Trinomial Modeling in One Binary Logit Regression

Stan Lipovetsky
GfK North America, 8401 Golden Valley Rd, Minneapolis, MN 55427

Abstract
A new approach of constructing a triple-category ordinal outcome model in one binary logistic regression is presented. Various applied problems are formulated with a dependent variable of three ordinal categorical levels, for instance, positive-neutral-negative segments of meaning. It is commonly considered in a multinomial model for a categorical variable of three possible outcomes. This work shows that the problem can be reduced to a much more simple and convenient binomial logit model. It can be done in the approach developed in the area of marketing research and known in terms of Best-Worst scaling or MaxDiff modeling. In this approach the positive-neutral data subset is stacked with the negative-neutral subset. In the latter one the predictor signs are changed to opposite. The binary dependent variable is kept equal one for both positive-negative outcomes and equals zero for neutral outcomes, respectively. In the constructed logit regression the positive category predictions are close to 1, negative close to 0, and neutral are in the middle of its continuous 0-1 scale. Theoretical features and practical application of the model are discussed and a numerical example is given.

Key Words: MNL, trinomial model, MaxDiff, binary logit regression.

1. Introduction

In many applied problems the response variable can be presented in three categorical levels. For instance, in marketing research, a dependent variable of overall satisfaction is commonly measured in the ordinal Likert scale, say, from 1 as the worst to 10 as the best value. Let several upper levels of it correspond to the satisfaction, and some bottom levels to the dissatisfaction. Then we have a scale of satisfaction-neutral-dissatisfaction or positive-neutral-negative segments. Modeling for such a scale can be performed in different regression and non-regression approaches (Conklin et al., 2004; Lipovetsky and Conklin, 2005; Lipovetsky, 2012). Double sigmoid functions with three levels are described in (Madhavan et al., 1995; Lipovetsky, 2010). Outcome of three levels can be considered in a general multinomial model setup, but it requires special software for modeling. It is interesting to note that division of data to three groups has been studied for linear pair regression modeling in many works, for instance, see (Leser, 1971; Gelman and Park, 2009, and references within). The last of these papers considers also the discrete outcomes. Triple outcome can be convenient for election modeling with undecided voters who can switch from neutral to one of the margins.

The current work shows that the problem of three level modeling can be reduced to a much simpler and convenient binomial logit model. It can be done in the approach developed in the area of marketing research and known in terms of Best-Worst scaling or MaxDiff modeling. It is a contemporary method for the prioritization of items proposed by Jordan Louviere (1991, 1993), and developed and applied in numerous works (for instance, Marley and Louviere, 2005; Marley et al., 2008; Bacon et al., 2007, 2008). In MaxDiff approach the positive-neutral data subset is stacked with the negative-neutral subset in which the signs of predictors are changed to opposite. The binary dependent variable equals one for positive-negative outcomes and zero for neutral outcomes, respectively. Logit regression model is constructed, where the positive category predictions are close to 1, negative close to 0, and neutral in the middle of its continuous 0-1 scale. More detail on MaxDiff data see, for instance in (Lipovetsky and Conklin, 2014 a, b; Lipovetsky et al., 2015).

2. MaxDiff and Trinomial Modeling

Consider an example of a dataset with 3062 respondents and Max-Diff exercise to prioritize 17 items. Respondents went through 10 tasks each where, in each task, they chose the best and worst item from a set of 4 of the 17 items. Design was balanced so that each item was seen an average of 2.35 times by each respondent. There were three different versions of the design so that for the overall sample the number of exposures of each item and each pair of items was balanced. In MaxDiff data pre-processing, we stack the data by task so that we have a final dataset with 3062*10 = 30,620 rows. Each row contains information on which items were shown in the task as well as which item was "best" and which was "worst".A general multinomial logit (MNL) defines choice among several outcomes and can be described by the probability of choice model:

$$p_k = \frac{\exp\left(a_1^{(k)}x_1 + ... + a_n^{(k)}x_n\right)}{\exp\left(a_1^{(1)}x_1 + ... + a_n^{(1)}x_n\right) + ... + \exp\left(a_1^{(m)}x_1 + ... + a_n^{(m)}x_n\right)}, \qquad (1)$$

where x_j are predictor variables, the parameters a_k define the probability of each k-th choice among all m of them (m=3 in trinomial outcome). For the sake of identification, one share's parameters (1) are taken as reference, or put to zero. Finding parameters of MNL is a complex numerical problem of nonlinear estimation.

