$ is the inverse distance-normalization function.

The mass function (3) satisfies the basic probability assignment (BPA) properties, which are $m(\emptyset) = 0$ and $\sum_{A \in 2^\Omega} m(A) = 1$. That is, the mass $m^{(i)}(\{\omega_q\})$ captures the proportion of all relevant and available evidence from individual i that supports the claim that an individual with characteristics x will choose alternative q. From Equations (3), each mass depends on two factors, which are 1) the distance between $x^{(i)}$ and x and 2) the rank of the alternative.

The parameters to be estimated are $\theta = \{\alpha_1, ..., \alpha_M, \sigma_1, ..., \sigma_T, \gamma\}$. Parameters $0 \leq \alpha_j \leq 1, j = 1, ..., M$ capture different weights for the mass functions of objects with different ranks. Since the higher ranked objects should have higher weights, $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_L$. Parameters $0 \leq \sigma_t \leq 1, t = 1, ..., T$ capture different weights for each characteristic of individual i in the vector $x^{(i)}$. A characteristic that is more important as a determinant of the choice selection should have a higher weight. Lastly, the parameter γ is a positive scale parameter for the inverse distance-normalization function.

In the belief function framework, the belief on a claim can be represented as a belief-plausibility interval. The belief function measures the extent to which the evidence implies the claim and is defined as $Bel(A) = \sum_{B \in A} m(B)$. The plausibility function measures to what extent the evidence does not contradict the claim; it is defined as $Pl(A) = \sum_{B \cap A \neq \phi} m(B)$. Here, the belief and plausibility of each alternative q from individual i are $Bel^{(i)}(\{\omega_q\}) = m^{(i)}(\{\omega_q\})$ and $Pl^{(i)}(\{\omega_q\}) = m^{(i)}(\{\omega_q\}) + m^{(i)}(\Omega)$, respectively.

The plausibility of each alternative q for individual i can thus be written as

$$Pl^{(i)}(\{\omega_q\}) = 1 - \sum_{j=1}^{L_i} (\alpha_j \phi(d^{(i)}))^{(1 - y_{jq}^{(i)})}, \; \forall q = 1, ..., M, \tag{4}$$

where $y_{jq}^{(i)} = \begin{cases} 1 \text{ if } j^{th} \text{ choice of individual } i \text{ is } \omega_q \\ 0 \text{ otherwise} \end{cases}.$

When all observations are independent and identically distributed (iid), all pieces of evidence from the k nearest neighbors can be combined using Dempster's rule. Cobb and Shenoy (2006) proposed the plausibility transformation method to convert Dempster-Shafer belief function models to Bayesian probability models that are consistent with Dempster's rule. The *plausibility probability function* is the normalized form of the combined plausibility function $Pl(\{\omega_q\})$ [6]. Furthermore, the plausibility of each singleton after combination by Dempter's rule is the product of the plausibilities from each piece of evidence. Therefore, the plausibility probability function is

$$Pl\_P_m(\{\omega_q\}) = K^{-1}Pl(\{\omega_q\}) = K^{-1}\prod_{i=1}^{k}\left[1-\sum_{j=1}^{L_i}(\alpha_j\phi(d^{(i)}))^{(1-y_{jq}^{(i)})}\right], \quad (5)$$

where $K = \sum_{r=1}^{M} Pl(\{\omega_r\})$ is the normalization constant.

Estimation: The vector of parameters $\theta = \{\alpha_1, ..., \alpha_M, \sigma_1, ..., \sigma_T, \gamma\}$ can be estimated by minimizing the mean squared error (MSE)[1]. To compute the MSE, we estimate $Pl\_P_M^{(i)}$ for each observation i given its characteristics $x^{(i)}$. Let $t_q^{(i)}$ be a vector representing the observed choice of individual i where the chosen element q equals to 1 and other elements equal to 0. The MSE is

$$MSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{q=1}^{M}(Pl\_P_m^{(i)}(\{\omega_q\}) - t_q^{(i)})^2. \quad (6)$$

The procedure is repeated for all possible k to find the optimal value of k that minimizes the MSE. For the prediction rule, the predicted choice of individual i is the choice with the highest plausibility probability $Pl\_P_M^{(i)}(\{\omega_q\})$.

