A Latent Class Conjoint Analysis for analysing graduates' profiles

Un modello Latent Class Conjoint per l'analisi dei profili dei laureati

Paolo Mariani, Andrea Marletta, Lucio Masserini and Mariangela Zenga

Abstract This paper aims to stabilize the relationship between universities and companies. Lombardy companies with at least 15 employees were asked them to manifest their preferences choosing among profiles of new graduates. A Latent Class Metric Conjoint Analysis is employed to evaluate the ideal graduate's profile for a job position and to detect the existence of subgroups of companies having homogeneous preferences about such features.

Abstract *Questo lavora mira ad analizzare la relazione tra aziende ed università. Alle imprese lombarde con almeno* 15 *dipendenti è stato chiesto di esprimere le loro preferenze fra alcuni profili dei neolaureati a loro sottoposti per una possibile nuova assunzione. Un modello Latent Class Conjoint è stato utilizzato per valutare il profilo ideale fra i candidati e individuare l'esistenza di sottogruppi di imprese.*

Key words: Labour Market, Latent Class Models, Conjoint Analysis, Electus

1 Introduction

During last years, the economic crisis exhibits effects about performances in business and particularly in the employment in all European countries. The impact of this crisis struck weaker segments of the labour market, in detail young person

Paolo Mariani
University of Milano-Bicocca, e-mail: paolo.mariani@unimib.it

Andrea Marletta
University of Milano-Bicocca, e-mail: andrea.marletta@unimib.it

Lucio Masserini
University of Pisa, e-mail: lucio.masserini@unipi.it

Mariangela Zenga
University of Milano-Bicocca, e-mail: mariangela.zenga@unimib.it

and people with less work experience. Most of the times, there is no possibility for turnover, so the younger people are not able to access the labour market. In Italy, according to Istat, from 2007 over 2014, the young unemployment rate (15-24 years) increased from 20.4% over to 44.7%.

In relation to the labour market, it reveals evident how the other side of the phenomena is represented by the companies and their expectations about the possibility for a new hiring. It appears useful to carry out information from the companies' point of view obtaining a deep analysis about what they are looking for.

In a perspective of synergy between education and labour market, a possible solution is represented by ELECTUS, an acronym standing for Education-for-Labour Elicitation from Companies' Attitudes towards University Studies, a research project involving several Italian universities [3]. The ELECTUS survey was conducted in 2015 using CAWI technique using a questionnaire containing two macro-sections. In the first part the entrepreneurs are asked to choose and rank four possible profiles of new graduates for five different job vacancies. In the second part the entrepreneurs are asked about their socio-demographic features.

The candidates' profile are characterized by six attributes: *Field of Study*, *Degree Mark*, *Degree Level*, *English Knowledge*, *Relevant Work Experience*, *Willingness to Travel on Business*.

The aim of this paper is to carry out a segmentation analysis of employers' preferences for graduates' profiles evaluated as candidates in a job position by using a Latent Class Metric Conjoint Analysis [1]. In fact, it has been highlighted that the tandem approach suffers from some problems: a) different clustering methods often yield different results, in terms of number of clusters and their composition; b) in presence of highly fractionated designs, as in our study, individual-level part-worth estimates are rather unstable and then may be untrustworthy when employed in successive clustering algorithms. On the other hand, the LCMCA, unlike cluster analysis, is a model-based approach in which model parameters and subgroups (segments or latent classes) are estimated simultaneously, and segments are composed of individuals whose part-worth utilities are similar.

The paper is organized as follows. Section 2 introduces Latent Class Metric Conjoint Analysis. Section 3 presents the results of the estimated model. Finally, Section 4 is reserved to discussion and final remarks.

2 Latent Class Conjoint Analysis

In this study, a Latent Class Metric Conjoint Analysis (LCMCA) [1] is employed is carried out to evaluate which characteristics of a graduate's profile employers prefer for a potential candidate in the job position of administrative clerk. In particular, in order to detect if there exist unobserved subgroups of employers having homogeneous preferences about graduates' characteristics for this position. Latent Class Metric Conjoint Analysis is a statistical modelling technique included in the more general class of Finite Mixture Models (FMMs) [4]. Following the FMMs approach, LCMCA relaxes the single homogeneous population assumption to allow for parameter differences across G latent classes and supposes that the marginal density function of the response variable y is given by a weighted sum over the G mixture components, with weights indicating the *a-priori* probability for an observation to come from a specific component [1]:

$$f(y_{ij}|\pi,x,z,\Sigma) = \sum_{g=1}^{G} \pi_{g|z} f_g(y_{ij}|x,z,\beta_g \Sigma_g) \tag{1}$$

