Markov Chain and Classification of Difficulty Levels Enhances the Learning Path in One Digit Multiplication

Behnam Taraghi, Anna Saranti, Martin Ebner, and Martin Schön

Graz University of Technology, Münzgrabenstr. 35/I, 8010 Graz, Austria
{b.taraghi,martin.ebner}@tugraz.at,
s0473056@sbox.tugraz.at, mrtnschn@googlemail.com

Abstract. In this work we focus on a specific application named "1x1 trainer" that has been designed to assist children in primary school to learn one digit multiplications. We investigate the database of learners' answers to the asked questions by applying Markov chain and classification algorithms. The analysis identifies different clusters of one digit multiplication problems in respect to their difficulty for the learners. Next we present and discuss the outcomes of our analysis considering Markov chain of different orders for each question. The results of the analysis influence the learning path for every pupil and offer a personalized recommendation proposal that optimizes the way questions are asked to each pupil individually.

Keywords: Learning Analytics, One digit multiplication, Knowledge discovery, Math, Markov chain, Primary school.

1 Introduction

Generally, recommender algorithms use the implicit data generated through monitoring the users' interactions with the underlying system to understand better the hidden users' preferences. Once the users' interests are known, the system can provide personalized recommendations to the users that best suites their needs and interests.

Recommender systems in e-learning applications follow the same approach [1]. The analysis of chronological user activities or user traces (sometimes called user navigation) helps to provide personalized recommendations to the users [2].

Learning applications in particular can benefit from such an approach too. Duval [3] pointed out that we have to think about learners' traces and their learning efforts. Siemens and Baker [4] defined learning analytics as the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning as well as the environments in which it occurs.

Graz University of Technology has been developing math trainers since 2010 with the aim to improve the basic math education for primary schools [5]. First of all the 1x1 trainer [6] was implemented followed by the multi-math-coach [7] as well as the addition / subtraction trainer [8]. All applications can be used for free at the URL: http://mathe.tugraz.at.

P. Zaphiris and A. Ioannou (Eds.): LCT 2014, Part I, LNCS 8523, pp. 322–333, 2014.

In primary school, learning the one digit multiplication table is one of the major goals in the four-year lasting education. The learning problem seems to be trivial at first glance, but by studying the literature several difficulties unfold: Language implications in general [9], the role of math as first non-native language [10], pure “row learning” [11] etc.

Therefore a web-based application was developed which on the one side can assist the learning process of the pupils and on the other side can enhance the pedagogical intervention of the teachers. The full implementation, the intelligent algorithm and the first results are described in [6].

Several educational, pedagogical and psychological surveys classify various pupils’ common errors in one digit multiplications. One common finding is the problem size effect [12][13]. Large multiplications such as 9*8 tend to have a higher error rate than smaller ones, such as 2*3. Therefore this set of questions is assumed to be more difficult to learn. [14], predict that errors in simple multiplication are more probable, if they contain the same digit as the correct result. Some studies investigated on easy and difficult groups of questions and denoted patterns of easy questions to learn, such as doubles, times five and square numbers [15, 16, 17]. However we could find no previous work dealing with the problem of one digit multiplication table computationally and analytically.

We have already made a first computational analysis on the provided dataset, using a Markov chain model and classification algorithms, to discover some common structures within the answers of the pupils. These structures help to better understand the pupils’ behavior especially when they answer to the set of difficult questions. The results and the analysis are published in [18]. In this paper we present in more detail the clustering algorithm that is used to identify different difficulty classes in one digit multiplication problems. Furthermore, the outcomes of the analysis from our new Markov model are presented and discussed. The results are then used to influence the learning path within the application and hence provide a proposal for basic recommendation of the asked question’s sequence.

