Interfaces with Other Disciplines

Using similarity measures for collaborating groups formation: A model for distance learning environments

Yannis A. Pollalis [a,*], George Mavrommatis [b]

[a] *Department of Economic Science, University of Piraeus, 18534 Piraeus, Greece*
[b] *Hellenic Military Academy, 16673 Vari, Attica, Greece*

Abstract

Adaptability to each separate user's needs and preferences is a common concern in modern e-learning systems. Among various adaptation techniques described in recent research, *collaboration support* seeks to create groups that efficiently work together in order to advance user's learning. This paper defines two similarity coefficients between users and learning objects and focuses on automatic creation of properly matching collaborating groups based on an algorithmic approach. By adopting methods derived from Group Technology, the method simultaneously selects appropriate learning objects to form a corresponding educational package for each group, thus assuring optimal value of user's learning.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Education; Distance learning; Information retrieval; Clustering methods; Group technology; Collaboration

1. Introduction

More than sixty years after the introduction of the first computer system, Internet has become the standard platform for e-learning environments. E-learning, the contemporary version of distance education, is mainly web-based, conducted by means of Internet-connected computers running special programs (learning content management systems, LCMS), which bring learners, teachers, courses and collaborative technologies into contact.

Adaptive e-learning systems, an alternative to the traditional approach in the development of e-learning platforms, use a variety of methods in order to adapt to the needs of each separate user. Modern, advanced information and communication technology has to be used in more ways than simply retrieving learning material. Research on adaptive e-learning is almost ten years old, yet adaptive learning environments are mainly research prototypes with little, if any, standards compliance (Paramythis and Loidl-Reisinger, 2004). There are two major drawbacks obstructing broad use (Brusilovsky, 2004): lack of integration and lack of re-use support. However, both drawbacks may be resolved successfully by the use of learning objects (LO) technology (Alvarado-Boyd, 2003).

A vast variety of definitions of LO can be found in literature. In a recently published paper a Learning Object is defined as a "*standalone, reusable, digital resource that aims at teaching one or more instructional objectives or concepts*" (Mavrommatis, 2006a). The Learning Objects Metadata (information used to describe a LO) framework is described in the Sharable Content Object Reference Model (SCORM, 2004) and several learning objects repositories (pools containing retrievable LO) are already in common practice for distance learning.

Brusilovsky (1998) presents a review of adaptation technologies; among them, *adaptive collaboration support* is defined as *the technology that uses system's knowledge about different users to form a matching collaborating group*. Most current web-based educational systems collect large amounts of information about the students but this

---
* Corresponding author. Tel.: +30 2104142353; fax: +30 2104142055.
E-mail addresses: yannis@unipi.gr (Y.A. Pollalis), gmav@unipi.gr (G. Mavrommatis).

information, so far, is not widely used by instructors (Mazza and Dimitrova, 2004). Supporting collaborative learning is one of the most recent approaches of adaptive educational systems. The use of collaborative methods can extend e-learning from an individual learner to a group of learners (Mödritscher et al., 2004).

Collaborative learning activities are based on constructivist learning theory (Wilson, 1996) and, although collaboration in the classroom has proven itself a successful learning method, online collaborative learners do not seem to enjoy the same benefits, mainly because distance learning technologies do not provide guidance nor direction during online discussion sessions (Soller and Lesgold, 2003). Most of the commercial, standards-based e-learning platforms currently used in higher education institution, allow very little collaboration by simply providing basic tools (Van Rosmalen et al., 2004). To make things even worse, Fung and Yeung (2000) found in literature fifteen research adaptive educational systems that were then reviewed to check their level of adaptivity. They were found to support a subset of the known adaptation technologies. Among them however none was reported to support adaptive collaboration.

On the other hand, the importance of collaboration is increasingly underlined by researchers and learning theorists: cooperative learning, communities of learners, social negotiation etc, are some examples (Wiley, 2003). Collaboration occurs when learners somehow work together to accomplish shared learning goals (Johnson et al., 2000).

In order to achieve maximum benefits, collaboration has to rely on well adjusted learning teams, therefore placing users (learners) randomly in a group and assigning them a task is not enough (Soller, 2001). Interacting with other people is crucial for a contemporary learning environment, and "interactive" learning (Tapscott, 1998) requires adaptive learning, which embodies adaptive collaboration. The first step in directing collaborative learning environments is, therefore, forming the right group(s) of learners. Additionally, re-use of a collaborative environment on a variety of courses is essential; therefore, it is necessary that adaptation tasks will be domain independent (Pollalis, 1996; Gaudioso and Boticario, 2003).

This paper presents a method for course construction that promotes collaboration in order to achieve a common educational objective for a community of learners. When doing the group work it is useful to form some subgroups bearing in mind common (or differentiating) student aspects. The method creates properly matching collaborating groups and at the same time selects appropriate learning objects to form the corresponding course's core for each group.

The paper is organized as follows: the following section contains a mathematical model, based on simple information vector spaces, where we present both Learners and Learning Objects as the basis for similarity coefficients within educational technology; in Section 3 the Educational Cells are formed by applying clustering methods to the Learning Objects–Learners array; an application-example is presented in Section 4; in chapter 5 a few additional options are presented, aiming to integrate the model within a modern distance learning environment; and, finally, some conclusions are drawn in Section 6.

2. Resemblance coefficients in learning technology

Task Analysis, probably the most important component of Instructional Design, includes methods like Learning Hierarchy Analysis, Learning Contingency Analysis (Jonassen et al., 1999), or even, Principled Skill Decomposition (van Merriënboer, 1997). In general, these methods presume that every knowledge field or complex cognitive skill to be taught can be broken down into constituent skills, finally leading to construction of a learning hierarchy (similar to an ontology). A detailed description of such methods, the assumptions that each method makes, together with their advantages and disadvantages are also presented in Jonassen et al. (1999).

