Mathematizing the Process of Analogical Problem Solving

Michael Gr. Voskoglou
Graduate Technological Educational Institute of Patras
voskoglou@teipat.gr, mvosk@hol.gr

Abstract
Much of our cognitive activity depends on our ability to reason analogically. When we encounter a new problem we are often reminded of similar problems solved in past and may use the solution procedure of an old problem to solve the new one (analogical problem solving). In this paper we develop two mathematical models for the description of the process of analogical problem solving. The first one is a stochastic model constructed by introducing a finite, ergodic Markov chain on the steps of the analogical reasoning process. Through this we obtain a measure of the solvers’ difficulties during the process. The second is a fuzzy model constructed by representing the main steps of the process as fuzzy subsets of a set of linguistic labels characterizing the individuals’ performance in each of these steps. In this case we introduce the Shannon’s entropy (total probabilistic uncertainty) - properly modified for use in a fuzzy environment - as a measure of the solvers’ performance. The two models are compared to each other by listing their advantages and disadvantages. Classroom experiments are also performed to illustrate their use in practice.

Keywords: Analogical Reasoning, Problem Solving, Finite Markov Chains, Fuzzy Sets

1 Introduction

Analogical Reasoning (AR) is a method of processing information that compares the similarities between new and past understood concepts, then using these similarities to gain understanding of the new concept. The basic intuition behind AR is that when there are substantial parallels across different situations there are likely to be further parallels. AR is ubiquitous in human cognition. Analogies are used in explaining concepts which cannot directly perceived (e.g. electricity in terms of the water flow), in making predictions within domains, in communication and persuasion, etc. Within cognitive science mental processes are likened to computer programs (e.g. neural networks) and such analogies serve as mental models to support reasoning in new domains. AR is important in general in creativity and scientific discovery. In Artificial Intelligence the Case-Based-Reasoning paradigm covers a range of different methods (including analogical reasoning) for organizing, retrieving, utilizing and indexing the knowledge of past cases (Voskoglou 2010, 2011b).

Solution of problems by analogy (analogical problem solving) is a special case of the general class of AR. However this strategy can be difficult to implement in problem solving, because it requires the solver to attend to information other than the problem to be solved (target problem). Thus the solver may come up empty-handed, either because he has not solved any similar problems in past, or because he fails to realize the relevance of previous problems. But, even if an analogue is retrieved, the solver must know how to use it to determine the solution procedure for the target problem.

Several studies (Holyoak 1985, Genter & Toupin 1986, Novick 1988, Genter & Markman 1997, Voskoglou 2003, etc) have provided detailed models for the process of AR which are broadly consistent with reviews of problem solving strategy training studies, in which factors associated with instances of successful transfer – that is, use of already existing knowledge to produce new knowledge - are identified. According to these studies the main subprocesses (steps) involved in AR include:

- **Representation** of the target problem.
- **Search-retrieval** of a related past problem.
- **Mapping** of the representations of the target and the related problem.
- **Adaptation** of the solution of the related problem for use with the target problem.

More specifically, before solvers working on a problem they usually construct a representation of it. A good representation must include both the surface and the structural (abstract, solution relevant) features of the problem. The former are mainly determined by what are the quantities...
involved in the problem and the latter by how these quantities are related to each other. The features included in solvers' representations of the target problem are used as retrieval cues for a related problem in memory (called a source problem). When the two problems share structural but not surface structures the source is called a remote analogue of the target problem (Holyoak 1984). Analogical mapping requires aligning the two situations – that is, finding the correspondences between the representations of the target and the source problem – and projecting inferences from the source to the target. Once the common alignment and the candidate inferences have been discovered the analogy is evaluated. The last step involves the adaptation of the solution of the analogous problem for use with the target problem, where the correspondences between objects and relations of the two problems must be used.

The successful completion of the above process is referred as positive analogical transfer. But the search may also yield distractor problems having surface but not structural (solution relevant) common features with the target problem and therefore being only superficially similar to it. Usually the reason for this is a non satisfactory representation of the target problem, containing only its salient surface features, and the resulting consequences on the retrieval cues available for the search process. When a distractor problem is considered as an analogue of the target, we speak about negative analogical transfer. This happens if a distractor problem is retrieved as a source problem and the solver fails, through the mapping of the representations of the source and target problem, to realize that the source cannot be considered as an analogue to the target. Therefore the process of mapping is very important in analogical problem solving playing the role of a "control system" for the fitness of the source problem.

In this paper we introduce two mathematical models (a stochastic and a fuzzy one) describing the analogical problem solving process and we present classroom experiments illustrating the use of these models in practice. This is an important contribution in the area of mathematical modelling and application resulting from research and relevant experiences in educational and classroom environment.

