Trinomial Response Modeling in One Logit Regression

Stan Lipovetsky[1]

Abstract Constructing a triple ordinal outcome model in one binary logistic regression is proposed. Various applied problems can be formulated with response variable of three ordinal categorical levels of negative–neutral–positive kind. Such a response is commonly considered in two ordinal logistic models or in two multinomial shares of three possible outcomes. The current work shows that the problem can be presented in a much simpler and convenient one binomial logistic regression model, so in one probability scale. This approach is based on a special data transformation used in the Best–Worst scaling or MaxDiff modeling, when the positive-neutral data subset is stacked with the negative–neutral subset and in the latter one the predictors' 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 obtained one logit regression, the positive category predictions are closer to 1, negative—closer to 0, and neutral are in the middle of its continuous 0–1 scale. Theoretical features and practical application of the model are discussed.

1 Introduction

In various applied problems the response variable is presented in three category ordinal levels. For instance, in marketing research a dependent variable of overall satisfaction is commonly measured in ordinal Likert 10-point scale from the worst to the best values. With several upper levels corresponding to satisfaction and bottom levels

✉ Stan Lipovetsky
    stan.lipovetsky@gfk.com

[1]   GfK North America, 8401 Golden Valley Rd., Minneapolis, MN 55427, USA

to dissatisfaction, it is a trinomial outcome of satisfaction–middle–dissatisfaction, or positive–neutral–negative segments modeled by various predictors. Modeling for such a trinomial variable is usually performed by a pair of ordinal logistic regressions or by two multinomial share models [5,18,19]. Many examples considered in the main statistical packages include: large-medium-small size of soda that people buy in relation to other meals and demographics; gold–silver–bronze medaling in Olympic sport, with relevant predictors of training hours, diet, age, and popularity of this kind of activity in athletes' home country; three categories of decision by college juniors asked if they are unlikely, somewhat likely, or very likely to apply to graduate school, with possible predictors of parental educational status, public or private undergraduate college, and GPA. Other examples of ordered response are presented in bond ratings, surveys with responses of agree–no opinion–disagree, high–mid–low levels of state spending on government programs, the levels none–partial–full of insurance coverage chosen, and employment status 'non-employed'–'employed part-time'–'fully employed' [4, pp. 824–827]. Tri-state neural networks [20] and fatigue profiles in experimental physiology [1] have been studied by logistic models with an intermediate and eventually reached levels of saturation. It is interesting to note that division of data to three groups has been also studied for continuous and discrete modeling ([3,6], and references within). Modeling for triple outcome can be performed in different regression and non-regression approaches [2,7,11,12]. Multinomial, ordinal logistic model, and other mentioned techniques are complicated numerical procedures which require special software.

The current work shows that the problem of three level modeling can be reduced to a much simpler and convenient one binomial logit model, so the categorical outcome can be transformed into one probability scale. It can be done in the approach based on a special data transformation used in the Best–Worst scaling or MaxDiff technique known in marketing research, which is a contemporary method for the prioritization of items proposed in [14–16], developed and applied in numerous works (for instance, [8–10,21,22]. In MaxDiff, 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. Logistic regression is constructed where the positive category predictions are close to 1, negative are close to 0, and neutral are in the middle of its continuous 0–1 scale.

2 MaxDiff and Trinomial Modeling

A general multinomial logit (MNL) defines a choice among several outcomes and can be described by the probability model:

$$p_k = \frac{\exp\left(a_1^{(k)}x_1 + \cdots + a_n^{(k)}x_n\right)}{1 + \exp\left(a_1^{(2)}x_1 + \cdots + a_n^{(2)}x_n\right) + \cdots + \exp\left(a_1^{(m)}x_1 + \cdots + a_n^{(m)}x_n\right)}, \quad (1)$$

where x_j are predictor variables ($j = 1, 2, \ldots, n$), and $a_j^{(k)}$ are parameters of the probabilities of each k-th choice among all m of them ($m = 3$ in the trinomial outcome).