The Learning Hierarchy (Gagné et al., 1992), is a central idea in Gagné's theory of learning: in order to plan instruction, one must first identify a specific learning objective and construct a learning hierarchy for that objective. This learning hierarchy also determines the prerequisites for a given learning objective.

By using such a method, a certain knowledge field can be broken down to its constituent parts – the nodes, thus creating an Information Space S that contains all component parts composing the knowledge field, that we call *properties*

$$S = \{s^i\}, \quad i \in M = \{1, 2, \ldots, m\}$$

It must be stressed here that the decision of whether the skills' analysis has reached a low enough level is up to the designer. Furthermore, Annett et al. (1971), as reported in Stanton (2006) point out that this part is possibly "one of the most difficult features of task analysis".

On the other hand, the "real world" is much more complicated compared to the outcome created by these methods. Other, more qualitative models, mainly recognized under the general title of Concept mapping, possibly make a better capture of the details and characteristics of a domain. Based on Ausubel's learning theories, Concept Mapping was first developed by Novak: "Concept maps are graphical tools for organizing and representing knowledge" (Novak and Cañas, 2006).

Concept maps are being used in education (Walker and King, 2002; van Zele et al., 2004), and a lot of work has been done on this field: instructional designers use maps for content and (more detailed) task analyses (Milam et al., 2000). Using Concept maps has several advantages (Daley, 2004) but also a few disadvantages

- increased complexity (Daley, 2004),
- no standard formalization (e.g., definition, linking words, etc) (Milam et al., 2000), and

- difficulty to be associated with mathematical modeling and derivation of numeric results (Bredeweg and Forbus, 2003).

The present paper adopts and uses the well grounded on sound learning theories – notion of Learning Hierarchy Analysis for the sake of modelling and simplicity. There is a growing re-investment in such methods, mainly due to the Learning Objects advent (Impedovo et al., 2006; HFI DTC, 2007). Instructional Designers often make such assumptions, in order to promote course creation. Besides, E.O. Wilson (1998, p. 54, as reported in Merrill, 2001) states that "the cutting edge of science is reductionism, the breaking apart of nature into its natural components".

Finally, it is well known that methods such as the Learning Hierarchy Analysis are tools of the learning theory of Objectivism. Although Objectivism, compared to Constructivism, is considered as "the other end" of the learning theories continuum, it has been reported that both can be used simultaneously in order to promote learning construction (Wilson, 1997; Moallem, 2001; Cronjé, 2006). The applicability of these issues related to the present paper are discussed in Section 5.

In what follows, a Vector Space Model was chosen to represent Learning Objects and Learners for the following reasons:

- it is simple, easily implemented, based on well-known and widely used concepts;
- it may be subject to a large variety of information manipulation and retrieval techniques;
- other more complicated models, for example weighted models, can be reduced to it.

Let L be the set of Learning Objects dealing with a certain knowledge field, where

$$L = \{\lambda_p\}, \quad p = 1, 2, \dots, t.$$

Every λ_p in L can be represented as an m-dimensional vector of $\{0, 1\}^m$

$$\lambda_p = (\lambda_p^1, \lambda_p^2, \dots, \lambda_p^m), \text{ where}$$

$$\lambda_p^i = \begin{cases} 1, & \text{if } \lambda_p \text{ covers property } s^i, \\ 0, & \text{otherwise.} \end{cases}$$

A modern e-learning environment can support a large number of students, each one with his/her own different knowledge of properties in S. Let

$$U = \{u_q\}, \quad q = 1, 2, \dots, k$$

be the set of learners. A user/learner $u_q \in U$ can be represented by a similar vector of $\{0, 1\}^m$

$$u_q = (u_q^1, u_q^2, \dots, u_q^m), \text{ where}$$

$$u_q^i = \begin{cases} 1, & \text{if user masters skill } s^i, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 2.1. A Learner $u_q \in U$ *masters* Information Space S if

$$u_q^i = 1 \quad \forall \; i = 1, \dots, m.$$

It is implied then, that for each user $u_q \in U$, vector $\overline{u_q}$ represents the corresponding user's educational needs in order to reach the educational target, which is mastering Information Space S.

Based on the above definitions, both Learners and Learning Objects are being treated uniformly as elements of the same Information Space S. In order to efficiently associate learners and educational material, we define the proper resemblance coefficients.

Definition 2.2. For each user $u_q \in U$ and each Learning Object $\lambda_p \in L$ we define the following:

   I. **a** is the number of properties contained in both vectors $\overline{u_q}$ and λ_p (matching '1's);
  II. **b** is the number of properties the Learner lacks that is not presented by the specific Learning Object ('1's in $\overline{u_q}$ but '0's in corresponding position of λ_p);
 III. **c** is the number of properties the Learner already masters (therefore, does not need) but they are being presented by the specific Learning Object (redundant properties, '0's in $\overline{u_q}$ but '1's in corresponding position of λ_p);
 IV. **d** is the number of properties contained in neither of the vectors $\overline{u_q}$ and λ_p (both '0's).

Definition 2.3. For each learner u_q in U, we define the Hamming weight of vector $\overline{u_q}$ as the learner's Distance $D(u_q)$ from mastering S (the educational target).

Directly derived from Definitions 2.1, 2.2 and 2.3 we deduce the following proposition:

Proposition 2.1. *The following conditions hold*:

   I. $D(u_q) = a + b$
  II. $D(u_q) \in [0, m]$
 III. $D(u_q) = 0 \Rightarrow$ *learner u_q masters Information Space S.*

The Jaccard's coefficient is a common and effective measure of similarity, widely used in numerical taxonomy, also widely used in Group Technology (Sarker, 1996). Recently, it has been reported to be the most stable among 20 well-known similarity coefficients (Yin and Yasuda, 2005). Incorporating the special characteristics of educational Information Space S, we adapt the Jaccard's similarity coefficient in order to measure similarities between two 0–1 vectors representing a Learner and a Learning Object.