2 The stochastic model

Mathematics does not explain the natural behaviour of an object, it simply describes it. This description however is so much effective, so that an elementary mathematical equation can describe simply and clearly a relation, that in order to be expressed with words could need entire pages. During the last two decades (1991-2011) we have worked in creating mathematical (stochastic and fuzzy) models for the better description and understanding of several processes appearing in the areas of Mathematical Education, Artificial Intelligence (Case-Based Reasoning) and Management (Voskoglou 2011a). Continuing this research we shall present in this article a stochastic and a fuzzy model for the description of the process of AR.

For the development of our stochastic model we assume that the AR process has the Markov property. This means that the probability of entering a certain step at a certain phase of the process, although it is not necessarily independent of previous phases, it depends at most on the step occupied in the previous phase. Our assumption is a simplification (not far away from the truth) made to the real system that enables the formulation of it to a form ready for mathematical treatment (assumed real system, e.g. see Voskoglou 2007; section 1). In fact, we introduce a finite Markov chain on the steps of the AR process described above. The states of this chain are: \( s_1 \) = representation, \( s_2 \) = search-retrieval, \( s_3 \) = mapping, \( s_4 \) = adaptation and \( s_5 \) = solution of the target problem. For general facts on Markov chains we refer freely to Kemeny & Snell (1976).

The starting state is always \( s_1 \). When the AR process is completed at \( s_5 \) it is assumed that a new problem is given for solution and therefore the process restarts from \( s_1 \). After the construction of the target problem’s representation the solvers proceed from \( s_1 \) to \( s_2 \). Being at \( s_2 \) and facing difficulties in finding a source problem they may return to \( s_1 \) asking for more information from problem’s representation. Then they proceed again to \( s_2 \) to continue the AR process.

After the retrieval of a source problem the solvers proceed from \( s_2 \) to \( s_3 \). If the source is considered to be analogous to the target problem, then they transfer from \( s_3 \) to \( s_4 \). Otherwise they return to \( s_2 \) searching for a new source problem. Notice that solvers who finally fail to retrieve an analogue
through the mapping process cannot proceed further. Therefore they return to \( s_1 \) waiting for a new problem to be given for solution.

After the adaptation of the solution of the source for use with the target problem the solvers proceed to the final state \( s_5 \) of the solution of the target problem. On the contrary, if during the adaptation process they realize that the source is in fact a distractor problem, they return to \( s_2 \) searching for a new source. Solvers who finally fail to adapt the solution of the source for use with the target problem they return from \( s_4 \) to \( s_1 \) waiting for a new problem to be given for solution. According to the above description the “flow-diagram” of the AR process is that shown in Figure 1.

![Figure 1: The “flow-diagram” of the process of AR](image)

Denote by \( p_{ij} \) the transition probability from state \( s_i \) to \( s_j \), for \( i,j=1,2,3,4,5 \). According to the above diagram the transition matrix of the chain is:

\[
A = \begin{bmatrix}
      s_1 & s_2 & s_3 & s_4 & s_5 \\
      0 & 1 & 0 & 0 & 0 \\
      p_{21} & 0 & p_{23} & 0 & 0 \\
      p_{31} & p_{32} & 0 & p_{34} & 0 \\
      p_{41} & p_{42} & 0 & 0 & p_{45} \\
      1 & 0 & 0 & 0 & 0 
\end{bmatrix}
\]

Obviously we have that \( p_{21}+p_{23}=p_{31}+p_{32}+p_{34}=p_{41}+p_{42}+p_{45}=1 \) (1)

It becomes also evident that in our Markov chain it is possible to go between any to states, not necessarily in one step, i.e. it is an \textit{ergodic chain}. For an ergodic chain it is well known that, as the number of its phases tends to infinity (\textit{long run}), the chain tends to an \textit{equilibrium situation} characterized by the equality \( P=PA \) (2), where \( P=[p_1, p_2, p_3, p_4, p_5] \) is the \textit{limiting probability vector} of the chain. The entries of \( P \) give the probabilities for the chain to be in each of its states in the long run. Obviously we have that \( p_1 + p_2 + p_3 + p_4 + p_5 = 1 \) (3).

