# Cluster-Based Prediction of Mathematical Learning Patterns

Tanja Käser[1], Alberto Giovanni Busetto[1,2], Barbara Solenthaler[1], Juliane Kohn[4], Michael von Aster[3,4,5], and Markus Gross[1]

[1] Department of Computer Science, ETH Zurich, Zurich, Switzerland
[2] Competence Center for Systems Physiology and Metabolic Diseases, Zurich, Switzerland
[3] Center for MR-Research, University Children's Hospital, Zurich, Switzerland
[4] Department of Psychology, University of Potsdam, Potsdam, Germany
[5] Department of Child and Adolescent Psychiatry, German Red Cross Hospitals Westend, Berlin, Germany

**Abstract.** This paper introduces a method to predict and analyse students' mathematical performance by detecting distinguishable subgroups of children who share similar learning patterns. We employ pairwise clustering to analyse a comprehensive dataset of user interactions obtained from a computer-based training system. The available data consist of multiple learning trajectories measured from children with developmental dyscalculia, as well as from control children. Our online classification algorithm allows accurate assignment of children to clusters early in the training, enabling prediction of learning characteristics. The included results demonstrate the high predictive power of assignments of children to subgroups, and the significant improvement in prediction accuracy for short- and long-term performance, knowledge gaps, overall training achievements, and scores of further external assessments.

**Keywords:** feature processing, pairwise clustering, prediction, learning, dyscalculia.

## 1 Introduction

Recently, computer-assisted learning has entered different fields of education. Computer-based therapy systems for learning disabilities have gained particular attention. Such systems present inexpensive extensions to conventional one-to-one therapy by providing an adaptive and fear-free learning environment. The effectiveness of computer-based therapy programs has been proven by several user studies targeting children with dyslexia [3,6,13] and developmental dyscalculia (DD) [11,12,15]. To improve diagnostics and intervention outcomes, knowledge of performance profile, knowledge gaps and learning behaviours of the student as well as an accurate performance prediction are essential. This is particularly important for students suffering from learning disabilities as the heterogeneity of these children requires a high grade of individualization. Current tutoring

K. Yacef et al. (Eds.): AIED 2013, LNAI 7926, pp. 389–399, 2013.

systems use approaches such as Bayesian networks [16], knowledge tracing [4], and performance factors analysis [19] to assess the knowledge of the student.

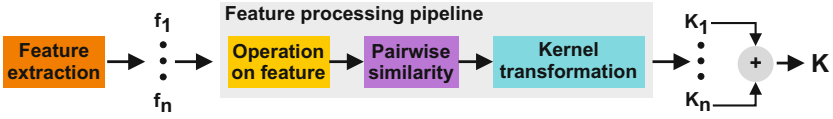
Given the high diversity of students using a tutoring system, training individualization proves highly beneficial and has been the focus of recent improvements. Clustering is a family of approaches which are useful to detect small and homogeneous groups of learners. In fact, clustering [22] and co-clustering [23] approaches successfully improved post-test score predictions. The precision of a knowledge tracing model can be increased using clustering [18] and multiple classification models can also improve performance prediction within a system [5]. Furthermore, ensemble methods offer a way to increase prediction accuracy by training different types of student models [2,17]. Clustering can also be used to gain insight on learning characteristics of the students. Bootstrap aggregated clustering [14] identified different subtypes of children with dyslexia. Other authors used offline clustering followed by online classification to analyse and predict the students' input behaviours [1,10].

The present study aims at predicting and analysing children's mathematical performance on the basis of distinguishable learning patterns extracted from similar subgroups of students. Our approach is articulated in two steps: In a first step, we cluster children according to individual learning trajectories. Compared to previous approaches, we use the subgroup information not only to improve prediction accuracy, but also to provide a valuable tool for experts to analyze individual learning patterns. The second step consists of a supervised online classification during training, enabling prediction of future performance. Whereas existing contributions address the task of predicting short-term performance and external assessment results, we introduce a method which also predicts learning characteristics such as knowledge gaps and overall training achievement. The reported results demonstrate that the prediction accuracy of several learning characteristics can be significantly improved by taking subgroup information into account. They allow for a further training individualization and thus contribute to a better support for children with learning difficulties.