Definition 2.4. The *General Resemblance* coefficient (GR) between Users and Learning Objects is the Jaccard's coefficient between $\overline{u_q}$ and λ_p

$$GR(u_q, \lambda_p) = J_{qp} = \frac{a}{a+b+c},$$

for each user $u_q \in U$ and Learning Object $\lambda_p \in L$, $q = 1, \ldots, k, p = 1, \ldots, t$.

Definition 2.5. For each user $u_q \in U$ and each Learning Object $\lambda_p \in L$, we define the Learning Object's *Relevance* (R) to the User's Educational Needs as the quotient

$$R(u_q, \lambda_p) = \frac{a}{a+b}, q = 1, \ldots, k, p = 1, \ldots, t.$$

Let $u_q \in U$ be a user and $\lambda_1, \lambda_2 \in L$ learning objects.

- Assuming $R(u_q, \lambda_1) > R(u_q, \lambda_1)$ means that $\frac{a_1}{a_1+b_1} > \frac{a_2}{a_2+b_2}$ which leads us to $a_1 > a_2$, since by Proposition 2.1 (I) it is $D(u_q) = a_1 + b_1 = a_2 + b_2$. Therefore, we arrive at another proposition

Proposition 2.2. $R(u_q, \lambda_1) > R(u_q, \lambda_2) \Rightarrow a_1 > a_2$.

- *Additionally,* $R(u_q, \lambda_1) = R(u_q, \lambda_2) \Rightarrow a_1 = a_2$ *and if* $GR(u_q, \lambda_1) > GR(u_q, \lambda_2)$ *then* $\frac{a_1}{a_1+b_1+c_1} > \frac{a_2}{a_2+b_2+c_2}$ *which leads us to* $c_1 < c_2$. *This leads to our final proposition*

Proposition 2.3. $R(u_q, \lambda_1) = R(u_q, \lambda_2)$ *and* $GR(u_q, \lambda_1) > GR(u_q, \lambda_2) \Rightarrow c_1 < c_2$

The above analysis takes under consideration solely the content, in order to classify Learning Objects. Most Collaborative Learning methods demand teams with members of (somehow) comparable abilities and knowledge (Knight and Bohlmeyer, 1990). Therefore, *what* a Learner *knows* before he starts acting, is crucial for team formation (Pollalis, 1996). A large variety of factors and properties that has been used for Group creation can be found in literature (Bekele, 2005; Muehlenbrock, 2006). We believe that, as distance learners hardly get into physical contact, characteristics like gender, nationality, age, shyness, religion, color, etc are of less importance than they are in the case of classic classroom collaboration. Therefore, we believe, the proposed model is adequate, though not restrictive at all.

In the following section we will create the Learning Compatibility Matrix upon which we use an Array-based clustering method. This method will help us create groups of learners, together with the corresponding learning material especially suitable for each group.

3. Parametric clustering for group formation

The idea that (demand-driven) production systems share many similarities with educational systems, is more than ten years old (Kester et al., 2001). By analogy to the machines-parts table in Manufacturing Cell Formation, an application of Group Technology to production, we define the Learning Objects–Learners table which we call the learning compatibility matrix (LCM).

$$\text{LCM} = (c_{ij}), i \in [1, t], j \in [1, k] \text{ such that}$$

$$c_{ij} = \begin{cases} 1, & \text{if } R \geqslant \partial^R \text{ and } GR \geqslant \partial^{GR}, \\ 0, & \text{otherwise.} \end{cases}$$

where $\partial^R, \partial^{GR} \in (0, 1]$.

Creation of Educational Cells (groups of Learners and corresponding set of LO) has to accomplish certain restrictions, shown in Table 1.

Clustering on every consistent LCM may produce a feasible solution. Therefore, one may use any consistent LC Matrix to form the Educational Cells. Selection of the proper LCM can be done subject to Matrices' density limitations which depends on number of desirable clusters, learners and learning objects per cluster

$$d = f(\text{no of clusters, learners per cluster, learning objects per cluster}).$$

The problem can be formulated as follows:

Perform clustering on LC Matrix created subject to the following restrictions:

Max $R(u_q, \lambda_p)$ and then Max $GR(u_q, \lambda_p)$

$$\forall j \in [1, k] : \sum_{i=1}^{t} c_{ij} > 1$$

$$\sum_{i=1}^{t} \sum_{j=1}^{k} c_{ij} \geqslant d \quad d \text{ is a minimum desirable absolute}$$

density of LCM.

Table 1
Inter-relations within a sample Learning Community

1st	Maximize Covering of the Educational Target. It is faced here as a matter of similarity between learner's educational demand and learning object's knowledge supply. It is measured by the Relevance coefficient R. The Relevance coefficient divides set L into subsets containing Learning Objects with content (properties) of increasing similarity to the Learner's demands (Proposition 2.2).
2nd	Minimize Redundancy of properties. A Learning Object may present properties a Learner already masters, therefore (s)he does not need. This is measured by the General Resemblance coefficient GR, *within* each class defined by *R* (Proposition 2.3).
3rd	Every Learner corresponds *at least* to one Learning Object. An LC Matrix with this property is said to be *consistent*.
4th	Uniform, as possible, distribution of participating Learners into Groups. Obviously selection of ∂-parameters affects density *d* of LC Matrix and therefore the quality of produced educational cells.

The algorithmic approach to solve the problem tries different combinations of ∂-parameters and selects, among consistent LC Matrices, the one with proper density of '1's to perform clustering.

We start by giving $\partial^R = 1$, its maximum value, showing complete compatibility between Learning Object and Learner. ∂^{GR} is also started from 1 showing that there are no redundant properties present. ∂^{GR} decreases until it reaches 0, then we decrease ∂^R and start again with $\partial^{GR} = 1$.