2 Methodology – Markov Chains

Markov chains are used to model stochastic processes such as navigation models. There are many works investigating on Web navigation and human navigation patterns on World Wide Web such as [19, 20, 21]. Another example is the Random Surfer model in Google’s PageRank algorithm that can be seen as a special case of a Markov chain [22], where Web pages are represented as states and hyperlinks as probabilities of navigating from one page to another. In our model the answer types to each question represent the states and the probabilities to the answer type of the subsequent same question in the sequence as transition links between states. In this section we introduce the Markov chains formally.

A finite discrete Markov chain of the first order is a sequence of random variables $X_1, X_2, X_3, \ldots, X_n$ for which the following Markov property holds:

$$P\,(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P\,(X_{n+1} = x_{n+1} \mid X_n = x_n). \quad (1)$$

We assume that the probabilities do not change as a function of time, hence the Markov chains are time-homogeneous.

A Markov chain of order k is described formally as follows:

$$\begin{gathered} P\,(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, X_n = x_n) = \\ P\,(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_{n-k+1} = x_{n-k+1}). \end{gathered} \quad (2)$$

The Markov chain of first order is characterized as memoryless, meaning that the next state depends only on the current state. Considering the Markov chain of order k, the probability of the next state depends on the k previous states.

The Markov model is represented as a matrix P of all stochastic transition probabilities between the states. Hence for n states, the Matrix P is of size n*n. Each row in the matrix represents the stochastic transition probabilities from one state to all the other states. As a result the sum of probabilities within a row is always 1.0.

3 Answer Types

In this paper we perform our analysis on the dataset from the database of “1x1 trainer” application. The application puts each question to the pupils at least two times. Based on whether the submitted answers in the user history are correct or not, the answers for each question are classified into one of six different answer types. If a pupil answers a question for the first time correctly (answer type R that stands for RIGHT), the same question is asked once again later to ensure that the answer is truly known by the pupil. If a question is answered correctly for the second time (answer type RR), the application assumes that the user had already truly known the answer to the asked question. In contrast, if the pupil answers the question incorrectly in the second round (answer type RW) the application keeps asking the same question later on till the pupil answers it correctly (answer type WR). Answer type W (stands for WRONG) implies the first incorrect answer to a question. Answer type WW implies an incorrect answer to a question after a preceding wrong answer.

Table 1 lists these six defined answer types and their definitions. The following example illustrates how the answer types are assigned to each given answer. Assuming the application has asked a pupil the question 9*3 5 times in his history and the pupil’s answers have been as follows: 27, 24, 26, 27, 27. The assigned answer types for this set of answers would be: R, RW, WW, WR, RR. The defined answer types build the states of the Markov chain model in our analysis.

Table 1. Six different answer types and their definitions. "R" stands for "Right" and "W" for "Wrong"

Answer type	Definition	Preceding answer	Current answer
R	First correct answer	-	R
W	First wrong answer	-	W
RR	Correct answer given a preceding correct answer	R	R
RW	Wrong answer given a preceding correct answer	R	W
WR	Correct answer given a preceding wrong answer	W	R
WW	Wrong answer given a preceding wrong answer	W	W

4 Difficult Questions

In this section we begin with the analysis of probabilities and reaction times to identify the questions that are most difficult for the pupils. The probabilities of the occurring answer types in the dataset reveal the most difficult questions. To identify the difficult questions most efficiently we divided the dataset to two subsets. The first subset includes only the R and W answer types. These are the questions that are put to the pupils by the application and answered by the pupils for the first time. The goal is to identify the questions that were mostly already known (hence easy) and those that were mostly already unknown (hence difficult) to the pupils before a learning process actually begins within this application. The second subset includes only RW, WR and WW answer types. These are the unknown questions answered by the pupils at least for the second time. The goal is to identify the most difficult as well as easiest questions within this subset of data. These are the questions that the pupils have repeated the most and least times till they got the correct answer.