For the binomial dependent variable the common tool for modeling is logistic regression

$$p = \frac{\exp\left(a_1x_1 + ... + a_nx_n\right)}{1 + \exp\left(a_1x_1 + ... + a_nx_n\right)} , \qquad (2)$$

where p is the binary outcome, x_j are predictor variables, and a_k are parameters estimated by data. In discrete choice modeling (DCM), for instance, with MaxDiff data, the binary outcome of the Best one versus non-best items (1 vs. 0) can be modeled by (2). After finding parameters, the choice probabilities can be estimated by (2) as the continuous values in the 0-1 interval.

Choice of the Worst item in MaxDiff can be considered in a similar DCM model (2). For a simultaneous estimation by all best and worst choices in one combined dataset the following property is applied: if to change signs of all predictor variables then the probability estimated by the logit model (2) equals 1-p which defines the absence of a binary event. Indeed, consider the transformation of sign change:

$$\frac{\exp\left(-(a_1x_1 + ... + a_nx_n)\right)}{1 + \exp\left(-(a_1x_1 + ... + a_nx_n)\right)} = 1 - \frac{\exp\left(a_1x_1 + ... + a_nx_n\right)}{1 + \exp\left(a_1x_1 + ... + a_nx_n\right)} = 1 - p. \qquad (3)$$

The design matrix for modeling of the worst item can contain all predictors with opposite sign, and the binary output is defined as the Worst versus non-worst items (also 1 vs. 0).

In practical MaxDiff modeling, two DCM design matrices with the rows defined by Best-Neutral segment and by Worst-Neutral segment (with opposite signs of predictors) are combined into one total matrix of choices (Louviere et al., 2008).

This approach can be applied to any data with trinomial positive-neutral-negative segments. In the combined matrix, the positive segment has the original values of predictors and the value 1 in the binary outcome, the negative segment has the opposite sign of the original values of predictors and the value 1 in the binary outcome, and the neutral segment has doubled rows of original neutral segment and the same with opposite signs of predictors and 0 value of the binary outcome.

For more explicit presentation, let us express the positive-neutral segment and the negative-neutral segment of data as:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} X_{positive} \\ X_{neutral} \end{pmatrix}, \qquad \begin{pmatrix} -1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} X_{negative} \\ X_{neutral} \end{pmatrix}. \tag{4}$$

Then total stacked segments for the binary logit model can be written as follows:

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} X_{positive} \\ X_{neutral} \\ -X_{negative} \\ -X_{neutral} \end{pmatrix}. \tag{5}$$

Such a stacked data can be modeled in one logistic regression (2) which defines choice of positive or negative outcomes versus non-chosen cases. With this design (5) the positive and negative values of the binary predictors push the outcomes with the values 1 to the sides of maximum and minimum (2) probability, respectively, while zero values tend to the middle part of the logit curve. Thus, a trinomial outcome can be considered via binary logit model. It is also useful to mention that logit models can be constructed in analytical closed-form solution as described in (Lipovetsky, 2014).

3. Numerical Example

A data from a marketing research project is taken with about nineteen hundred observations, two dozen predictors, and a dependent variable transformed by some thresholds to the trinomial outcome (within it 320 negative, 926 neutral, and 681 positive cases). The outcome can be seen as an ordinal categorical variable, so we can try to apply ordinary least squares. Another approach consists in using three regular binary logit models of each outcome versus the other values. Obtaining three logit models (2) and predictions by them, we additionally normalize the predicted values within each response to one. The third technique corresponds to direct modeling of the MNL regression (1). And the fourth approach is the newly suggested trinomial outcome considered via MaxDiff data transformation and binary logit modeling.

Having one scale of logistic probability for all three classes we can identify the best threshold values to classify observations using Receiver Operating Characteristic, or ROC curve. An example of it is shown in Fig.1. With sliding thresholds we identify the cut-off for neutral class, and the other two classes, respectively.

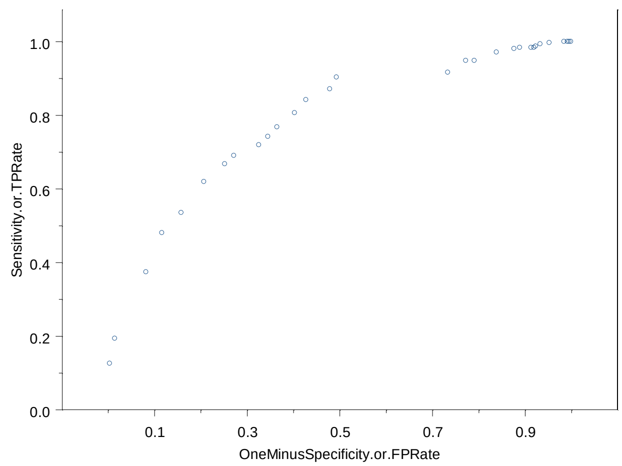

Figure 1: ROC for Negative class: Sensitivity vs. Specificity, or True vs. False Positive Rate.