According to Merrill (2002): "Learning is facilitated when existing knowledge is activated as a foundation for new knowledge". With the proposed approach, we are not prohibiting the LOs from *using* concepts already known to the learner. By setting $\partial^{GR} = 1$ we are trying to prohibit LOs from *presenting* concepts already known to the learner. However, if one wishes to *present* material already familiar together with new concepts, then one may vary ∂^{GR} within an interval $(0, x]$, where $x < 1$.

In order to vary both coefficients we use e, a step-value, which may be given a desired value. For each selection of ∂-parameters, a version of LCM is constructed. If LC Matrix is consistent, we count the number of '1's within the table, seeking to exceed d. Clustering is performed on LC Matrix with number of '1's exceeding d. The whole procedure is described by Algorithm *Educational-Cells*, and outlined in Fig. 1 that follows.

```
Algorithm Educational-Cells
set e, d to the desired values
∂R = 1
∂GR = 1
finished = false
while NOT finished
{
    construct LCM
    if LCM NOT consistent
       Call theta-change
    else
    {
       density = count '1's in LCM
       if density ⩾ d
       {
          store LCM
          finished = true
       }
    }
}
if consistent LCM found Call Clustering//on LCM to
form Educational Cells
End Educational-Cells
```


Fig. 1. The procedure outline.

Procedure theta-change
if $\partial^{GR} > 0$
$\partial^{GR} = \partial^{GR} - e$
else
{
   if $\partial^{GR} > 0$
   {
   $\partial^{GR} = \partial^{R} - e$
   $\partial^{GR} = 1$
   }
   else finished = true
}**End** theta-change

Procedure *Clustering* may implement any array-based clustering method. In order to make things more transparent, an application-example is presented in the following section.

4. A learning community groups its members towards a common objective

As an example, let us assume a certain knowledge field, where Task Analysis yielded a 5-dimensional Information Space $S = \{s^1, s^2, s^3, s^4, s^5\}$. The Repository contains $t = 10$ Learning Objects and the Community has $k = 6$ Learners. All Learning Objects' and Learners' properties vectors, together with computed corresponding General Resemblance (GR) and Relevance (R) coefficients between each couple, are shown in Table 2.

Algorithm *Educational-Cells* produces several consistent LCM instances. Instances with $\partial^R \in [1.0, 0.6]$ and $e = 0.1$, are displayed in Table 3.

For the demonstration purposes of this paper we demand a total of $n_1 = 3$ clusters created, each one contain-

ing $n_2 = 2$ users and $n_3 = 3$ learning objects. The minimum number of 1's the array has to contain in order to achieve proper clustering with these specifications is $d = n_1\, n_2\, n_3 = 18$. Therefore, we may select the following LCM instance generated with $\partial^R = 0.6$ and $\partial^{GR} = 0.4$ with density = 18, upon which we perform clustering.

$$LCM = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Of course, one may relax the condition concerning density of LCM if one wishes to promote another condition, for example fix a certain value for one (or both) ∂-parameter(s).

The process that leads to Educational Cells formation, consists of three major steps

1. Computation of both similarity coefficients R and GR for all Users and Learning Objects pairs and store the values. This is done only once during initialization, in polynomial time.
2. Run algorithm *Educational-Cells*. The algorithm's heart consists of a couple of nested loops where the control variables vary from 1 to zero with step e. Each loop makes $\lfloor \frac{1}{e} \rfloor$ iterations, that leads to a total of $\lfloor \frac{1}{e} \rfloor^2$ iterations in the worst case.

Table 2
Inter-relations within a sample Learning Community

				Learner 1	Learner 2	Learner 3	Learner 4	Learner 5	Learner 6
				10000	01000	00111	00110	11000	01100
Learning objects	1	10111	R	0.75000	1.000000	0.500000	0.666667	1.000000	1.000000
			GR	0.60000	1.000000	0.200000	0.400000	0.750000	0.750000
	2	11100		0.500000	0.500000	1.000000	0.666667	0.333333	0.333333
				0.400000	0.400000	0.666667	0.500000	0.200000	0.200000
	3	00111		0.750000	0.750000	0.000000	0.333333	1.000000	0.666667
				0.750000	0.750000	0.000000	0.200000	1.000000	0.500000
	4	11001		0.500000	0.500000	1.000000	1.000000	0.333333	0.666667
				0.400000	0.400000	0.666667	1.000000	0.200000	0.500000
	5	10100		0.250000	0.500000	0.500000	0.333333	0.333333	0.333333
				0.200000	0.500000	0.333333	0.250000	0.250000	0.250000
	6	01110		0.750000	0.500000	0.500000	0.333333	0.666667	0.333333
				0.750000	0.400000	0.250000	0.200000	0.500000	0.200000
	7	00110		0.500000	0.500000	0.000000	0.000000	0.666667	0.333333
				0.500000	0.500000	0.000000	0.000000	0.666667	0.250000
	8	11000		0.250000	0.250000	1.000000	0.666667	0.000000	0.333333
				0.200000	0.200000	1.000000	0.666667	0.000000	0.250000
	9	10000		0.000000	0.250000	0.500000	0.333333	0.000000	0.333333
				0.000000	0.250000	0.500000	0.333333	0.000000	0.333333
	10	01100		0.500000	0.250000	0.500000	0.333333	0.333333	0.000000
				0.500000	0.200000	0.333333	0.250000	0.250000	0.000000

Table 3
Consistent learning compatibility matrices

∂^R	∂^{GR}	LCM density
1.0–0.8	All values	No consistent solution
0.7	1.0–0.8	No consistent solution
	0.7	9
	0.6	11
	0.5 - 0.0	12
0.6	1.0 - 0.8	No consistent solution
	0.7	9
	0.6	13
	0.5	14
	0.4	18
	0.3–0.0	19

Table 4
A solution

Learners	Learning objects
1, 5	1, 3, 6
2, 6	1, 3, 4
3, 4	2, 4, 8

Solution specifications
$\partial^R = 0.6$
$\partial^{GR} = 0.4$
Density = 18

3. Perform Clustering on LC Matrix. The cell formation problem is a combinatorial optimization problem that is NP-hard, therefore using some heuristic method for solving it may be wise. The advantage of matrix formulation is that the solution procedures are computationally efficient (Lin et al., 1996).