From relation (2) one gets easily the following equations:

\[
\begin{align*}
p_1 &= p_2 p_{21} + p_3 p_{31} + p_4 p_{41} + p_5 \\
p_2 &= p_1 + p_3 p_{32} + p_4 p_{42} \\
p_3 &= p_2 p_{23} \\
p_4 &= p_3 p_{34}
\end{align*}
\]
\[ p_5 = p_4 p_{45} \]

Adding the first four of the above equations and using relation (1) one finds the fifth equation, which therefore is equivalent with the others. Solving the linear 5X5 system of the first four equations and of equation (3) by the Cramer’s rule one finds that

\[ p_j = \frac{(p_{31} - 1)(p_{41} - 1) + p_{23} p_{32} (p_{41} - 1) - p_{23} p_{42} p_{34}}{D}, \quad p_2 = \frac{(p_{31} - 1)(p_{41} - 1)}{D}, \quad p_3 = \frac{-p_{23}(p_{41} - 1)}{D}, \quad p_4 = \frac{p_{23} p_{34}}{D}, \quad p_5 = \frac{-p_{23} p_{34}(p_{41} - 1) - p_{23} p_{34} p_{42}}{D} \]

where \( D = (2+p_{23})(p_{31} - 1)(p_{41} - 1) + p_{23}(2p_{32} - 1)(p_{41} - 1) + p_{23} p_{42} (1 - 2p_{42}) \) is the determinant of the system. Further, it is well known that in an ergodic chain the mean number of times in state \( s_i \) between two successive occurrences of \( s_j \) is given by \( \frac{p_{i}}{p_{jj}} \) (Kemeny & Snell 1976; Theorem 6.2.3). Therefore, since the process starts again after state \( s_5 \) (as a new problem is given for solution), the mean number of times between two successive occurrences of \( s_5 \) is given by \( m = \sum_{i=1}^{4} \frac{p_i}{p_5} = \frac{1 - p_5}{p_5} \). The value of \( m \) is an indicator of the solvers’ difficulties during the AR process. Another indicator is the time spent for the solution of each problem. However, assuming that the time available for the solution of each problem is prefixed, it becomes evident that \( m \) is a measure for solvers’ difficulties during the AR process. The bigger is \( m \) the more the solvers’ difficulties during the AR process.

AN APPLICATION IN CLASSROOM

In order to illustrate the use of the above model in practice we performed the following experiment, where the subjects were students of the graduate Technological Educational Institute of Patras, Greece, being at their second term of studies. We formed two groups, with 20 students of the School of Management and Economics in the first and 20 students of the School of Technological Applications (prospective engineers) in the second group.

Four mathematical problems were given for solution to both groups (see Appendix). The first of these problems was related to combinatorial analysis and probability, the second was related to the theory of matrices and the third one was an application of derivatives to economics. All the above are topics in the students’ first term course of mathematics. The fourth was a problem involving arithmetic i.e., an already developed at school ability. In each case and before receiving the target problem students received two other problems together with their solution procedures. They read each problem and its solution procedure and then solved the problem themselves using the given procedure. Subjects were allowed 10 minutes for each problem and they were not given the other problem until after 10 minutes had elapsed. The first of these problems was a remote analogue to the target problem, while the other was a distractor problem. Next the target problem was given and was asked from the subjects to try to solve it by adapting the solution of one of the previous problems (time allowed 20 minutes). Our instructions stressed the importance of showing all of one's work on paper and emphasized that we were interested in both correct and incorrect solution attempts.

Examining students’ papers after the end of the experiment we calculated the following means:

- 4.2 students from the first group faced difficulties in retrieving a source problem, but they came through after looking back to their representations of the target problem (5.1 students from the second group).
- 15.1 students from the first group considered through the mapping the collected source as an analogue to the target problem, while the rest of them (4.9 students) searched for a new source. Finally 3.7 from the 4.9 students considered the new source as an analogue to the
target problem, while the rest of them (1,2 students) failed to retrieve an analogue through the mapping process (14,8 and 1,6 students from the second group respectively).

- Thus 15,1+3,7=18,8 students from the first group proceeded finally to the step of adaptation (14,8+1,6=16,4 students from the second group). From these students 11,1 adapted successfully the solution of the analogue for use with the target problem, while 1,5 students (who had previously considered both the remote analogue and the distractor as source problems) failed to do so.(12,1 and 1,5 students from the second group respectively). The rest (6,2 students from the first and 2,8 from the second group) returned to $s_2$ to retrieve a new source and through $s_3$ they came back to $s_4$.

- Finally 3,3 from these 6, 2 students of the first group adapted successfully the solution of the analogue for use with the target problem and 2,9 failed to do so (1,6 and 1,2 students from the second group respectively). Thus 11,1+3,3=14,4 students from the first and 12,1+1,6=13,7 students from the second group solved finally the target problems.

