ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2009, Vol. 43, No. 5, pp. 292–295. © Allerton Press, Inc., 2009.
Original Russian Text © V.M. Moskovkin, 2009, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, 2009, No. 10, pp. 19–21.

Simulation Expert System for Making Students' College Decisions

V. M. Moskovkin

Received July 9, 2009

Abstract—A brief survey has been performed of foreign investigations on the modeling of the process of decision making about student choices of colleges and universities; a suitable simulation expert system has been developed at a conceptual level.

Key words: simulation expert system, Google Scholar search engine, simulation models of college choice, integrated indicator of education attraction.

DOI: 10.3103/S0005105509050057

To the best of our knowledge, there are no Soviet-era scientific publications devoted to the choice of higher education institutions by secondary-school graduates, because there has been no demand for such research, as university and college education was free and widely accessible and a tradition for such studies has not yet developed over the post-Soviet period.

The situation in the west appears to be absolutely different. Our experiments with the search engine Google Scholar, which were organized as an advanced search using an exact phrase, showed that there is an established set of terms in this field (Table 1).

It can be seen from Table 1 that over three months the number of responses to the queries for the key terms of our research increased by about 5–10%. It is also clear that the number of responses to queries containing the word "college" significantly exceeds the number of queries for terms containing the word "university."

As early as 18 years ago, Larry Litten wrote that the process of college decision making was hard and pressing for many teenagers and their parents. Each of these groups spend an average of about 50 hours per year on the collection and analysis of information, without allowance for personal visits to the campuses of these higher education institutions, which normally cost about $1500 [1].

The most comprehensive review in this field was published by L. Henrickson in 2002 [2]. He pointed out that over 40 years about 1900 publications were devoted to "College choice" and "College access" and performed a detailed analysis of the three categories of models developed in this field of research:

—process models,

—theoretical models,

—methodological models.

The **process models** considered are three- to five-stage developmental models, in particular, a standard three-stage model [3], in which at the first stage the would-be student contemplates his intentions about college enrollment (from high school grade 8 to grade 12);

the second stage involves acquiring and studying information about the colleges the students are considering; determination of the set of higher learning institutions meeting his selection criteria; this can be based not only on limited characteristics, such as location or

Table 1. Queries for terms in the field of college choice made by Google Scholar search engine and responses to them

Terms	Number of responses	
	08.04.2009	08.07.2009
College choice	7010	7250
College access	3190	3460
University choice	862	982
University access	2390	2580
University decision-making process	75	85
College decision-making process	no measurements have been performed	150
University selection process	74	93
College selection process	no measurements have been performed	656

educational programs and syllabus, the student's perception of the particular college, and the tuition fee;

and the third stage includes the process of actual admission (acceptance) or rejection of the student by the college (admission process).

The **theoretical models** considered include models of higher education, as well as sociological and econometric models.

Further, L. Henrickson notes that over 60% of the works devoted to students' college decision making are based on empirical data and their later processing by statistical methods. He called this class of models **methodological**. Fifty percent of the studies of this class employ logarithmic probability functions combined with maximum probabilistic estimates and in addition twelve per cent of the publications use a factor analysis [2].

L. Henrickson associates the development prospects in this field of research with the use of nonlinear dynamics methods and complexity theory (which would enable the consideration to take into account feedback and heterogeneity), as well as of sophisticated computational algorithms. Some of the new computational methods include advanced methods of data storage (e.g., GIS) or new methods for computer programming and modeling (e.g., computer modeling using genetic algorithms, neural networks, agent-based models, or complex adaptive systems) [2]. Further, he dwells in detail on agent-based modeling and notes that this is a totally novel computing technique in social sciences (only two articles have been found devoted to this topic) based on C_{**} object-oriented programming. The agents in this model are a student and a college, each of which acts relatively independently in making his/its decision, abiding by simple rules of interaction but the resulting macro enrollment patterns are very complicated.

Numerous foreign empirical studies carried out within the frame of sociological model development are based on the typification of information sources connected with the students' college choice and the quantitative assessment of the college choice factors on the basis of the weights of higher educational establishments generated by students [4, 5]. For example, Galotti in [4] uses a seven-point scale for assessing the importance of information sources (from 1 to 7) and a scale from 0 to 10 for rating the significance of various factor weights in college choice made by students.

This literary survey of publications sources in the field of students' choices of colleges and universities gives us better understanding of the main factors of college choice when the students are able to assess the significance of these factors themselves. This is very important when developing simulation models of college choice. We have not, however, discovered any work that synthesized the features of simulation modeling with online facilities for the purpose of developing simulation expert systems for seeking and choosing colleges and universities for further education on a global scale. We have developed such a system on a conceptual level.

* * *

It will be assumed that high school graduates in their choice of the worldwide universities on a qualitative level usually take into account the following factors: college rating (prestige and status), tuition cost, cost of living (living wage and the cost of a market basket) in the university/college location, living conditions (accommodation) and security conditions in the respective community.

On the basis of these five factors let us devise the simplest formalized procedure to determine the attractiveness of studying at the *i*-th college/university. To this end, all these factors must be represented in the form of quantitative indicators. The first three indicators correspond to the measured (estimated) statistical data (hard data), the remaining two are based on questionnaires (survey data). Let us consider these cases one by one.

1. **University ranking** is determined by means of an integrated index of one of the international university rankings (Chinese university ranking, Spanish webometric ranking, Times Higher Education ranking (Great Britain) etc.). For the *i*-th college/university this will be designated by I_i. If there are no available data about integrated indexes, rankings themselves can be used (university ranking positions). For example, for the "∈ – -500 of the Chinese university ranking, it will be assumed to be $I_1 = l$, $I_{500} = 500$.