4.1 Subset of R and W Answer Types

Figures 1 illustrates the questions that are quite easy (known by a high percentage of pupils) and the ones that are rather difficult (unknown by a high percentage of pupils). The 1x1 trainer application provides 90 multiplication problems (1*9 to 10*9), hence 10 rows and 9 columns in the heatmap. The heatmap illustrates that the multiplications where 1, 2, 5, and 10 occur as operands can be classified as easy or most known questions, whereas 3, 4, 6, 7, 8 and 9 as operands build multiplications that can be classified as difficult or most unknown. From the first view it could be inferred that the pupils knew most of the questions beforehand. 7*8 and 6*8 are the first two questions that the most number of pupils have difficulties.

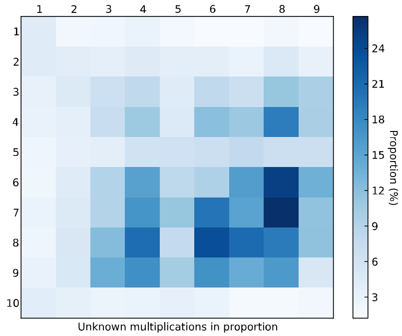

Fig. 1. Heatmap of the asked unknown questions (multiplications) proportionally within the subset of R and W answer types. The rows correspond to the first operand of the multiplication and the columns correspond to the second operand.

If we consider the reaction times consumed by the pupils - especially for the set of known (easy) answers (R answer type) - we can observe that the pupils need more time for the identified difficult unknown questions than for the identified easy ones. Figure 2 illustrates this result in a heatmap. It shows the average time consumption for each question individually from the set of known questions. This observation confirms our results about the identified difficult and easy questions.

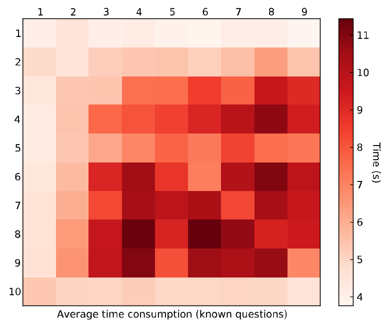

Fig. 2. Heatmap of the average time consumption from the set of asked known questions (multiplications) within the subset of R and W answer types. The rows correspond to the first ope and of the multiplication and the columns correspond to the second operand.

4.2 Subset of RW, WR and WW Answer Types

Figure 3 illustrates the histogram of the 30 most difficult questions within this subset. 6*8 and 7*8 are again the first two questions that have the highest probabilities. Figure 4 is a heatmap that illustrates the questions that are quite easy to learn (low probabilities) and the ones that are rather difficult (high probabilities). It is can be seen again that the multiplications where 1, 2, 5, and 10 occur as operands can be classified as easy to learn questions whereas 3, 4, 6, 7, 8 and 9 as operands build multiplications that can be classified as difficult.

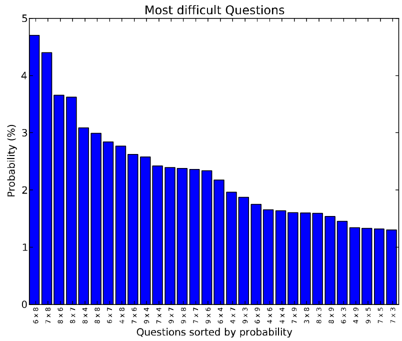

Fig. 3. Histogram of the most 30 difficult questions within the subset of RW, WR and WW answer types.

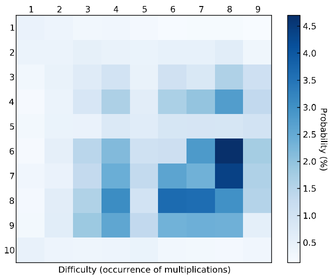

Fig. 4. Heatmap of the probabilities of questions (multiplications) within the subset of RW, WR and WW answer types. The rows correspond to the first operand of the multiplication and the columns correspond to the second operand.