Figure 2: The “movements” of the students of the first group

The “movements” of the students of first group are shown in Figure 2. We observe that we have 35,3 in total “arrivals” to $s_2$ and 31,1 “departures” from $s_2$ to $s_3$, therefore $p_{23}=\frac{31,1}{35,3}\approx0,881$. In the same way one finds that $p_{21}=\frac{3,35}{9,4}\approx0,119$, $p_{32}=\frac{1,31}{9,4}\approx0,038$, $p_{33}=\frac{4,9}{31,1}\approx0,158$, $p_{34}=\frac{25}{14,4}\approx0,804$, $p_{41}=\frac{4,4}{25}\approx0,176$, $p_{42}=\frac{6,2}{25}\approx0,248$ and $p_{43}=\frac{14,4}{25}\approx0,576$.Eng
Replacing the values of the transition probabilities in the formulae of the model we find that the limiting probability vector for the first group is \( P \approx [0, 157 \ 0, 231 \ 0, 232 \ 0, 121] \) and that \( m \approx 7,264 \) times.

Operating the analogous calculations for the second group we find that \( P \approx [0, 154 \ 0, 237 \ 0, 23 \ 0, 119] \) and \( m \approx 7,404 \) times.

The elements of \( P \) give the several probabilities about the “behaviour” of each group during the AR process. Also, since 7,264<7,404, the performance of the first group was slightly better.

According to the design of our experiment students had to choose the source problem between two given problems: A remote analogue to the target and a distractor problem. However, often things are not so simple. In fact, the individuals have usually to search in their memories to retrieve the source among several past problems sharing common surface and/or structural characteristics with the target. We could of course add in our experiment one or more problems among the candidate source problems. Nevertheless, this manipulation would make the calculation of the transition probabilities between states of the chain more complicated, because the students’ movements would be extended to several directions.

3 The fuzzy model

We shall develop now a fuzzy model by representing the main steps of the AR process as fuzzy subsets of a set of linguistic labels characterizing the individuals’ performance in each of these steps. For general facts on fuzzy sets we refer freely to Klir & Folger (1988). For basic definitions and concepts the reader may also look at Voskoglou (2009; section 1). In order to make the development of our model technically simpler we shall consider the step of representation of the target problem as a sub step of the search-retrieval of the source problem.

Let us consider a group of \( n \) analogical problem solvers, \( n \geq 2 \). Denote by \( A_i, i=1,2,3 \) the steps of search-retrieval, mapping and adaptation respectively. Denote also by \( a, b, c, d, \) and \( e \) the linguistic labels of negligible, low, intermediate, high and complete success respectively of the analogical problem solvers in each of the \( A_i \)’s.

Set \( U = \{a, b, c, d, e\} \) and let \( n_{ia}, n_{ib}, n_{ic}, n_{id} \) and \( n_{ie} \) denote the number analogical problem solvers who faced negligible, low, intermediate, high and complete success at step \( A_i, i=1,2 \)

Transition across levels in the process of learning: A fuzzy model,3. We define the membership function \( m_{Ai} \) for each \( x \) in \( U \), as follows:

\[
m_{Ai}(x) = \begin{cases} 
1, & \text{if } \frac{4n}{5} < n_{ix} \leq n \\
0.75, & \text{if } \frac{3n}{5} < n_{ix} \leq \frac{4n}{5} \\
0.5, & \text{if } \frac{2n}{5} < n_{ix} \leq \frac{3n}{5} \\
0.25, & \text{if } \frac{n}{5} < n_{ix} \leq \frac{2n}{5} \\
0, & \text{if } 0 \leq n_{ix} \leq \frac{n}{5}
\end{cases}
\]

Then \( A_i \) is represented as a fuzzy subset of \( U \) by \( A_i = \{(x, m_{Ai}(x)) : x \in U\}, i=1, 2, 3 \).

In order to represent all possible solvers’ profiles (overall states) during the AR process we consider a fuzzy relation, say \( R \), in \( U^3 \) of the form

\[
R = \{(s, m_R(s)) : s=(x, y, z) \in U^3\}.
\]

Since the degree of solvers’ success at a certain step depends upon the degree of their success in the previous step and in order to determine properly the membership function \( m_R \) we give the following definition:

A profile \( s=(x, y, z) \), with \( x, y, z \) in \( U \), is said to be well ordered if \( x \) corresponds to a degree of success equal or greater than \( y \), and \( y \) corresponds to a degree of success equal or greater than \( z \). For
example, \((c, c, a)\) is a well ordered profile, while \((b, a, c)\) is not. We define now the membership degree of a profile \(s\) to be
\[
m_R(s) = m_A(x)m_A(y)m_A(z),
\]
if \(s\) is well ordered, and 0 otherwise. In fact, if for example profile \((b, a, c)\) possessed a nonzero membership degree, how it could be possible for a solver, who failed at the step of mapping, to perform satisfactorily at the step of adaptation?