Among various array-based clustering methods (Joines et al., 1996), Rank Order Clustering (ROC) is the better known of the array-based clustering algorithms and has been widely applied in various scientific fields. ROC has a computational complexity of $O(mn(m + n))$ where m and n are the numbers of rows and columns (Kusiak, 1985). This can be reduced to $O(mn(\log(mn)))$, by using the proper sorting algorithms (Boubezari and Bozena, 1996).

ROC2 and MODROC are extensions to ROC, the first aiming to deal with the problem of storage, and the latter to reduce the computational effort as well, by using the so-called block-and-slice approach, that is continuously reducing the number of rows and columns in each iteration. Both have led to drastic improvement in ROC results (Offodile et al., 1994; Onwubolu, 1998).

Additionally, problems like "bottleneck machines" (that cell formation faces, the ROC method included) do not exist in the case of Educational Cells, because of the nature of the problem itself.

Of course, one must not rule out the possibility that other methods, for example the Bond Energy Analysis (BEA), may perform equally or even better (Chu and Tsai, 1990).

For the demonstration needs of our paper we will use MODROC2, a variation adapted for the purpose by combining MODROC and ROC2. ROC2 rearranges columns and rows using the word-value of each column and row. This is done repeatedly until the matrix remains unchanged, thus produces a matrix in which both columns and rows are arranged in order of decreasing lexicographic value.

MODROC afterwards, identifies the cluster formed in the upper left part of the matrix and removes the corresponding columns (one may remove both columns *and* lines if one wishes the groups to be assigned with completely different sets of LOs). Then, ROC2 is reapplied to

the sub-matrix. This is done until no columns are left in the matrix.

Using this procedure produces clusters as shown in Table 4.

5. Further applicability issues: towards the formation of most valuable distance learning environments

In the preceding sections we have presented a method that outputs groups of learners with similar educational background, together with a corresponding set of Learning Objects for each group. These sets are serving each group as their Course's Core (CC) towards the overall common objective (Fig. 2).

The method takes into account the learners' knowledge and the available learning objects content, and can be easily automated, thus providing support and advice to tutors and learners. Yet, there are a few points arising towards
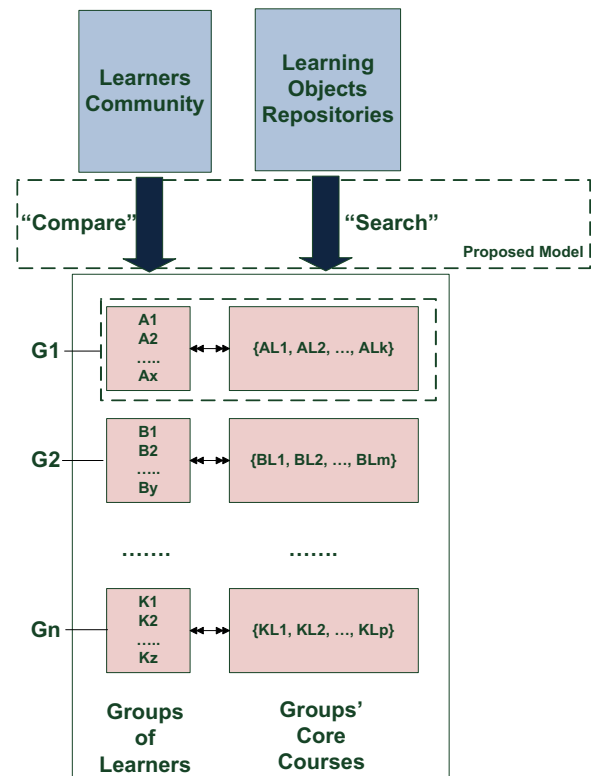

Fig. 2. The educational cells.

integrating the model with the current e-learning trends, and we will try to provide some answers, together with some possible extensions.

(a) *Learning Objects referencing*. It has been argued that a Learning Object can be characterized by three sets of properties (Mavrommatis, 2006b):
  - Content properties: the skills, concepts, objectives it present,
  - Quantitative properties: anything that can be presented by a numeric 'cost', like size, level of difficulty, expected studying time, etc.,
  - Qualitative properties: the type of instruction (lecture, case study), the media the LO utilizes (graphic, video, text, audio, java program), and so on.

We have already used the Content properties for our model. Having defined all three categories, one may add more criteria to his search by simply adding corresponding restrictions to the problem formulation. Adding some extra criteria reduces the search space, and therefore makes solution to the problem easier.

(b) *The learning environment*. Learning environments of the future will be comprised of distributed repositories of LOs from which, *via adequate search sub-environments*, users from all parts of the world will be able to attain specific knowledge, at specific time and based on his/her specific needs (Veen, 2005; Bonk et al., 2004).

It is widely recognized that modern web environments fit perfectly with the ideas of Constructivism (Savery and Duffy, 1995; Miller and Miller, 1999): learning must have an overall purpose, that is made clear to the learners; process control is transferred to the learner; learners use the content as information resources rather than teaching material; learners have the freedom to browse all over the content in order to be able to view multiple perspectives, gradually constructing their knowledge; collaboration is crucial; learners must be challenged by a loosely structured, open and free learning environment; freedom must not leave learner helpless though, or they will get lost: for that purpose, both automated procedures and teacher–facilitator guidance may be used.

The learning environment that the proposed method fits perfectly in, is a dynamic, mostly asynchronous one. People are grouped, assigned some conceptually proper learning objects, going through them at their own pace and sequence, searching the repositories for more content, being guided by their tutor, being assigned collaborative tasks, supported by a proper group formation.