## 2 Method

Our model uses online and offline cluster information. Firstly, we cluster children after the complete training to identify subgroups with similar mathematical learning patterns. Secondly, we classify children to a particular subgroup after each training session to predict future performance. In the following, we first describe the experimental setup and specify the extracted features as well as the feature processing pipeline used for clustering and classification. We then explain clustering, classification and performance prediction in detail.

**Experimental Setup.** The training environment consists of Calcularis [11,12], a tutoring system for children with DD or difficulties in learning mathematics. The program transforms current neuro-cognitive findings into the design of different instructional games, which are classified into two parts. Part A focuses on the training of different number representations, while part B trains addition and




| *Operation* | | *Similarity measures* | | *Kernel transformations* | |
|---|---|---|---|---|---|
| **Inv:** | $1/\mathbf{f}$ | **JC:** | $(\mathbf{F_i} \cap \mathbf{F_j})/(\mathbf{F_i} \cup \mathbf{F_j})$ | **JK:** | $2^S - 1$ |
| **LogInv:** | $1/log(\mathbf{f})$ | **L1:** | $\|\|\mathbf{f_i} - \mathbf{f_j}\|\|_1$ | **GK:** | $exp(-S^2/(2 \cdot \sigma^2))$ |
| **Beta:** | $betacdf(\mathbf{f}, .5, .5)^a$ | **SD:** | $min(\mathbf{f_i} - \mathbf{f_j})^b$ | **RK:** | $exp(-S/\sigma)$ |

[a] Cumulative distribution function of Beta distribution with $\alpha, \beta = 0.5$.
[b] Shortest path between skills on the skill net .

**Fig. 1.** Feature processing pipeline (top) and processing modules employed on feature **f** (**F** in case of a set feature) (bottom). The modules can be combined arbitrarily.

subtraction at different difficulty levels. All games in A and B are played in the number ranges 0-10, 0-100 and 0-1000 ($A_{10}$, $A_{100}$, $A_{1000}$, $B_{10}$, $B_{100}$, $B_{1000}$). The employed student model is a dynamic Bayesian network, consisting of a directed acyclic graph representing different mathematical skills $s$ and their dependencies. The controller acting on the skill net is rule-based and allows forward and backward movements (increase and decrease of difficulty levels).

The data used in the presented analysis was collected by an on-going user study with 88 participants (68% females). 50 participants (72% females) were diagnosed with DD, and 38 participants (63% females) were control children (CC). All participants were German-speaking and visited the 2nd-5th grade of elementary school (mean age: 8.71 (SD 0.91), mean age CC: 8.06 (SD 0.48), mean age DD: 9.21 (SD 0.85)). The children trained with the program for 6 weeks with a frequency of 5 times per week, during sessions of 20 minutes. The collected log files contain 27 complete training sessions per child. On average, each child solved 1430 (SD 212) tasks during the 6 weeks.

**Feature Extraction and Processing.** We identified a set of recorded features, which describe local and global properties of the user's training performance. The set contains cumulative as well as per skill measures, and covers performance, error behaviour and timing. Table 1 lists the features, which are evaluated after each training session. Having continuous and discrete feature types as well as different scales, we process the features to make them comparable (Fig. 1, top). Depending on their nature, features are processed before calculating pairwise similarities $s_{ij}$ (between all samples). The resulting similarity matrices $\mathbf{S_i}$ are transformed into a Kernel and summed up to obtain the similarity matrix **K**. Finally, **K** is transformed to a distance matrix **D** using a constant shift (**D** = #features - **K**). The employed processing modules are listed in Fig. 1 (bottom).

**Clustering.** An inherent property of the controller design of Calcularis is its adaptability. Rather than following a specified sequence of skills to the goal, learning paths are individually adapted for each child. Form and maxima of the network paths vary depending on the learning characteristics of a student

**Table 1.** Extracted features and abbreviations (bold) used in the following

| Feature | Description |
|---|---|
| **H**ighest **S**kills | Indices of highest skills per part (A and B). |
| **N**umber of **P**assed **S**kills | Total number of skills passed. |
| **P**layed **S**kills | Indices of played skills per part (A and B). Set feature. |
| **P**ass **T**imes | Accumulated time (from start of training) in seconds until passing a skill. Not passed skills are set to $\infty$. |
| **S**amples per **S**kill | Number of samples needed to pass a skill. Not passed skills are set to $\infty$. |
| **K**ey **S**kills* | Indices of problem skills. Set feature. |
| **A**nswer **T**imes | Mean answer time per skill. Not played skills set to $\infty$. |
| **P**erformance **P**er **S**kill | Mean performance (correct trials/all trials) per skill. Not played skills are set to 0. |

* Key skill $s$: If a user went back to a precursor skill at least once before passing $s$.