Special Cases: The evidence-theoretic k-NN model for heterogenous rank-ordered data can be consistently applied to partial and complete rank-ordered data. In particular, when $L_i = L < M$ for all i, the partial rank-ordered model is recovered. When $L_i = M$ for all i, we have the complete rank-ordered model. Moreover, the model is consistent with the original k-NN model for the traditional multinomial choice data. That is, when $L = 1$, we get the traditional multinomial choice model with only one observed choice.

Variations in Model Specification: The k-NN method for the rank-ordered choice data can be modified to capture several aspects of heterogeneity in the data. In particular, each parameter in $\theta = \{\alpha_1, ..., \alpha_M, \sigma_1, ..., \sigma_T, \gamma\}$ can be modified to be alternative-specific. For instance, the scale parameter γ can be generalized to γ_q in order to capture the different chance of occurring of each alternative. The σ_t can also be generalized to σ_{qt} to capture the different contribution of each

[1] We used the fmincon procedure in Matlab, with the active set algorithm.

characteristic t in predicting each alternative q. It should be noted that adding more parameters allows the model to capture more characteristics of the data. However, it also causes inefficiency especially for studies with a small number of observations.

4 Predicting an Individual's Source of Loan

In this section, we report on the application of the above method to predict an individual's primary source of loan given his set of characteristics x.

The data used in this study are the 2010 cross-sectional data from the Panel Household Socio-Economic Survey (SES) conducted by the National Statistical Office of Thailand. The dependent variable is the source of loan. In the SES, each individual was provided with eight choices of loan sources and asked whether he had borrowed any money in the past year. If the individual had borrowed, the survey asked for his or her two largest sources of loan in order.

In this study, we performed and compared four types of Evidence-theoretic k-NN rule models including the multinomial model with equal weights (MEW), the rank-ordered model with equal weights (REW), the multinomial model with optimized weights (MOW) and the rank-ordered model with optimized weights (ROW).

Multinomial models use the information only from the primary choice to estimate the vector of parameters θ. Rank-ordered choice models use the information from both primary and secondary choices. Formally, multinomial choice models are rank-ordered models with $\alpha_2 = 0$. The equal weight assumption restricts all the weight $\sigma_t = 1$, for all t. This restriction implies that all characteristics in x have an equal contribution to the loan choice prediction. Optimizing the weights allows the weights to vary across characteristics. Therefore, the prediction using the optimized weight models relies more on the characteristics with higher weights. That is, equal weight models are optimized weight models with $\sigma_t = 1$, for all t. It should be noted each characteristic in x was normalized so that the weights σ_t can be compared across t. In addition, in this study, we allowed the scale parameter γ in the inverse distance-normalization function $\phi(\cdot)$ to vary with each individual's primary choice. Specifically, γ_q is the scale parameter for each individual with $\omega^{(i1)} = q$. The estimated parameters for each of the four models are reported in Table 1.

In Table 1, consider the models REW and ROW. The parameter $\alpha_2 \neq 0$ indicates that including the information from the secondary loan choice helps the model to predict the primary choice more accurately. Consider the models MOW and ROW. The parameters $\sigma_t \neq 1$, for all t show that characteristics in x are not equally important for the loan source prediction. The characteristics with highest contribution to the prediction accuracy include *total saving, college, total income* and *urban.*

To ensure that the parameters in Table 1 minimize the MSE, it is necessary to check that the MSE function is smooth and convex with respect to all parameters. Fig. 1 shows the MSE contour plot for parameters α_1 and α_2 for the ROW model.