Undoubtedly, learning is contextual. A model has to be context-free, though.

(c) *"Full course" building*. As we have already mentioned, Course's Core (CC) is a result of compromise, aiming to balance several criteria and, of course,

achieve effective clustering. The method "as is", therefore, selects sets of LOs that may
  - cover certain new properties more than once. This is fully compatible with constructivist environments, since it provides the learner with various aspects of the subject.
  - cover some properties that a learner is already familiar with. Although this may be sometimes desirable, we are trying to reduce redundancies, as already discussed in chapter 3. If a learner wishes, (s)he may skip this part of content, though.
  - leave uncovered some properties that a student needs to learn. This can be easily resolved by having the learner or the system select one or more extra LOs covering these specific skills.

The proposed algorithm selects learners that are, as much as possible, similar; selects content that is, as much as possible, the same for each group's member. Of course, it is highly anticipated that some students within the same group may need to study more than others, but this happens all the time. After all, this is part of adaptation.

As it comes with sequencing, the content does not have to be presented it in a strict, predetermined, hierarchically built way (the expert's representation of knowledge, as produced by our task analysis): sequencing varies as learners construct their own knowledge (McGuire, 1996) by going through the learning objects in any order, decided separately by each student.

The system, of course, may be using the task analysis results to *recommend* one or more appropriate pathways within the content. Additionally a learner may use the search capabilities of the environment to find more 'similar' LOs. For an analytic, strict sequencing method, and a similarity measure between Learning Objects see (Mavrommatis, 2006a).

(d) *Collaboration experiences*. According to Paquette (2004), in order for a highly distributed, flexible educational environment to be efficient, collaboration has to be well-coordinated, conducted mainly via asynchronous interactions among learners, with a number of, strategically fitted, synchronous events. The proposed model may be used for the formation of all three general types of groups (Gross Davis, 1993), depending mainly upon tutor's decision and the complexity of the subject matter (Fig. 3)

1. *informal learning groups*, that is temporary clusters within a class session;
2. *formal learning groups*, lasting from a single class session to several weeks;
3. *study teams*, the long-term groups usually working over the course.

There is no clear consensus whether the collaborating groups should be homogeneous or heterogeneous (Zurita et al., 2005). In (Johnson et al., 2000) ten cooperative learning methods are reported as having received the most attention. Several methods require heterogeneous group
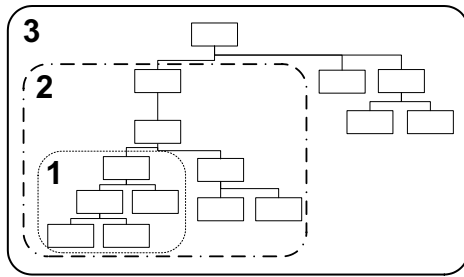
Fig. 3. A group's lifecycle.

members, some require homogeneity, and other methods form their groups in both manners during the various stages of the learning process.

Jigsaw (Aronson and Patnoe, 1997) requires heterogeneous group formation, but also, uses temporary homogeneous groups, the so-called *expert groups*.

In Team Assisted Individualization (TAI) (Slavin et al., 1986) students receive instructions from the tutor in small homogeneous groups and then practice in heterogeneous teams *at their own pace and individualized sequence on materials appropriate to their specific needs* (as described in the US Department of Education Website).

In Learning Together (Johnson and Johnson, 1999) students work in small heterogeneous teams. They cooperate to achieve a definite goal by sharing material and ideas.

The proposed model in the first place creates groups of students in which their learning state is as close as possible. Depending on the collaboration method a tutor uses, then one may create heterogeneous groups by choosing one or more students from each of the initially created Groups $G_1, G_2, \ldots, G_n$ (Fig. 2).

6. Conclusions and future research

Distance learning technologies and e-learning methods seem to have a common denominator: using interconnected information systems, which through special programs and platforms bring users/learners in contact with courses and teachers (LCMS). E-learning technologies are already in relatively mature stage, while research in the area advances in what lately is known as Learning Objects.

This paper's main goal is to automate learning groups and environments by matching educational means and objectives. It is obvious that one of the most important parameters of collaboration is the proper formation of the groups of learners. In addition, this paper tries to show that, after the emergence of the Learning Objects notion, the next step needed today is proper use of well-known and widely accepted tools: learning theories, instructional design, information retrieval, taxonomy, etc. By these means we may be able to create modern, pluralistic, integrated educational environments, moving one step beyond the – somehow restrictive- environment that is currently offered by LCMSs.

The presented information space is comprised of equally weighted constituent skills, so that each skill either exists in the relevant vector or not. Thus, in classifying LOs we consider all included/presented concepts/skills with same weight '1'. In other situations where we might have more than one Learning Objects (sometimes, many more than two) with the same vector properties, we need to re-rate it in order to have detailed classification, leading to a less number of possible solutions. Thus, it is possible to use the model proposed in this paper in such a way to assign a 'weight' (instead of '0–1') to every included skill in information space which will describe e.g., its importance as a subject of learning or its priority for each learner.

Furthermore, in order to reduce the possible solutions provided by our proposed model, we can select LOs that cover the different needs of every learner/user, e.g., required study time, level of difficulty, etc, as already mentioned.

Also, future research on similar models of forming user-based distance learning environments (UBDLEs) should additionally focus on providing more criteria to interrupt the selection algorithms used in such models and identify which criteria are more reliable than others. For example, our criterion was the formation of a matrix with certain density but there are others, also significant, such as the coverage rate of specific skills required by users, the amount of duplication of knowledge already possessed by the user, etc.