(see Fig. 4). These variations suggest that clustering the children on the basis of their trajectories identifies subgroups of children with similar mathematical learning profiles. Furthermore, the use of the trajectory features allows for modelling the development of mathematical learning over time.

Children are clustered after 27 training sessions using trajectory features. These features take into consideration how far the children came during the training (and how fast they arrived there) as well how they reached this point. The selected features are **PT** evaluated per part and number range (6 dimensions: $A_{10}$, $B_{10}$, $A_{100}$, $B_{100}$, $A_{1000}$, $B_{1000}$) and **PS** (set features for part A and B). **PT** is processed using **LogInv** $\rightarrow$ **L1** $\rightarrow$ **GK** which yields the similarity matrix $\mathbf{K}_1$, while the pipeline **JC** $\rightarrow$ **JK** used for **PS** results in $\mathbf{K}_2$ and $\mathbf{K}_3$. The combined similarity matrix **K** ($\mathbf{K} = \mathbf{K}_1 + \mathbf{K}_2 + \mathbf{K}_3$) is finally transformed to the distance matrix **D** (**D** = 3 - **K**) used for clustering.

As the measurements are characterized by relations, we performed pairwise-clustering (PC) [9] on **D**. Through a kernel transformation, dissimilarity values can be interpreted as distances between points in a (usually higher-dimensional) Euclidean space. As shown by the Constant Shift Embedding transformation, PC exhibits a cost which is equivalent to that of K-means in the Euclidean embedding of the similarity data [21]. The optimal number of clusters is determined by the Bayesian Information Criterion (BIC) [20], calculating the effective number of parameters as the normalized trace of the kernel transformation matrix [8].

**Classification.** We classify students after each training session and use the according cluster information for perfomance prediction. The features used for clustering represent global measures and are thus not optimized for early classification. As all children start the training at the lowest skill level ($A_{10}$), their trajectories tend to be similar during early training and do not provide information about future performance. Therefore, we use additional features taking into account local differences. While **HS**, **NPS**, **PS** and **KS** are cumulative features, **PT**, **SS**, **AT** and **PPS** are evaluated per skill. All features and their processing pipelines are displayed in Fig. 2. The obtained similarity matrices $\mathbf{K}_i$ are

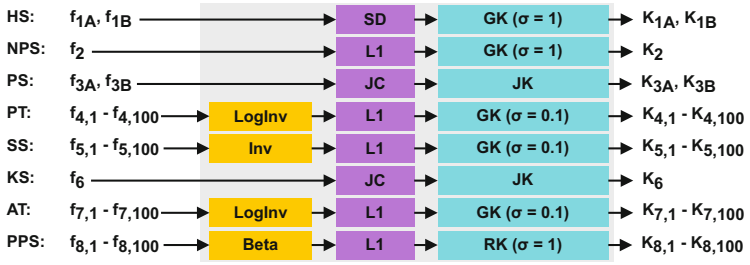



**Fig. 2.** Extracted features and according processing pipelines

transformed to distance matrices $\mathbf{D_i}$ through a constant shift ($\mathbf{D}_i = 1 - \mathbf{K}_i$). Feature processing yields a set of more than 400 distance matrices. Feature selection is performed by ranking the features according to their degree of correlation to the correct labels (of the clustering). An optimal matrix $\mathbf{T}$ is computed, which is a square-matrix containing the pairwise hamming distances between the labels of the samples: $\mathbf{T}(i, j) = 0$, if the samples $i$ and $j$ belong to the same cluster, and $\mathbf{T}(i, j) = 1$ otherwise. For each matrix $\mathbf{D_i}$, we compute the distance $dt$ to the optimal matrix with the Frobenius norm: $dt = ||(\mathbf{T} - \mathbf{D_i})||_F$. The features are then sorted in ascending order by their distance $dt$. For classification, the best combination $b$ of the 10 features with minimal distance to the optimal matrix $\mathbf{T}$ ($2^{10}$ possibilities) is used. The distance matrix $\mathbf{D}$ is obtained by adding up the distance matrices $\mathbf{D_i}$ of the features $\mathbf{f_i}$ contained in $b$. Classification is performed by using a k-nearest neighbours scheme on $\mathbf{D}$. The best combination $b$ and the optimal $k$ are found using a 9-fold cross validation. The classification accuracy is computed on the same folds (not nested).