Next, for reasons of brevity, we shall write \(m_s\) instead of \(m_R(s)\). Then the possibility of the profile \(s\) is defined by
\[
\pi(s) = \max\{m_s\},
\]
where \(\max\{m_s\}\) denotes the maximal value of \(m_s\), for all \(s\) in \(U^3\).

In other words the possibility \(\pi\) expresses the "relative membership degree" of \(s\) with respect to \(\max\{m_s\}\). Calculating the possibilities of all profiles one obtains a qualitative view of the group’s performance during the AR process.

Further, the amount of information obtained by an action can be measured by the reduction of uncertainty resulting from this action. Accordingly the individuals’ uncertainty during the AR process is connected to their capacity in obtaining relevant information. Therefore a measure of uncertainty could be adopted as a measure of the group’s abilities in AR. For example, such a measure that we have used in an earlier paper, when developing an analogous fuzzy model for the process of learning (Voskoglou 2009) is the total possibilistic uncertainty \(T\) of the group.

Another measure of (probabilistic) uncertainty and the associated information was established by Shannon (1948). When expressed in terms of the Dempster-Shafer mathematical theory of evidence, this measure takes the form
\[
H = \frac{1}{\ln n} \sum_{s=1}^{n} m_s \ln m_s,
\]
where \(n\) is the total number of elements of the corresponding fuzzy set (Klir 1995; p.20), and it is called the Shannon entropy or the Shannon-Wiener diversity index. In the above formula the sum is divided by \(\ln n\) in order to normalize \(H\), so that its maximal value is 1 regardless the value of \(n\).

Adopting \(H\) as a measure of the group’s abilities in AR it becomes evident that the lower is the value of \(H\) (i.e. the higher is the reduction of the corresponding uncertainty), the better the group’s abilities. An advantage of adopting \(H\) as a measure instead of \(T\) is that \(H\) is calculated directly from the membership degrees of all profiles \(s\), in contrast to \(T\) that presupposes the calculation of the possibilities of all profiles first.

Assume now that one wants to study the combined results of behaviour of \(k\) different groups, \(k \geq 2\), during the same process. For this we introduce the fuzzy variables \(A_1(t), A_2(t)\) and \(A_3(t)\) with \(t=1, 2, ..., k\). The values of these variables represent fuzzy subsets of \(U\) corresponding to the steps of the AR process for each of the \(k\) groups; e.g. \(A_1(2)\) represents the fuzzy subset of \(U\) corresponding to the step of search-retrieval for the second group (\(t=2\)).

It becomes evident that, in order to measure the degree of evidence of combined results of the \(k\) groups, it is necessary to define the possibility \(r(s)\) of each profile \(s\) with respect to its membership degrees for all groups. For this, we introduce the pseudo-frequencies \(f(s)\)
\[
= \sum_{t=1}^{k} m_t(s)\quad \text{and we define} \quad r(s) = \frac{f(s)}{\max\{f(s)\}}, \quad \text{where} \quad \max\{f(s)\}\quad \text{denotes the maximal pseudo-frequency.}
\]
Obviously the same method could be applied when one wants to study the combined results of behaviour of a group during different analogical problem solving processes.

**AN APPLICATION OF THE FUZZY MODEL**

The classroom experiment described in the previous section was repeated a few days later with two different groups of 20 students of the Technological Educational Institute of Patras being at their second term of studies. The only difference was that this time we added one more problem among the candidate source problems in each of the four cases, which was unrelated to the target problem in terms of both their surface and structure features. For this problem we followed the same
process with the other problems, i.e. we gave a solution procedure to students and we allowed 10 minutes to solve it themselves by using the given procedure.

Our characterizations of students’ performance at each step of the AR process involved:

- Negligible success, if they didn’t obtain (at the particular step) positive results for the given problems.
- Low success, if they obtained positive results for 1 only of the given problems.
- Intermediate success, if they obtained positive results for 2 problems.
- High success, if they obtained positive results for 3 problems.
- Complete success, if they obtained positive results for all the given (4 in total) problems.

