A NON-LINEAR OPTIMIZATION TASK FOR FORMING OPTIMAL SETS OF SELECTIVE DISCIPLINES

Introduction
Systems of e-learning are getting wide-spread now; most of them became an inherent part of the educational process. Automated enrollment of students on optional disciplines becomes a typical feature of such systems.

The problem is that teachers and tutors are quite free to add new educational materials and to create new courses. These courses may be very similar each to other and may even duplicate each other. So it may be difficult for students to choose the most appropriate selective disciplines. Therefore systems for automated enrollment of students on optional disciplines should be complemented with features of recommendation systems addressing both to students and to faculties.

In the simplest case these recommendations can be elaborated by solving a problem similar to the well-known knapsack problem. Let \( D = \{d_1, \ldots, d_n\} \) be a set of optional disciplines, \( c_i, i=1, \ldots, n \) be the usefulness of the \( i \)-th discipline and \( q_i, i=1, \ldots, n \) be a number of credits for the \( i \)-th discipline. We suppose that students make their choices independently, and moreover we can admit that an unique optimization problem is invariant to a specific student. So we introduce the boolean variable

\[
x_i = \begin{cases} 1, & \text{if the student chooses } i-th discipline} \\ 0, & \text{otherwise} \end{cases}
\]

Then the following optimization task arises:

\[
\sum_{i=1}^{n} c_i x_i \rightarrow \max,
\]

\[
\sum_{i=1}^{n} q_i x_i \leq W,
\]

\[
x_i \in \{0,1\}
\]

where \( W \) is a given limit for the total amount of credits.

But the stuff is not so simple, and we have to consider different variations of this basic model. This paper considers an issue that the value of any discipline depends not only upon this discipline itself but of selection of other disciplines as well.

Main part
We have to take into account not only the values of separate disciplines but some functions of usefulness of the whole sets of selected courses. We will denote the usefulness of the set \( S=(s_1, \ldots, s_n) \) as \( u(S) \), or \( u(s_1, \ldots, s_n) \). This function, of course has not to be linear. We distinguish three cases:

- function \( u \) is additive: \( u(S) = \sum_i u(s_i) \)
- function $u$ is sub-additive: $u(S) < \sum_i u(x_i)$
- function $u$ is super-additive: $u(S) > \sum_i u(x_i)$

So we regard a more complicated optimization task which is generally non-linear:

$$u(x_1, ..., x_n) \rightarrow \max,$$

$$\sum_{i=1}^n q_i x_i \leq W,$$

$$x_i \in \{0,1\}$$

Of course, solving this problem by means of the straightforward search is usually unrealistic because of the dimensionality problem. We regard the following heuristic approaches:

- using genetic algorithms shown to be good for solving different optimization tasks [1]. As a goal function we propose to use the discrepancies between the rankings of sets obtained by the algorithm and those recommended by experts. An important issue is that we may regard not the whole set of all possible pair comparisons but some subset of them only;

- using techniques of Data Mining [2] for discovering typical patterns of discipline sets; Apriori algorithm appears to be appropriate for this task;

- reducing values of disciplines which is very similar to those already included to the selected set; an approach similar to the one based on the fuzzy rule “if the distance is AVERAGE then the relevance is HIGH” where AVERAGE DISTANCE and HIGH RELEVANCE are fuzzy concepts defined by their membership functions [3] is being developed;

- combining of these approaches.

Conclusions

It appears to be useful to build a recommendation systems aimed to choosing optimal sets of selective disciplines within the e-learning environment. It should be taken into account that the value of any discipline depends not only upon this discipline itself but of selection of other disciplines as well. A non-linear optimization task is suggested as the model of this problem; some heuristic approaches for solving this task on the base of soft computing are regarded.

References