This is relevant to another point under research, namely the quality of solution given by the algorithm. Quite a few quantitative performance measures have been developed in Group Technology to evaluate the final cell formation solutions (Yin and Yasuda, 2005). Some of them may be good for a start, but the problem of defining one or more metrics, especially adapted to the educational technology, is still open. For example, a metric that would, somehow, make use of the "properties" that need to be learned by all the members of a collaboration group might be a good idea for our purpose. After having defined such a metric, next step will almost naturally be the testing of the algorithm on a large sample that would also draw conclusions about the sensitivity of results to parameters.

Finally, among the different domains using learning objects, university, corporate and military are of the most importance (Collis and Strijker, 2004). All three include collaboration into the learning process and objectives achievement. Additionally, all three, especially the corporate and military establishments, use LOs to keep records of their employees' Knowledge, Skills and Attributes (KSAs). Thus, there is a major and challenging research field unfolding ahead of us and this paper has tried to be part of such advancement. In other words, it is the integration of Information Technology and Education with Human Resources and Management, as parts of the Semantic Web (Hendler et al., 2002) in order to advance to the era of the Global Knowledge Society.

References

Alvarado-Boyd, S., 2003. Traveler's Guide to the Learning Object Landscape. NMC: The New Media Consortium.

Aronson, E., Patnoe, S., 1997. The Jigsaw Classroom, second ed. Longman, NY.

Bekele, R., 2005. Computer-Assisted Learner Group Formation Based on Personality Traits. Ph.D. dissertation, University of Hamburg.

Bonk, C., Kim, K.-J., Zeng, T., 2004. Future directions of blended learning in higher education and workplace learning settings. In: Bonk, C.J., Graham, C.R. (Eds.), Handbook of Blended Learning: Global Perspectives, Local Designs. Pfeiffer Publishing, San Francisco, CA.

Boubezari, S., Bozena, K., 1996. A new reconfigurable test vector generator for built-in self-test applications. Journal of Electronic Testing: Theory and Applications 8, 153–164.

Bredeweg, B., Forbus, K., 2003. Qualitative modeling in education. AI Magazine 24 (4), 35–46.

Brusilovsky, P., 1998. Adaptive educational systems on the world-wide-web: A review of available technologies. In: Proceedings of workshop WWW-Based Tutoring at 4th International Conference on Intelligent Tutoring Systems (ITS'98).

Brusilovsky, P., 2004. KnowledgeTree: A distributed architecture for adaptive e-learning. May 17–22, 2004, New York, New York, USA. ACM 1-58113-912-8/04/0005.

Chu, C.-H., Tsai, M., 1990. A comparison of three array-based clustering techniques for manufacturing cell formation. International Journal of Production Research 28 (8), 1427–1433.

Collis, B., Strijker, A., 2004. Technology and human issues in reusing learning objects. Journal of Interactive Media in Education (4), Special Issue on the Educational Semantic Web <www-jime.open.ac.uk/2004/4>.

Cronjé, J., 2006. Paradigms regained: Toward integrating objectivism and constructivism in instructional design and the learning sciences. Educational Technology Research and Development 54 (4), 387–416.

Daley, B., 2004. Using concept maps in qualitative research. In: Cañas, A.J., Novak, J.D., González, F.M. (Eds.), Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain 2004.

Fung, A., Yeung, J., 2000. An object model for a web-based adaptive educational system. In: ICEUT2000, Educational Uses of Communication and Information Technologies, Beijing, China, August 21–25, 2000.

Gagné, R.M., Wager, W.W., Briggs, L.J., 1992. Principles of instructional design, 4th ed. Holt, Rinehart and Winston, New York.

Gaudioso, E., Boticario, J.G., 2003. Towards web-based adaptive learning communities. In: Proceedings of Artificial Intelligence in Education. IOS Press, Sydney, Australia.

Gross Davis, B., 1993. Tools for Teaching. Jossey-Bass Inc., SF, California, ISBN 1-55542-568-2.

Hendler, J., Berners-Lee, T., Miller, E., 2002. Integrating applications on the semantic web. Journal of the Institute of Electrical Engineers of Japan 122 (10), 676–680.

HFI DTC, 2007. Human Factors Integration Defence Technology Centre, <http://www.hfidtc.org/WP2/wp2background.htm> (accessed April, 1st, 2007).

Impedovo, S., Lucchese, M.G., Modugno, R., 2006. Progress in e-learning: toward the Knowledge Based Society. In: Proceedings of the 6th WSEAS International Conference on Distance Learning and Web Engineering, Lisbon, Portugal, September 22–24, 2006.

Johnson, D., Johnson, R., 1999. Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning. Allyn & Bacon, ISBN 978-0205287710.

Johnson, D., Johnson, R., Stanne, M.B., 2000. Cooperative Learning Methods: A Meta-Analysis, University of Minnesota retrieved May 2006, <http://www.co-operation.org/pages/cl-methods.html>.

Joines, J.A., King, R.E., Culbreth, C.T., 1996. A comprehensive review of production-oriented manufacturing cell formation techniques. International Journal of Flexible Automation and Integrated Manufacturing 3 (3–4), 225–265.

Jonassen, D., Tessmer, M., Hannum, W., 1999. Task Analysis Methods for Instructional Design. Lawrence Erlbaum Associates, London.

Kester, L., Kirschner, P., Van Merrienboer, J.J.G., Baumer, A., 2001. Just-in-time information presentation and the acquisition of complex cognitive skills. Computers in Human Behavior 17, 373–391.

Knight, G.P., Bohlmeyer, E.M., 1990. Cooperative learning and achievement: Methods for assessing causal mechanisms. In: Sharan, S. (Ed.), Cooperative Learning – Theory and research. Praeger, New York, NY, pp. 1–22.

Kusiak, A., 1985. The part families problem in flexible manufacturing systems. Annals of Operations Research 3, 279–300.

Lin, T., Dessouky, M., Kumar, K., Ng, S., 1996. A heuristic-based procedure for the weighted production cell formation problem. IIE Transactions 28, 579–589.

Mavrommatis, G., 2006a. Learning objects and objectives towards automatic learning construction. European Journal of Operational Research 187 (3), 1449–1458.