Considering the reaction times within this subset for correct answers (WR answer types) we can observe the same results as we did in the first subset (R and W answer types). Figure 5 illustrates this result in a heatmap. It shows the average time consumption for each question individually from the set of correct answers (WR answer type). This observation reinforces our beliefs about the identified difficult and easy questions.

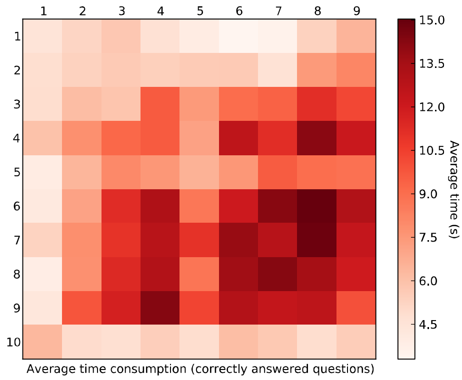

Fig. 5. Heatmap of the average time consumption from the set of correct answered question (multiplications) within the subset of RW, WR and WW answer types. The rows correspond to the first operand of the multiplication and the columns correspond to the second operand.

5 Classification of Questions

The goal is to classify the questions into different difficulty levels. The categorization bases on one dimensional data that represent the occurrence probabilities of answer types within the observed dataset. Beginning with the hypothesis that there are three difficulty levels, we used the k-means algorithm [23] to compute the three clusters.

Figure 6 shows one possible hard classification of the data. Each data point is assigned to a specific cluster; the decision boundaries between them are linear. The algorithm is sensitive to the configuration of its initial iteration i.e. the assumption about the position of the cluster means. It is suggested that one runs the k-means algorithms several times with different starting cluster means (usually randomly chosen samples from the dataset). Due to the fact that we had prior knowledge and interpretation for our data, each time we ran the algorithm we divided the sample's values in three equal intervals. In a second step we picked randomly a sample (that had a value in a particular interval) to be the initial mean point of its corresponding cluster. Different runs in our training data reveal that the number of questions that are classified as difficult varies between four to eight. Figure 6 shows the first difficulty level identified by 3-means algorithm. It contains eight questions that own the highest probabilities within the dataset. The second difficulty level contains twenty two questions. The biggest cluster contains the set of easiest questions involving sixty questions with the least probabilities.

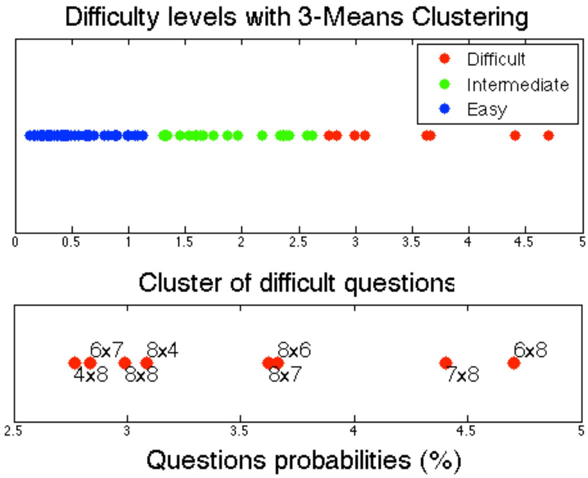

Fig. 6. Clusters representing the difficulties of questions as computed by the 3-means algorithm. The upper figure demonstrates the three identified difficulty levels. The lower figure shows the questions within the first level difficulty cluster.

In a second step we tried to identify the optimal number of clusters that best classifies one digit multiplication questions with respect to their difficulty. We ran the k-means algorithm using different number of clusters. The choice of the initial mean of each cluster was made in the same manner as in the three clusters case. Due to the fact that our training data contain only ninety samples and we don't have a test set, we used k-fold cross-validation to provide an estimate of the test error [24]. The quality measure that was used to indicate a potential improvement in our choice between different number of clusters was the averaged cumulative distance. The algorithm stops when the cumulative distance converges to prevent overfitting. The algorithm suggests that the optimal number of clusters is six. The cluster representing the most difficult problems contains 4 questions (level one) followed by 10 questions in level two.