Table 1. Comparison of the four models

	MEW	REW	MOW	ROW
Alphas - α_j				
Primary choice	0.12	0.11	0.12	0.11
Secondary choice	0.00	0.03	0.00	0.04
Gammas - γ_q				
Commercial bank	0.01	0.02	0.00	0.00
BAAC	0.00	0.00	0.00	0.00
GHB	0.00	0.00	0.01	0.01
Village Fund	0.04	0.04	0.35	0.67
Co-ops/Credit Union	0.00	0.00	0.01	0.01
Other financial inst.	0.04	0.04	0.25	0.25
Friend/relative	0.03	0.02	0.50	0.54
Other source	0.81	1.00	1.00	1.00
Weights - σ_t				
Age	1.00	1.00	0.01	0.12
Total income	1.00	1.00	0.52	0.43
Total saving	1.00	1.00	1.00	1.00
Female	1.00	1.00	0.01	0.00
Urban	1.00	1.00	0.32	0.22
College	1.00	1.00	0.97	0.86
Employed	1.00	1.00	0.00	0.00
Agricultural household	1.00	1.00	0.21	0.02
House owner	1.00	1.00	0.00	0.00

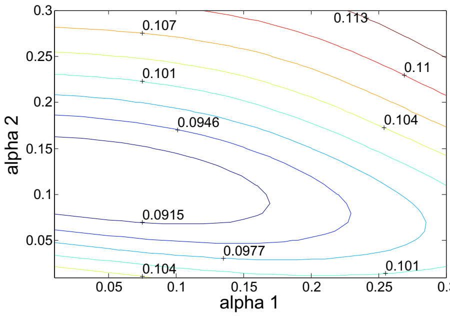

Fig. 1. MSE contour plot for parameter α_1 and α_2 for the ROW model

The optimal number of neighbour $k = 37$ for all k-NN models. It should be noted that k was endogenously determined in the model and the number needs not to be the same across models. However, changes of the MSE with respect to k have the same pattern across the four models in this study as shown in Fig. 2.

Table 2 compares the performances of the four k-NN models with the multinomial logit (MLogit) model, which is commonly used for choice prediction. The performance comparison statistics used in this study are the out-sample MSE

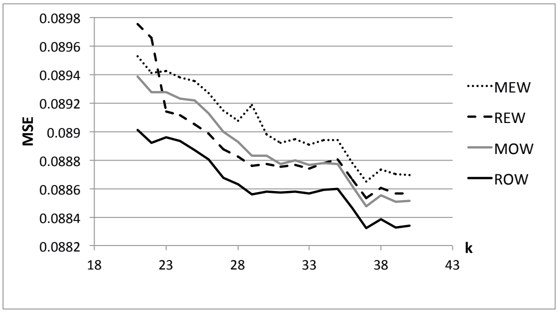

Fig. 2. MSE of the MEW, REW, MOW and ROW models for different values of k

Table 2. 5-fold cross validation results

	MLogit	MEW	REW	MOW	ROW
MSE	0.0958	0.0918	0.0918	0.0916	0.0916
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0022)
Classification error	0.6263	0.6168	0.6078	0.6134	0.6048
	(0.0232)	(0.0232)	(0.0232)	(0.0232)	(0.0232)
McNemar's χ^2 stat	10.26	2.57	2.12	1.30	-
	(0.0013)	(0.1089)	(0.1454)	(0.2542)	

*For MSE and classification error, standard deviations in parentheses.
**For McNemar's test, p-values in parentheses.

and classification errors estimated using 5-fold cross validation [11]. The rank-ordered models have smaller MSE than the multinomial models and the optimized weight models have smaller MSE than the equal weight models. Moreover, the results also show that models with a smaller MSE also yield a lower classification error. Using the McNemar's test [1] to compare all models with the ROW model, we can see that the ROW model has significantly higher prediction power than the multinomial logit model. However, the McNemar's test did not give significant results for other cases. In this dataset, only 20.68% of the data borrowed from the second source, which can explain why the performance improvement from using the rank-ordered model was rather small. It can be expected that more information on the non-primary ranks would increase the performance improvement.