Examining students’ papers of the first group we found that 9, 6 and 5 students had intermediate, high and complete success respectively at the step of search-retrieval in terms of choosing the correct problem (i.e. the remote analogue to the target) as the source problem. This means that \( n_{1a}=n_{1b}=0, n_{1c}=9, n_{1d}=6 \) and \( n_{1e}=5 \). Thus, according to the definition of \( m_{A_1}(x) \), the step of search-retrieval corresponds to a fuzzy subset of \( U \) of the form: 
\[
A_1 = \{(a,0),(b,0),(c, 0.5),(d, 0.25),(e, 0.25)\}.
\]

In the same way we represented the steps of mapping and adaptation as fuzzy subsets of \( U \) by 
\[
A_2 = \{(a,0),(b,0),(c, 0.5),(d, 0.25),(e,0)\} \quad \text{and} \quad A_3 = \{(a, 0.25),(b, 0.25),(c, 0.25),(d,0),(e,0)\}
\]

Next, we calculated the membership degrees of the 5\(^3\) (ordered samples with replacement of 3 objects taken from 5) in total possible students’ profiles (see column of \( m_s(1) \) in Table 1). For example, for \( s=(c, c, a) \) one finds that
\[
m_s = m_{A_1}(c) \cdot m_{A_2}(c) \cdot m_{A_3}(a) = (0.5)\cdot(0.5)\cdot(0.25) = 0.06225.
\]

It turned out that \( (c, c, a) \) was one of the profiles possessing the maximal membership degree and therefore the possibility of each \( s \) in \( U^3 \) is given by 
\[
r_s = \frac{m_s}{0.06225}.
\]

Using this formula we calculated the possibilities of all profiles (see column of \( r_s(1) \) in Table 1).

Finally we calculated the Shannon entropy in terms of the values of column \( m_s(1) \) in Table 1, where \( n=125 \) and we found that \( H=0.289 \).

Working as above for the second group we found that 
\[
A_1 = \{(a,0),(b,0.25),(c, 0.5),(d, 0.25),(e,0)\},
\]
\[
A_2 = \{(a, 0.25),(b, 0.25),(c, 0.5),(d,0),(e,0)\} \quad \text{and} \quad A_3 = \{(a, 0.25),(b, 0.25),(c,0.25),(d,0),(e,0)\}.
\]

The membership degrees of all possible profiles of the second group are shown in column of \( m_s(2) \) of Table 1. It turned out that the maximal membership degree was again 0.06225, therefore the possibility of each \( s \) is calculated by the same formula as for the first group. The possibilities of all profiles are shown in column of \( r_s(2) \) of Table 1, while for the Shannon entropy we found that \( H=0.312 \). Thus, since 0.289<0.312, the general performance of the first group was slightly better.

Next, in order to study the combined results of behaviour of the two groups, we introduced the fuzzy variables \( A_i(t) \), \( i=1, 2, 3 \) and \( t=1, 2 \), as we have described them in the model. Then the pseudo-frequency of each student profile \( s \) is given by \( f(s) = m_{A_1}(1) + m_{A_2}(2) \) (see corresponding column in Table 1). It turns out that the highest pseudo-frequency is 0.124 and therefore the possibility of each student’s profile is given by \( r(s) = \frac{f(s)}{0.124} \). The possibilities of all profiles having non-zero pseudo-frequencies are presented in the last column of Table 1.

4 Discussion and conclusions

In this paper we developed two mathematical models for the description of the process of AR:
A stochastic model by introducing a finite ergodic Markov chain on the steps of the AR process and a fuzzy model by representing the main steps of the AR process as fuzzy subsets of a set of linguistic labels characterizing the individuals’ performance in each of these steps. Both models give important quantitative information about the abilities of a group of analogical problem solvers”. In the stochastic
model the calculation of the transition probabilities between states of the chain is getting more complicated when the source must be chosen among more than two given problems, because the individuals’ “movements” in this case are extended to more directions. On the contrary, there is not any particular difficulty in this case with the fuzzy model. Moreover the fuzzy model gives a qualitative view of the group’s performance through the calculation of the possibilities of all individuals’ profiles during the AR process. Finally, an additional advantage of the fuzzy model is that it gives to the researcher the opportunity to study the combined results of the behaviour of two or more groups during the AR process or alternatively to study the combined results of the behaviour of the same group during different analogical problem solving processes. On the other hand the characterization of the analogical problem solvers’ performance in terms of a set of linguistic labels which are fuzzy by themselves is a disadvantage of the fuzzy model, because this characterization depends on the researcher’s personal criteria (see for example in section 3 the criteria used in our experiments for characterizing the students’ performance). Therefore the combined use of the two models seems to be the best solution in achieving a worthy of credit mathematical analysis of the AR process.