2. **The tuition cost** at the *i*-th college/university will be designated as $\$C_i$.

3. **Living wage** in the municipality where the *i*-th college/university is located will be designated as $\$LW_i$. Here, it should be taken into consideration that there may be several colleges/universities in the same city, therefore, the LW values for those colleges/universities will be the same.

4. **Living conditions** at the *i*-th college/university are determined on the basis of, say, a five-point score, BLC_i. Thus, $BLC_i = 5$ corresponds to excellent living conditions and $BLC_i = 1$ reflects very poor living conditions (seven- or ten-point scores can also be used).

5. **Security conditions** in the municipality where the *i*-th college/university is located will also be determined by means of five-point score, BSC_i: $BSC_i = 5$ corresponds to a very high security level and $BSC_i = 1$ reflects a very low security level.

Now it is necessary to use a procedure for matching actually measurable (estimated) data from the first three indicators to the scores of the two last indicators. Here, attention should be paid to the fact that the scores of the last two indicators are devised as "stimulators." The integrated index of a worldwide university ranking is also a "stimulator" (the higher the integrated index,

Table 2. University/college classification based on attractiveness for students

Variation range of the integrated indicator UAI_i	Attractiveness for students
0–1	very low
1–2	low
2–3	medium
3–4	high
4–5	very high

the higher the college or university's attraction will be for students), it can be converted to the 5-point score by the following formula:

$$BI_i = \frac{4(I_i - I_{min})}{(I_{max} - I_{min})} + 1. \quad (1)$$

If I_i stands for a ranking position being an "stimulator" then

$$BI_i = \frac{4(I_i - I_{max})}{(I_{min} - I_{max})} + 1. \quad (2)$$

For example, in a Chinese university ranking ($I_1 = I_{min} = 1$, $I_{max} = 500$) Harvard University, which ranks first, will get the best score ($BI_1 = 5$), and the university at the bottom of the list will have the worst score ($BI_{500} = 1$).

The second and third indicators are "destimulators" and the following formulas will hold for their conversion to scores:

$$BC_i = \frac{4(C_i - C_{max})}{(C_{min} - C_{max})} + 1, \quad (3)$$

$$BLW_i = \frac{4(LW_i - LW_{max})}{(LW_{min} - LW_{max})} + 1. \quad (4)$$

The results of calculations by formulas (1)–(4) are not rounded to integer estimation scores, otherwise significantly differing values of the first three indicators will yield the same score.

Note that a formula of type (1) is used for the conversion of actual statistical data to seven-point score estimates in the methods for the calculation of the countries' competitiveness index of the World Economic Forum.

Now, on an additive basis, the integrated index of attractiveness of the *i*-th college/university to prospective students allowing for the weights of specific indicators (University Attractiveness Index, UAI):

$$UAI_i = \gamma_1^* BI_i + \gamma_2^* BC_i + \gamma_3^* BLW_i + \gamma_4^* BLC_i + \gamma_5^* BSC_i, \quad (5)$$

where $\sum_{i=1}^{5} \gamma_i = 1$.

It can be shown that

$$\min\{UAI_i\} = \sum_{i=1} \gamma_i = 1,$$

$$\max\{UAI_i\} = 5\sum_{i=1}^{5} \gamma_i = 5,$$

i.e., the integrated index of college/university attractiveness for a student will vary over the same range as its specific indicators. Therefore, on the basis of a uniform scale for this indicator, the following five-level college/university classification system can be developed, ranking their attractiveness (Table 2).

For generalized estimates of worldwide college/university attractiveness, the weighing factors can be obtained by the methods of pairwise comparisons and expert estimation but this is not at all necessary for assessments of individual institutions.

* * *

The high-school graduate can make his own choice of acceptable weights. For example, a financially well-off high-school graduate will assign a high priority to the college ranking and security conditions and all the rest of the indicators will be less significant for him. By varying the weights and trying various systems of college/university ranking, a high-school graduate can play out different scenarios of his college choice, thus obtaining a simulation expert system for students' college choice.

The set of colleges/universities used in making the choice must be determined by the potential student's preferences by means of limits imposed on specific indicators. For example, out of the first two hundred universities in the TOP-500 list according to some system of ranking universities, the high-school graduate wishes to choose universities satisfying the following limitations: $C_i \leq = \$4000$ a year; $LW_i \leq \$50$/day; $BLC_i = 3$ or 4; and $BSC_i = 4$ or 5. For this set of limitations on particular indicators the maximum of the integrated indicator UAI_i must be sought, i.e., we come to a problem of linear programming.

REFERENCES

1. Litten, L.H., *Ivy Bound: High Ability Students and College Choice*, New York: College Board, 1991.
2. Henrickson, L., Old Wine in a New Wineskin. College Choice, College Access Using Agent-Based Modeling, *Social Science Computer Review*, 2002, vol. 20, no. 4, pp. 400–419.
3. Litten, L.H., Different Strokes in the Applicant Pool: Some Refinements in a Model of Student College Choice, *Journal of Higher Education*, 1982, vol. 55, no. 4, pp. 383–402.
4. Galotti, K.M. and Mark, M.C., How do High School Students Structure an Important Life Decision? A Short-Term Longitudinal Study of the College Decision-Making Process, *Research in Higher Education*, 1994, vol. 35, no. 5, pp. 589–607.
5. Veloutsou, C., Lewis, J.W., and Paton, R.A., University Selection: Information Requirements and Importance *International Journal of Educational Management*, 2004, vol. 18, no. 3, pp. 160–171.