where y_{ij} is the vector of response variable which refers to the rating expressed by employer i to conjoint profile j; $\pi = (\pi_1, \pi_2, \ldots, \pi_{G-1})$ are $G-1$ independent mixing proportions of the mixture, such that $0 \leq \pi_g \leq 1$; x is a matrix containing the M conjoint dummy variables which defines the profiles evaluated; β_g is the vector of the estimated conjoint part-worth coefficients for subgroup g, and $\Sigma = (\Sigma_1 \Sigma_2, \ldots, \Sigma_g)$ is $J \times J$ vector of covariance matrices of the error terms estimated for each subgroup. Moreover, given that the weights depend on a set of explanatory variables, also referred to as concomitant variables, z defines the vector of variables which characterise employers. Given the metric response variable, each of the conditional distributions, f_g, is conventionally specified as a conditional multivariate normal distribution. Instead, the prior probability of group membership varies as a multinomial logistic regression model, in function of the concomitant variables, as it follows:

$$\pi_{k|z} = \frac{exp(\gamma_{0g} + z\gamma_{1g})}{\sum_{g=1}^{G} exp(\gamma_{0g} + z\gamma_{1g})} \tag{2}$$

where γ_{0g} is the intercept while γ_{1g} contains the vector of regression coefficients, quantifying the effect of the concomitant variables on the prior probability for class g. For identification purpose, usually $\gamma_{01} = 0$ and $\gamma_{11} = 0$, and designate the first category as a reference class. Moreover, the following constraints hold: $\sum_{g=1}^{G} \pi_{g|z} = 1, \pi_{g|z} > 0$. Once the estimates of all the model parameters of the mixture of probability density are obtained, the posterior probability that an observation belongs to class g, denoted with p_{ig}, can be calculated by updating the previous according to the Bayes' theorem, as follows:

$$\hat{p}_{ig} = \frac{\hat{\pi}_{g|z} \hat{f}_{ig}(y_{ij}|x,z,\beta_g \Sigma_g)}{\sum_{g=1}^{G} \hat{\pi}_{g|z} \hat{f}_{ig}(y_{ij}|x,z,\beta_g \Sigma_g)} \tag{3}$$

where $\sum_{g=1}^{G} \hat{p}_{ig} = 1$ and $0 \leq \hat{p}_{ig} \leq 1$. The posterior probabilities provide a probabilistic allocation of observations to the latent classes and can be used to classify data by assigning each employer to the class with the maximum posterior probability. Parameter estimation was carried out via Maximum Likelihood (ML) by using the Expectation-Maximization (E-M) algorithm [2], in which conjoint part-worth coefficients and class membership are obtained simultaneously. The conventional Akaike's information criterion (AIC) was used for choosing the number of latent classes.

3 Application and Results

After estimating several LCMCAs, starting from the aggregate solution, with one class ($G = 1$), to the most complex one, with five latent classes ($G = 5$), the solution with $G = 3$ seems to represent the latent structure underlying the employers' ratings quite well.

Table 1 provides the estimated regression coefficients (or part-worths) for the $G = 3$ latent classes solution. Given that categorical attributes were preventively converted into the appropriate number of dummy variables, the intercept represents the average rating of the reference profile and refers to a graduate in foreign languages, with a Master's degree, a low final grades, with a knowledge of English language, no working experience, and not willing to business trips. At a first glance, it is clear the substantial difference between the aggregate part-worth coefficients and those in each sub-model of the three-class solution. Moreover, the aggregate model is also the one with the highest value of AIC and this confirms that a single set of regression coefficients estimated for all the employers may produce misleading results. The first class is the one with the lowest average rating corresponding to the reference profile (2.56) and identifies especially Economics as the most preferred degree, whereas Law, Statistics and Engineering are also appreciated but to a lesser extent. Among the employers included in this class, it seems preferable for a candidate to have some kind of work experience, excluding the internship. On the other hand, low final grades and willing to long-term business trips produce a lower preference. The second class is the one with the higher average rating (8.25) and this indicates that the reference profile is already highly appreciated. Given such a high average score, part-worth coefficients are almost all negative. In particular, employers within such class evaluate Economics, Engineering, Mathematics and computer sciences as less important degrees. On the contrary, a Bachelor's degree and a medium final grades increase employers' preference. The third class is the one with the intermediate average rating and its value is also very similar to that of the aggregate model (4.46). Mathematics and computer sciences and Economics are the most preferred degrees. Political science is also evaluated positively but to a lesser extent. For employers in this class, a previous work experience both as a stable experience and internship experience is relevant.