For the sake of identification, the first share parameters are taken as reference, so put to zero.

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

$$p = \frac{\exp(a_1 x_1 + \cdots + a_n x_n)}{1 + \exp(a_1 x_1 + \cdots + a_n x_n)}. \qquad (2)$$

After finding parameters, the choice probabilities are estimated by (2) as the continuous values in the 0–1 interval. In discrete choice modeling (DCM), particularly with MaxDiff data, the binary outcome of the Best item versus non-best items (1 vs. 0) can be modeled by the logistic regression (2). 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 probability estimated by the logit model (2) equals 1-p which corresponds to absence of a binary event. Indeed, consider the transformation of sign change:

$$\frac{\exp(-(a_1 x_1 + \cdots + a_n x_n))}{1 + \exp(-(a_1 x_1 + \cdots + a_n x_n))} = 1 - \frac{\exp(a_1 x_1 + \cdots + a_n x_n)}{1 + \exp(a_1 x_1 + \cdots + a_n x_n)} = 1 - p. \qquad (3)$$

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 [17].

This approach can be applied to data with trinomial outcome of positive-neutral-negative segments considering the positive segment as the Best one, and negative as Worst one. 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 a more explicit presentation, let us express the positive-neutral segment and the negative-neutral segment of data in matrix form:

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

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

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

This stacked data is modeled in one logistic regression (2) which defines choice of positive or negative outcomes versus non-chosen cases. With 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 probability (2), respectively, while zero values yield the medium part of the logit curve.

Thus, a trinomial outcome can be considered via the binary logistic model, in one probability scale. Another important feature of modeling in one logit regression consists in possibility to obtain the analytical closed-form solution as described in [8,13].

3 Numerical Example

Consider data from a real marketing research project with nineteen hundred observations, two dozen predictors, and trinomial outcome dependent variable of 320 negative, 926 neutral, and 681 positive cases. We can try to apply ordinary least squares, also to use three binary logit models of each level versus the others, with additional normalizing the predicted values for each response to one. The third technique corresponds to direct modeling of the MNL regression (1), and the fourth approach is the suggested trinomial modeling via the binary logit. By the scale of probability predicted in logistic regression for all three levels we can 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 segment, and the other two segments, respectively.
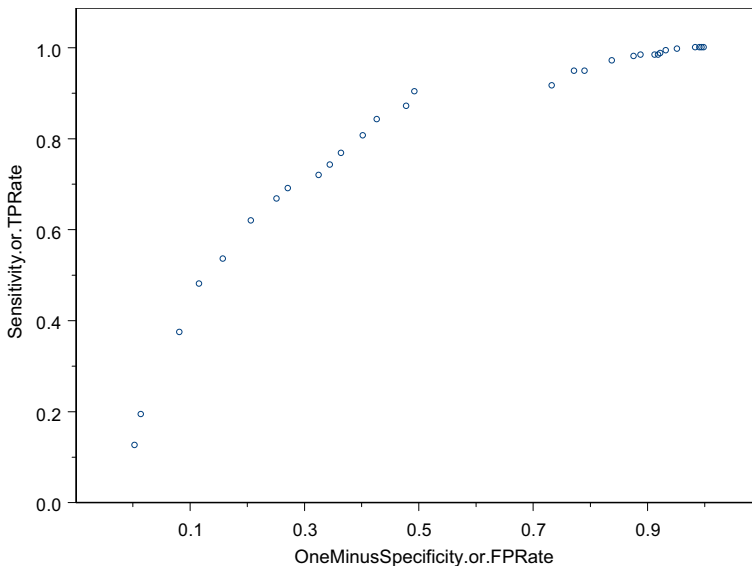

Fig. 1 ROC for Negative class: Sensitivity versus Specificity, or true positive rate (TPR) versus false positive rate (FPR)

Table 1 Prediction of trinomial outcome by several models

Class	Linear regression			Logit models normalized			MNL model			Trinomial binary logit		
	Neg	Neut	Pos	Neg	Neut	Pos	Neg	Neut	Pos	Neg	Neut	Pos
Negative	152	161	7	99	210	11	155	153	12	166	147	7
Neutral	124	615	187	51	693	182	79	675	172	123	620	183
Positive	4	109	568	6	119	556	5	130	546	4	116	561
Hit rate			0.6928			0.6995			0.7141			0.6990

Table 1 presents the results of 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 trinomial binary model yields the best prediction of the negative values for the smallest total count in the data.