Figure 7 illustrates the six identified difficulty levels.

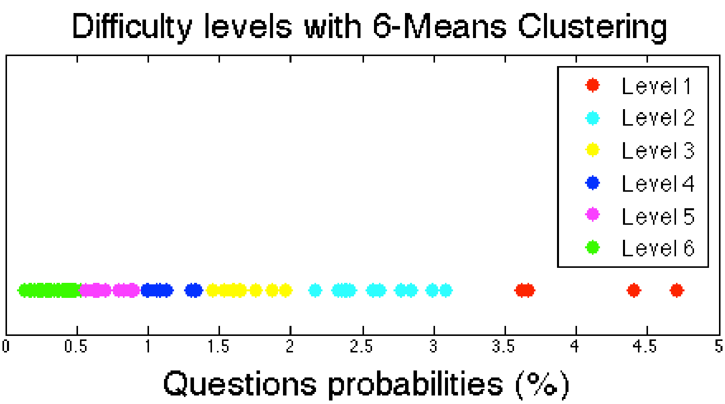

Fig. 7. Clusters representing the difficulty levels of questions as computed by the 6-means algorithm.

6 Markov Chain Analysis

In our Markov chain model the answer types to each question represent the states. The probability to the answer type of the subsequent same question in the sequence characterizes the transition link between these states. In other words, for each of the ninety questions we apply our Markov chain model individually.

Figure 8 illustrates the three dimensional plot of the Markov chain of six orders and all questions. For k >=2 the plots show the portion of each last (k-th) answer type for each question individually. As expected, most of the pupils answer the questions correctly once they get the questions for the first time. This can be observed in the first plot (high proportion of answer type R for k = 1). In the second round (k = 2) the majority of answer types are RR, which means that the pupils mostly answer correctly once they get the same question for the second time. The proportion of R and W answer types comes to 0% while k >=2. This is in accordance with our definition of R and W answer types. They imply the correct and wrong answers to a question for the first time (without any preceding answer). A common observation over all questions is that the proportion of answer type RW decreases whereas the proportion of answer types WR and WW increases in k >=4. This occurs because a question that is asked repeatedly (ascending k) has a higher probability of not being answered correctly in the past (preceding steps). Looking precisely to each question individually we can observe a remarkable difference in the proportion of k-th answer type in different difficulty levels. This proportion value acts as a measure in our recommendation proposal to weight each question depending on k (the step in which the application must decide which question to put as next).

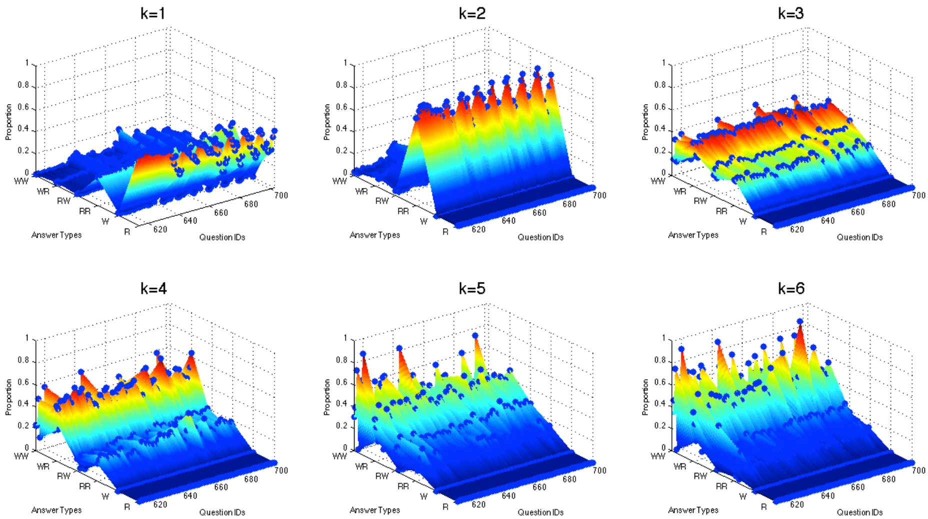

Fig. 8. Markov chains of all questions for k = 1 to 6. The x axis represents the question ids. The y axis represents the last (k-th) answer type within the chain of the length k. The z axis represents the proportion of last (k-th) answer type for each question and order k.