The Maximum likelihood estimates of the multinomial logistic regression model allows to determine which variables affect the latent classes' membership. Among the available variables which describe the characteristics and the context of the company, only the variables 'Recruitment of staff within one year', 'Company run by a manager' and 'Company committed also in the foreign market' seem to affect class membership when using the usual 0.05 as significance level. However, considering 0.10 as significance level also the variables 'Hired personnel over the past 3 years' and 'Education of the last administrative hired: graduated' contribute to explain class membership. In particular, the probability of being in Class 2 is higher for companies which plan to recruit staff within one year ($p = 0.007$) but lower for those that hired personnel over the past 3 years ($p = 0.070$) and those which have already taken a graduated as administrative ($p = 0.076$). On the other hand, the

Table 1 Maximum likelihood estimates of part-worth coefficients for each latent class

	Latent	class		
	1	2	3	Aggregate
Intercept	2.56**	8.25**	4.46**	4.79**
Philosophy and Literature	2.60**	-1.65**	-2.20	-0.88
Education sciences	1.54*	-0.02	-0.48	0.15
Political sciences	1.21	-1.23**	1.45**	0.97
Economics	6.52**	-5.21**	2.18**	1.92**
Law	4.29**	-3.75**	-0.47	-0.25
Statistics	3.99**	-2.77**	-0.63	0.42
Engineering	3.81**	-4.74**	-2.25**	-1.16*
Mathematics and computer sciences	2.68**	-5.41**	2.92**	0.45
Psychology	2.80**	-3.66**	-2.42**	-1.61*
Bachelor's degree	-0.23	0.59**	0.47	0.49
Low final grades	-1.62**	-1.13**	-1.01**	-1.25**
Medium final grades	-0.92*	1.70**	-0.94**	-0.50
No knowledge of English language	0.49	-0.75**	-1.57**	-0.87**
Internship experience	-0.56	0.84**	1.37**	0.46
Occasional working experience	2.10**	-0.25**	0.82	0.66
Stable working experience	1.04*	-2.28**	2.36**	1.07**
Willing to short-term business travels	-0.44	-1.46**	0.85**	-0.15
Willing to long-term business travels	-1.33*	-0.88**	0.53	-0.10

probability of being in Class 3 is higher for companies which plan to recruit staff within one year ($p < 0.0001$), for companies run by a manager ($p = 0.036$) and for companies committed also in the foreign market ($p = 0.033$) but lower for those that hired personnel over the past 3 years ($p = 0.072$). It seems adequate to identify the three latent groups by their peculiar features. In particular, the first group (26.4% of the sample) is characterized by companies lead by not a managerial view, working in a service sector in prevalence in domestic market, they neither will do recruitment new staff in the next year neither hired personnel over the past three years. This group could be named *Domestic Consolidated Companies*. The best profile for the AC required by these companies results to be related to a classical view of the position: a well graduated in Economics with some kind of working experience. The second group (17.1% of the sample) is composed by big sized companies, they will recruit staff in the next year, but they did not hire personnel over the past three years. This group could be called as *Static Companies*: respect to the expected profile of new graduates they seem to prefer new graduates with a not suitable major for AC (Language) with Bachelor's degree and a medium final grade, the English knowledge and working experience are not required. In some way, they prefer new graduates who must be fully trained. The third group (56.5% of the sample) is represented by small or medium enterprises, guided by a manager and committed also in the foreign market with a willingness to recruit new staff in the next year. They can be named as *Dynamic Companies*. The new graduate profile fits the description of these firms: it has to be a student in Economics or Political Sciences major, with a higher final grade and an English knowledge suitable to communicate with foreign people. Important requirements are also working experience and willing to business trips.

4 Discussion and final remarks

Using the survey ELECTUS, a segmentation of employers' preferences for graduates' profiles for administrative clerk is carried out by using a Latent Class Metric Conjoint Analysis. In general the features of the profiles for the new graduates' job are very different for every sub-groups but all respondents agree that a low final grades at graduation is not a preferable. Certainly the characteristics of the companies could influence the preferences about graduates' characteristics: the membership of the latent groups seems in fact to be effected by peculiar factors. In fact, the *Domestic Consolidated Companies*, run by not a managerial view, working in a service sector in prevalence in domestic market, without willingness to recruit in the past and in the future, require a well graduated students in Economics with some kind of working experience for AC position. The *Static Companies*, composed by big sized companies with the willingness to recruit in the future but not in the past, prefer new graduates who must be fully trained. The *Dynamic Companies*, represented by SME lead by a manager and committed also in the foreign market with a willingness to recruit new staff in the next year, look for a new graduate in Economics or Political Sciences major, with a higher final grade and an English knowledge suitable to communicate with foreign people with working experience and willing to business trips.

References

1. DeSarbo, W.S, Wedel, M.A., Vriens, M., Ramaswamy, V.: Latent Class Metric Conjoint Analysis. Marketing Letters, 3(2), 137-288 (1992).
2. Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EM-Alogrithm. Journal of the Royal Statistical Society: Series B, 39(1), 1-38 (1977).
3. Fabbris, L. and Scioni, M. Dimensionality of scores obtained with a paired-comparison tournament system of questionnaire item. In A. Meerman and T. Kliewe, eds., Academic Proceedings of the 2015 University-Industry Interaction Conference (2015).
4. McLachlan, G. J., Peel, D.: Finite Mixture Models. New York: Wiley (2000).
5. Wedel, M.A.: Concomitant variables in finite mixture models. Statistica Neerlandica, 56(3), 362-375 (2002).