Table 1 presents results of the described four approaches in cross-tables of the observed and predicted three classes. The bottom row in Table 1 shows the hit rate, or proportion of the total correct prediction on diagonal of cross-tables to the total number of observations. We see that the hit rate is very similar by all models, with a slightly better value for MNL which uses twice more parameters for data fitting. Interesting to note that the trinomial binary model yields the best prediction of the negative values of the smallest count in the data.

Table 1: Prediction of trinomial outcome by several models.

	linear model			logit models			MNL model			MaxDiff logit		
class	neg	neut	pos	neg	neut	pos	neg	neut	pos	neg	neut	pos
neg	152	161	7	99	210	11	155	153	12	166	147	7
neut	124	615	187	51	693	182	79	675	172	123	620	183
pos	4	109	568	6	119	556	5	130	546	4	116	561
hit rate			0.693			0.699			0.714			0.699

In opposite cases, of RR<1 or t-statistics negative, we have B as the inhibitors for the choice of the product A. More analysis can be performed on this data. We can create a heat map to identify key pairs of items that have high synergy. Finding hotspots (white) where t-values are high is easy, however, note that the matrix is not symmetric, thus, the items may not be mutually synergistic.

4. Summary

The work describes a convenient approach to modeling a trinomial ordinal categorical outcome via binary logit regression. Theoretical features and practical application of this model are discussed. The described technique is presented in more detail in (Lipovetsky, 2015) and can be useful in various problems and help researchers in practical data fit and analysis.

References

Conklin, M., K. Powaga, and S. Lipovetsky. 2004. Customer Satisfaction Analysis: Identification of Key Drivers, European J. of Operational Research, 154, 819-827.

Gelman, A., and D.K. Park. 2009. Splitting a Predictor at the Upper Quarter or Third and the Lower Quarter or Third, The American Statistician, 63, 1, 1-8.

Leser, C.E.V. 1966. Econometric Techniques and Problems, Charles Griffin & Co, London.

Lipovetsky, S., and M. Conklin. 2005. Regression by Data Segments via Discriminant Analysis, J. of Modern Applied Statistical Methods, 4, 63-74.

Lipovetsky, S., and M. Conklin. 2014a. Best-Worst Scaling in Analytical Closed-Form Solution, The Journal of Choice Modelling, 10, 60–68.

Lipovetsky, S., and M. Conklin. 2014b. Finding Items Cannibalization and Synergy by BWS Data, The Journal of Choice Modelling, 12, DOI: 10.1016/j.jocm.2014.08.001.

Lipovetsky, S., D. Liakhovitski, and M. Conklin. 2015. What's the right sample size for my MaxDiff study? Sawtooth Software Conference Proceedings, Orlando, FL.

Lipovetsky, S. 2010. Double logistic curve in regression modeling, J. of Applied Statistics, 37, 1785-1793.

Lipovetsky, S. 2012. Regression Split by Levels of the Dependent Variable, J. of Modern Applied Statistical Methods, 11, 319-324.

Lipovetsky, S. 2014. Analytical closed-form solution for binary logit regression by categorical predictors, J. of Applied Statistics, DOI: 10.1080/02664763.2014.932760.

Lipovetsky, S. 2015. Trinomial Response Modeling in One Logit Regression, Annals of Data Science, DOI: 10.1007/s40745-015-0039-7.

Louviere J.J. (1991) Best-Worst Scaling: A Model for the Largest Difference Judgments, Working Paper, University of Alberta.

Louviere, J.J. 1993. The Best-Worst or Maximum Difference Measurement Model: Applications to Behavioral Research in Marketing, The American Marketing Association's Behavioral Research Conference, Phoenix, Arizona.

Louviere, J.J., D.A. Hensher, and J. Swait. 2000. Stated Choice Methods: Analysis and Applications, Cambridge University Press, Cambridge.

Louviere, J.J.,_D. Street, L. Burgess, N. Wasi, T. Islam, and A. Marley. 2008. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information, Journal of Choice Modelling, 1, 128-164.

Madhavan, P., B. Stephens, and W. Low. 1995. Tri-state neural network and analysis of its performance, Int. J. of Intelligent Automation and Soft Computing, 1, 235-246.

Marley, A.A.J., and J.J. Louviere. 2005. Some Probabilistic Models of Best, Worst, and Best-Worst Choices, J. of Mathematical Psychology, 49, 464–480.

Marley, A.A.J., T.N. Flynn, and J.J. Louviere. 2008. Probabilistic models of set-dependent and attribute-level best-worst choice, J. of Mathematical Psychology, 52, 281–296.