References


Voskoglou, M. Gr. (2009), Transition across levels in the process of learning: A fuzzy model, *International Journal of Modelling and Application*, University Blumenau, Brazil, 1, 37-44.


Table 1: Profiles with non zero pseudo-frequencies

<table>
<thead>
<tr>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
<th>$m_1$</th>
<th>$r_1$</th>
<th>$m_2$</th>
<th>$r_2$</th>
<th>$f(s)$</th>
<th>$r(s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>b</td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>b</td>
<td>a</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>c</td>
<td>0.062</td>
<td>1</td>
<td>0.062</td>
<td>1</td>
<td>0.124</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>a</td>
<td>0.062</td>
<td>1</td>
<td>0.062</td>
<td>1</td>
<td>0.124</td>
<td>1</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>c</td>
<td>a</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>c</td>
<td>b</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>c</td>
<td>b</td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>a</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>b</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>c</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>a</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>b</td>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>b</td>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.258</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>d</td>
<td>c</td>
<td>a</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.5</td>
<td>0.062</td>
<td>0.5</td>
</tr>
<tr>
<td>d</td>
<td>c</td>
<td>b</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.5</td>
<td>0.062</td>
<td>0.5</td>
</tr>
<tr>
<td>d</td>
<td>c</td>
<td>c</td>
<td>0.031</td>
<td>0.5</td>
<td>0.031</td>
<td>0.5</td>
<td>0.062</td>
<td>0.5</td>
</tr>
<tr>
<td>e</td>
<td>c</td>
<td>a</td>
<td>0.031</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>e</td>
<td>c</td>
<td>b</td>
<td>0.031</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>e</td>
<td>c</td>
<td>c</td>
<td>0.031</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.031</td>
<td>0.25</td>
</tr>
<tr>
<td>e</td>
<td>d</td>
<td>a</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>e</td>
<td>d</td>
<td>b</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
<tr>
<td>e</td>
<td>d</td>
<td>c</td>
<td>0.016</td>
<td>0.258</td>
<td>0</td>
<td>0</td>
<td>0.016</td>
<td>0.129</td>
</tr>
</tbody>
</table>

(The outcomes of the table are written with accuracy up to the third decimal point)

APPENDIX: Problems given to the classroom experiments

CASE 1

Target problem: A box contains 8 balls numbered from 1 to 8. One makes three successive drawings, putting back the corresponding ball to the box before the next drawing. Find the probability of getting all the balls drawing out of the box different to each other.

The probability is equal to the quotient of the total number of the ordered samples of 3 objects from 8 (favourable outcomes) to the total number of the corresponding samples with replacement (possible outcomes).

Remote analogue: How many numbers of 2 digits can be formed by using the digits from 1 to 6 and how many of them have their digits different?

Solution procedure given to the students: Find the total number of the ordered samples of 2 objects from 6 with and without replacement respectively.

Distractor problem: A box contains 3 white, 4 blue and 6 black balls. If we draw out 2 balls, what is the probability to be of the same colour?

Solution procedure given to the students: The number of all favourable outcomes is equal to the sum of the total number of combinations of 3, 4 and 6 objects taken 2 at each time respectively, while the number of all possible outcomes is equal to the total number of combinations of 13 objects taken 2 at each time.
Unrelated problem (used only with the fuzzy model): Find the number of all possible anagrammatisms of the word "SIMPLE". How many of them start with S and how many of them start with S and end with E?

Solution procedure given to the students: The number of all possible anagrammatisms is equal to the total number 6! of permutations of 6 objects. The anagrammatisms starting with S are 5! And the anagrammatisms starting with S and ending with E are 4!

CASE 2

Target problem: Consider the matrices:

A = \[
\begin{bmatrix}
1 & -\hat{a} & -\hat{a} \\
0 & 1 & -\hat{a} \\
0 & 0 & 1 \\
\end{bmatrix}
\]

for all integers k, and B = \[
\begin{bmatrix}
0 & -\hat{a} & -\hat{a} \\
0 & 0 & -\hat{a} \\
0 & 0 & 0 \\
\end{bmatrix}
\]

Prove that \(A^n = A + (n-1)(B + \frac{n}{2}B)\), for every positive integer \(n\).

Since \(A=I+B\), where I stands for the unitary 3X3 matrix, and \(B^3 = 0\), is \(A^n=(I+B)^n=I+nB+\frac{n(n-1)}{2}B^2=I+nB+\frac{n(n-1)}{2}B=A+(n-1)(B+\frac{n}{2}B)\).

Remote analogue: Let \(\alpha\) be a nonzero real number. Prove that \(\alpha^n = \sum_{i=0}^{n} \binom{n}{i} (\alpha-1)^i\), for all positive integers \(n\).