4 Summary

The work describes a convenient approach to modeling trinomial ordinal categorical outcome via one binary logit regression. Theoretical features and practical application of this model are discussed, including a possibility of constructing logit regression in analytical closed-form solution. The considered approach suggests a new technique which is useful for problems requiring presentation of the trinomial events in one probability scale and helps in practical data fit and analysis.

References

1. Cairns SP, Robinson DM, Loiselle DS (2008) Double-sigmoid model for fitting fatigue profiles in mouse fast- and slow-twitch muscle. Exp Physiol 93:851–862
2. Conklin M, Powaga K, Lipovetsky S (2004) Customer satisfaction analysis: identification of key drivers. Eur J Oper Res 154:819–827
3. Gelman A, Park DK (2009) Splitting a predictor at the upper quarter or third and the lower quarter or third. Am Stat 63(1):1–8
4. Green WH (2012) Econometric analysis. Pearson Education, Boston
5. Johnson VE, Albert JH (1999) Ordinal data modeling. Springer, New York
6. Leser CEV (1966) Econometric techniques and problems. Charles Griffin & Co, London
7. Lipovetsky S, Conklin M (2005) Regression by data segments via discriminant analysis. J Modern Appl Stat Methods 4:63–74
8. Lipovetsky S, Conklin M (2014) Best–worst scaling in analytical closed-form solution. J Choice Model 10:60–68
9. Lipovetsky S, Conklin M (2014) Finding items cannibalization and synergy by BWS data. J Choice Model 12:1–9. doi:10.1016/j.jocm.2014.08.001
10. Lipovetsky S, Liakhovitski D, Conklin M (2015) What's the right sample size for my MaxDiff study? Sawtooth Software conference proceedings, Orlando
11. Lipovetsky S (2010) Double logistic curve in regression modeling. J Appl Stat 37:1785–1793
12. Lipovetsky S (2012) Regression split by levels of the dependent variable. J Modern Appl Stat Methods 11:319–324
13. Lipovetsky S (2015) Analytical closed-form solution for binary logit regression by categorical predictors. J Appl Stat 42:37–49
14. Louviere JJ (1991) Best–worst scaling: a model for the largest difference judgments, Working Paper, University of Alberta
15. Louviere JJ (1993) The best–worst or maximum difference measurement model: applications to behavioral research in marketing, The American Marketing Association's Behavioral Research conference, Phoenix
16. Louviere JJ, Hensher DA, Swait J (2000) Stated choice methods: analysis and applications. Cambridge University Press, Cambridge

17. Louviere JJ, Street D, Burgess L, Wasi N, Islam T, Marley A (2008) Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. J Choice Model 1:128–164
18. McCullagh P (1980) Regression models for ordinal data. J R Stat Soc Ser B 42:109–142
19. McCullagh P, Nelder JA (1997) Generalized linear models, 2nd edn. Chapman & Hall/CRC, Boca Raton
20. Madhavan P, Stephens B, Low W (1995) Tri-state neural network and analysis of its performance. Int J Intell Autom Soft Comput 1:235–246
21. Marley AAJ, Louviere JJ (2005) Some probabilistic models of best, worst, and best–worst choices. J Math Psychol 49:464–480
22. Marley AAJ, Flynn TN, Louviere JJ (2008) Probabilistic models of set-dependent and attribute-level best–worst choice. J Math Psychol 52:281–296