Mavrommatis, G., 2006b. Machine-supported Instructional design. IMECS 2006, In: International MultiConference of Engineers and Computer Scientists 2006, Hong Kong, 20–22 June, 2006, ISBN-13: 978-988-98671-3-3.

Mazza, R., Dimitrova, V., 2004. Visualising student tracking data to support instructors in web-based distance education. In: Proceedings of 13th International Conference on World Wide Web (WWW'04), pp. 154–161, retrieved <http://www.www2004.org/proceedings/docs/2p154.pdf>.

McGuire, E., 1996. Knowledge representation and construction in hypermedia environments. Telematics and Informatics 13 (4), 251–260.

Merrill, M.D., 2001. Toward a theoretical tool for instructional design. Instructional Science 29 (4–5), 291–310.

Merrill, M.D., 2002. First principles of instruction. Educational Technology Research and Development 50 (3), 43–59.

Milam, J., Santo, S., Heaton, L., 2000. Concept Maps for Web-Based Applications, ERIC Technical Report, University of Virginia.

Miller, S., Miller, K., 1999. Using instructional theory to facilitate communication in web-based courses. Educational Technology & Society 2 (3).

Moallem, M., 2001. Applying constructivist and objectivist learning theories in the design of a web-based course: Implications for practice. Educational Technology & Society 4 (3). Available online <http://ifets.ieee.org/periodical/vol_3_2001/moallem.html>.

Mödritscher, F., Garcia-Barrios, V.M., Gutl, C., 2004. The past, the present and the future of adaptive e-learning. An approach within the scope of the research project AdeLE. In: Proceedings ICL 2004, Villach, Austria, 2004.

Muehlenbrock, M., 2006. Learning group formation based on learner profile and context. International Journal on E-learning 5 (1), 19–24.

Novak, J.D., Cañas, A.J., 2006. The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01, Florida Institute for Human and Machine Cognition, 2006. Available from: <http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>.

Offodile, F., Mehrez, A., Grznar, J., 1994. Cellular manufacturing: A taxonomic review framework. Journal of Manufacturing Systems 13 (3), 196–220.

Onwubolu, G., 1998. Redesigning jobshops to cellular manufacturing systems. Integrated Manufacturing Systems 9 (6), 377–382.

Paquette, G., 2004. Instructional Engineering for Learning Objects Repositories Networks. In: CALIE 04: International Conference on Computer Aided Learning in Engineering Education, Invited paper.

Paramythis, A., Loidl-Reisinger, S., 2004. Adaptive learning environments and e-learning standards. Electronic Journal of E-Learning 2 (1), 181–194.

Pollalis, Y., 1996. Effective integration of information management and organizational behavior principles. Journal of Education for MIS 4 (1).

Sarker, B., 1996. The resemblance coefficients in group technology: A survey and comparative study of relational metrics. Computers & Industrial Engineering 30 (I), 103–116.

Savery, J., Duffy, T., 1995. Problem based learning: An instructional model and its constructivist framework. Educational Technology 35 (5), 31–38.

SCORM, 2004. Advanced Distributed Learning (ADL), Sharable Content Object Reference Model (SCORM®) 2004 Overview, 2004.

Slavin, R., Leavey, M., Madden, N., 1986. Team Accelerated Instruction: Mathematics. Charlesbridge, Watertown, MA.

Soller, A., 2001. Supporting social interaction in an intelligent collaborative learning system. International Journal of Artificial Intelligence in Education 12, 40–62.

Soller A., Lesgold, A., 2003. A computational approach to analyzing online knowledge sharing interaction, Amy Soller and Alan Lesgold. In: 11th International Conference on Artificial Intelligence in Education, Sydney Australia, 2003 (AIED 2003).

Stanton, N., 2006. Hierarchical task analysis: Developments, applications, and extensions. Applied Ergonomics 37 (1), 55–79.

Tapscott, D., 1998. Growing Up Digital: The Rise of the Net Generation. McGraw Hill, New York.

van Merriënboer, J.J.G., 1997. Training Complex Cognitive Skills: A Four-component Instructional Design Model for Technical Training. Englewood Cliffs, New Jersey.

Van Rosmalen, P., Brouns, F., Tattersall, C.,Vogten, H. Van Bruggen, J, Sloep, P., Koper, E.J.R., 2004. Towards an open framework for adaptive, agent-supported e-learning. International Journal of Continuing Engineering Education. <http://hdl.handle.net/1820/76>.

van Zele, E., Lenaerts, J., Wieme, W., 2004. Improving the usefulness of concept maps as a research tool for science education. International Journal of Science Education 26 (9), 1043–1064.

Veen, W., 2005. 2020 Visions/Wim Veen's Projection, Prof. Wim Veen, Delft University of Technology, The Netherlands (Online Educa Berlin 2005 Special Site).

Walker, J.M.T., King, P.H., 2002. In: Proceedings of the 2002 American Society for Engineering Education Annual Conference and Exposition.

Wiley, D., 2003. Learning Objects: Difficulties and Opportunities, online, retrieved at <http://wiley.ed.usu.edu/docs/lo_do.pdf>.

Wilson, B. (Ed.), 1996. Constructivist Learning Environments: Case Studies in Instructional Design. Educational Technology Publications, New Jersey.

Wilson, B., 1997. Reflections on constructivism and instructional design. In: Dills, C.R., Romiszowski, A.A. (Eds.), Instructional Development Paradigms. Educational Technology Publications, Englewood Cliffs, NJ.

Yin, Y., Yasuda, K., 2005. Similarity coefficient methods applied to the cell formation problem: a comparative investigation. Computers & Industrial Engineering 48, 471–489.

Zurita, G., Nussbaum, M., Salinas, R., 2005. Dynamic grouping in collaborative learning supported by wireless handhelds. Educational Technology & Society 8 (3), 149–161.