5 Conclusions

The evidence-theoretic k-NN rank-ordered choice model was shown to outperform the traditional multinomial choice model, which shows the benefit of including the additional information from each individual's non-primary choices. The weight matrix contributes significantly to the prediction accuracy, indicating that all the characteristics are not equally informative.

Despite the non-parametric nature of the model, a number of assumptions were made. It is important to discuss a few alternatives for the model specification as it may improve the model performance for different studies. The first assumption of this model is related to the distance function $d(\cdot)$. The second assumption is on the confidence measure. This study uses the plausibility probability function $Pl\_P_m(\cdot)$. Alternatives are the belief, the plausibility and the pignistic functions [6]. The last assumption is on the optimization rule. The optimization rule used in this study is the mean squared error (MSE) minimization. Alternative criteria are entropy or the modified MSE as suggested in Denoeux and Zouhal (2001) [9].

The evidence-theoretic k-NN model has been extended in several aspects, many of which can be applied to this rank-ordered choice model. An advantage of using the evidence-theoretic method is that it can be modified to cope with uncertain and imprecise data in which a set of alternatives is observed for each rank. For example, if we can only observe that individual i dislikes choice ω_q, then we know that all other available choices are preferred to ω_q but we do not know the ranking of those choices. In this case, the first rank would contain more than one alternative and the model can take advantage of the evidence theoretic method more fully.

Moreover, the belief function approach makes it possible to combine pieces of evidence from several different sources. Therefore, the output from the belief function classifier can conveniently be combined with evidence from other classifiers or with other information such as expert opinions.

References

1. Alpaydin, E.: Introduction to Machine Learning. MIT Press (2004)
2. Baltas, G.: A Model for Multiple Brand Choice. European Journal of Operational Research 154(1), 144–149 (2004)
3. Beggs, S., Cardell, S., Hausman, J.: Assessing the Potential Demand for Electric Cars. Journal of Econometrics 17(1), 1–19 (1981)
4. Bhat, C.R., Srinivasan, S.: A Multidimensional Mixed Ordered-response Model for Analyzing Weekend Activity Participation. Transportation Research Part B: Methodological 39(3), 255–278 (2005)
5. Bhat, C.R., Srinivasan, S., Sen, S.: A Joint Model for the Perfect and Imperfect Substitute Goods Case: Application to Activity Time-use Decisions. Transportation Research Part B: Methodological 40(10), 827–850 (2006)
6. Cobb, B.R., Shenoy, P.P.: On the Plausibility Transformation Method for Translating Belief Function Models to Probability Models. International Journal of Approximate Reasoning 41(3), 314–330 (2006)
7. Denoeux, T.: A k-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 25(5), 804–813 (1995)
8. Denoeux, T.: Analysis of Evidence-Theoretic Decision Rules for Pattern Classification. Pattern recognition 30(7), 1095–1107 (1997)
9. Denoeux, T., Zouhal, L.M.: Handling Possibilistic Labels in Pattern Classification Using Evidential Reasoning. Fuzzy Sets and Systems 122(3), 409–424 (2001)
10. Maddala, G.S.: Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press (1986)

11. Mosteller, F., Tukey, J.W.: Data Analysis, Including Statistics. The Collected Works of John W. Tukey: Graphics 123(5), 1965–1985 (1988)
12. The Mathworks, Optimization Toolbox: User's Guide (R2014b). The MathWorks, Inc. (2014)
13. Train, K.: Data analysis, Including Statistics Discrete Choice Methods with Simulation. Cambridge University Press (2009)
14. Zouhal, L.M., Denoeux, T.: An Evidence-theoretic k-NN Rule with Parameter Optimization. Systems, Man, and Cybernetics, Part C: Applications and Reviews, 28(2), 263-271 (1998)