Solution procedure given to the students: Write \(\alpha = 1+(\alpha-1)\) and apply the Newton’s formula \((x+b)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} b^i\), setting \(x=1\) and \(b=\alpha-1\).

Distractor problem: If A and B are as in the target problem, calculate \((A+B)^2\).

The students were asked to operate the corresponding calculations.

Unrelated problem (used only with the fuzzy model): Prove that \(1+2+\ldots+n=\frac{n(n+1)}{2}\), for all positive integers \(n\).

The students were asked to apply induction on \(n\).

CASE 3

Target problem: The price of sale of a good depends upon its total demand \(Q\) and it is given by \(P(Q) = \frac{1}{2}Q-50\), while the cost of production of the good is given by \(C(Q) = \frac{1}{4}Q^2 + 35Q + 25\). Find the quantity \(Q\) of the good’s total demand maximizing the profit from sale.

The revenue from sale is equal to \(P(Q)Q\) and therefore the profit from sale is given by \(K(Q) = P(Q)Q - C(Q)\). The maximum of function \(K(Q)\) is calculated by using the well known theorem of derivatives.

Remote analogue: A car is entering to a road having initial speed 50 Km/h, which is changed according to the relation \(U(t)=3t^2-12t+50\), where \(t\) represents the time (in minutes) during which the car is moving on this road. Find the minimal speed of the car on this road.

The students were asked to apply the well known theorem of derivatives in order to calculate the minimum of the function \(U(t)\).

Distractor problem: The price of sale of a good depends upon its total demand \(Q\) and it is given by \(P(Q) = 25-Q^2\). The price is finally fixed to 9 monetary units and therefore the consumers who would be willing to pay more than this price benefit. Find the total benefit to consumers (Dowling 1980, paragraph 17.7: Consumer’s surplus).
Solution procedure given to the students: For \( P=9 \) and since \( Q \geq 0 \), it turns out that \( Q=4 \). Drawing the graph of the function \( P(Q) \) (parabola) it is easy to observe that the total benefit to consumers is equal to \( \int_{0}^{4} P(Q)dQ = 4.9 \) monetary units.

**Unrelated problem** (used only with the fuzzy model): Find the area under the curve \( y=4x^2+2 \).

Solution procedure given to the students: The area is given by \( \int (4x^2+2)dx \).

**CASE 4**

**Target problem**: A producer has a stock of wine greater than 500 and less than 750 kilos. He has estimated that, if he had the double quantity of wine and transfused it to bottles of 12, or 25, or 40 kilos, it would be left over 6 kilos at each time. Find the quantity of the stock.

If \( Q \) is the quantity of stock, then, since the lowest common multiple of 12,25 and 40 is 600, \( 2Q-6 \) is a multiple of 600, therefore \( 2Q=606 \), or \( 2Q=1212 \), or \( 2Q=1818 \), etc. But \( 500<Q<750 \), therefore \( Q=603 \) kilos.

**Remote analogue**: An employer occupies less than 50 workers. If he occupied the triple number of workers and 3 more, then he could distribute them in bands of 8 or 12 or 15 workers. How many workers he occupies?

Solution procedure given to the students: If \( x \) is the number of workers, then, since the lowest common multiple of 8, 12 and 15 is 120, \( 3x+3 \) is a multiple of 120, or \( x+1 \) is a multiple of 40.

**Distractor problem**: A producer has a stock of 3400 and 5025 kilos respectively of two different kinds of wine and he decides to distribute these quantities to the maximal possible number of customers. After this distribution, they remained 25 kilos from each kind of wine in his barrels. How many of his customers he succeeded to satisfy with this manipulation?

Solution procedure given to the students: The number of customers is equal to the greatest common divisor of 3400-25 and 5025-25.

**Unrelated problem** (used only with the fuzzy model): The number of students of a school is between 300 and 400. When they tried marching in rows of 10 the last row had 9 students, while when they tried marching in rows of 9 the last row had 7 students. How many are the students?

Solution procedure given to the students: Let \( a=100x+10y+z \) be the number of students of the school. Then \( a=10t-1 \) for some positive integer \( t \). Therefore \( z=9 \) and \( x=3 \). Further \( a=9s+7 \) for some positive integer \( s \), or \( a-7=9s \). But, since 9 divides \( a-7 \), 9 divides also the sum of the digits of \( a-7 \), i.e. \( (3+y+9)-7=9k \) for some positive integer \( k \), or \( y+5=9k \). But \( 0<y \leq 9 \), therefore \( y=4 \).