7 Future Work

The goal of our research was to develop applications for basic math education, which allow individualized learning. Each child should be assisted on its own and personal way. The data analyses presented in this publication will help to improve the application in two different ways: the current empirical estimated difficulties of the question have to be adapted to the difficulty levels we have identified through this work. Assuming a pupil answers the question X from the difficulty level n correctly. The next question (Y) should be selected from the difficulty level m whereas either m=n (the same) or m=n+1 (the next higher difficulty level n+1). If a pupil answers the question X from the difficulty level n incorrectly, the next question (Y) should be selected from the difficulty level m whereas either m=n (the same) or m=n-1 (the preceding lower difficulty level n-1). The proportion measure p from the analysis of Markov chains introduced in this work will be used as selection criterion from the set of new difficulty level m. For each candidate question Y within the set of difficulty level m and step k (the k-th time the pupil will answer to the question Y) the question Y that owns the highest proportion rate referring to answer type WR or RR will be selected. The chosen answer type depends on the preceding answer type (W or R for step k-1) to that question (Y).

8 Conclusion

In this work we analyzed the dataset from “1x1 trainer” application that was designed for primary school children to learn one-digit multiplications. We identified the easiest and the most difficult questions by looking through the probabilities of different answer types of the pupils in two different subsets. The reaction time of the pupils for answering the questions was also taken into consideration. The result from both data sets was almost the same. The multiplications where 1, 2, 5, and 10 occur as operands can be classified as easy to learn, whereas 3, 4, 6, 9, and especially 7, 8 operands build multiplications that can be classified as difficult. We classified the questions into three difficulty levels (difficult, intermediate and easy) using k-means algorithm. We gained eight difficult, twenty two intermediate and sixty easy questions totally. The identified class of difficult questions contains the following eight multiplications: 6*8, 7*8, 8*6, 8*7, 8*4, 8*8, 6*7 and 4*8. As can be seen, the difficult set is characterized mainly by the operand 8.

Next we identified an optimal number of clusters that best classifies one digit multiplication questions in respect to their difficulty. The algorithm suggests that six clusters can optimally represent the different difficulty levels of the questions.

Our Markov model analysis for each question leads to a measure that can be used in the recommendation proposal to improve the question selection algorithm in the application.

References

1. Linton, F., Schaefer, H.P.: Recommender systems for learning: Building user and expert models through long-term observation of application use. User Modeling and User-Adapted Interaction 10, 181–207 (2000)
2. Taraghi, B., Grossegger, M., Ebner, M., Holzinger, A.: Web Analytics of user path tracing and a novel algorithm for generating recommendations in Open Journal Systems. Online Information Review 37(5), 672–691 (2013)
3. Duval, E.: Attention Please! Learning Analytics for Visualization and Recommendation. In: Proceedings of LAK 2011: 1st International Conference on Learning Analytics and Knowledge (2011)
4. Siemens, G., Baker, R.S.J.: Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In: LAK 2012 Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, pp. 252–254 (2012)
5. Ebner, M., Schön, M.: Why Learning Analytics in Primary Education Matters! Bulletin of the Technical Committee on Learning Technology 15(2), 14–17 (2013); Karagiannidis, C., Graf, S. (eds.)
6. Schön, M., Ebner, M., Kothmeier, G.: It's Just About Learning the Multiplication Table. In: Shum, S.B., Gasevic, D., Ferguson, R. (eds.) Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK 2012), pp. 73–81. ACM, New York (2012), doi:10.1145/2330601.2330624
7. Ebner, M., Schön, M., Taraghi, B., Steyrer, M.: Teachers Little Helper: Multi-Math-Coach. In: Nunes, M.B., McPherson, M. (eds.) Proceedings of the IADIS International Conference e-Learning 2013, pp. 183–190. IADIS Press, Prague (2013)
8. Ebner, M., Schön, M., Neuhold, B.: Learning Analytics in basic math education – first results from the field. eLearning Papers 36, 24–27 (2014)
9. Gerster, H.D.: Schwierigkeiten bei der Entwicklung arithmetischer Konzepte im Zahlenraum bis 100. In: Fritz, A., Ricken, G., Schmidt, S. (eds.) Rechenschwäche. Lernwege, Schwierigkeiten und Hilfen bei Dyskalkulie, pp. 248–268. Beltz, Weinheim (2009)
10. Landerl, K., Butterworth, B., Lorenz, J.H., Möderl, K.: Rechenschwäche - Rechenstörung - Dyskalkulie: Erkennung - Prävention – Förderung. Leykam, Graz (2003)
11. Fuson, K.C.: Using a base-ten blocks learning/teaching approach for first- and second-grade- place-value and multidigit addition and subtraction. Journal for Research in Mathematics Education 21(3), 180–206 (1990), `http://webschoolpro.com/home/projectlead/Research%20Articles%20and%20links/Base%20Ten%20Blocks.pdf`
12. Geary, D.C., Widaman, K.F., Little, T.D.: Cognitive addition and multiplication: Evidence for a single memory network. Memory & Cognition 14, 478–487 (1986)
13. Stazyk, E.H., Ashcraft, M.H., Hamann, M.S.: A network approach to mental multiplication. Journal of Experimental Psychology: Learning, Memory, & Cognition 8, 320–335 (1982)
14. Verguts, T., Fias, W.: Neighbourhood effects in mental arithmetic. Psychology Science 47, 132–140 (2005)
15. Chambers, D.: Direct modeling and invented procedures: Building on children's informal strategies. Teaching Children Mathematics 3, 92–95 (1996)
16. Garnett, K.: Developing fluency with basic number facts: Intervention for students with learning disabilities. Learning Disabilities Research & Practice 7, 210–216 (1992)
17. Thornton, C.: Strategies for the basic facts. In: Payne, J. (ed.) Mathematics for the Young Child, pp. 133–151. National Council of Teachers of Mathematics, Reston (1990)

18. Taraghi, B., Ebner, M., Saranti, A., Schön, M.: On Using Markov Chain to Evidence the Learning Structures and Difficulty Levels of One Digit Multiplication. In: Proceedings of the 4nd International Conference on Learning Analytics and Knowledge, Indianapolis, USA, pp. 68–72 (in print, 2014)
19. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: Proceedings of the Seventh International Conference on World Wide Web 7, WWW7, pp. 107–117. Elsevier Science Publishers B. V., Amsterdam (1998)
20. Deshpande, M., Karypis, G.: Selective Markov models for predicting web page accesses. ACM Trans. Internet Technol. 4(2), 163–184 (2004)
21. Lempel, R., Moran, S.: The stochastic approach for link-structure analysis (salsa) and the tkc effect. Comput. Netw. 33(1-6), 387–401 (2000)
22. Borges, J., Levene, M.: Evaluating variable-length Markov chain models for analysis of user web navigation sessions. IEEE Trans. on Knowl. And Data Eng. 19(4), 441–452 (2007)
23. Bishop, C.: Pattern Recognition and Machine Learning, pp. 424–430. Springer Science and Business Media, LLC, New York (2006)
24. Bishop, C.: Pattern Recognition and Machine Learning, pp. 32–33. Springer Science and Business Media, LLC, New York (2006)