**Performance Prediction.** The cluster information can be used to predict the student's performance. We identified a set of interesting features (see Tab. 2) that we like to predict. These features can be attributed to four different areas:

1. *Long-term training performance* (**PAS**, **NR**, **HS**): End level reached within the tutoring system.
2. *Short-term training performance* (**NSS**, **NSR**): Prediction of student responses.
3. *Individual knowledge gaps* (**KS**, **KNR**): Identification of particular deficient areas of knowledge.
4. *External test results* (**EPT**): Prediction of external post-test scores. In the **HRT** [7], children are provided with a list of 40 addition (subtraction) tasks ordered by difficulty. The goal is to solve as many tasks as possible within 2 minutes. The mean scores were 21.4 (53% correct) for addition and 19.6 (49% correct) for subtraction. In the **AT** [15], children are presented serially 20 addition (subtraction) tasks and there is no time limit. The mean scores were 16.6 (83% correct) for addition and 14.5 (72% correct) for subtraction.

Prediction of features is performed using cluster information (as described in Tab. 2). The prediction of long-term training performance is interesting for

**Table 2.** Predicted features along with error measures. $\mathbf{f_p}$ denotes the predicted value, $\mathbf{f_t}$ the actual value of the feature, and CE the classification error: $\#(\mathbf{f_p} \neq \mathbf{f_t})/\#played$.

| | Description | Error measures |
|---|---|---|
| **PAS** | Indices of passed skills during training. A skill is predicted as passed, if the cluster majority passed it. | **JC** |
| **NR** | Indices of passed number ranges during training. A range is predicted as passed, if the cluster majority passed it. | **JC** |
| **HS** | Indices of highest skills passed by cluster majority during training (separately for part A and B). | **SD** |
| **NSS** | # samples needed to pass a skill (cluster mean). Predicted only for skills passed by cluster majority. | median($\mathbf{L1}/\lvert\mathbf{f_t}\rvert$) |
| **NSR** | # samples needed to pass a number range (cluster mean). Predicted only for ranges passed by cluster majority. | median($\mathbf{L1}/\lvert\mathbf{f_t}\rvert$) |
| **EPT** | Absolute and relative (#correct tasks/#tasks) post test score (cluster mean): **HRT+**, **HRT-**, **AT+**, **AT-**. | **L1** |
| **KS** | Indices of key skills. A skill is classified as key skill, if the cluster majority has problems. | CE, Recall, Precision |
| **KNR** | Indices of key number ranges. A range is classified as key number range, if it contains at least one key skill. | CE, Recall, Precision |

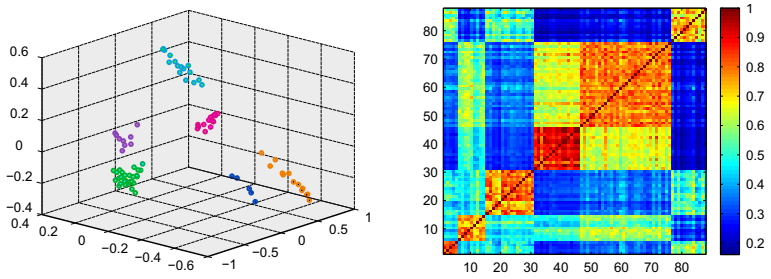



**Fig. 3.** Resulting clusters in 3 dimensions (left) and according similarity matrix (right). High similarities are displayed in red.

analysis as the predicted features are correlated to the learning trajectories. The identification of knowledge gaps helps to find subtypes of mathematical learning patterns and can be used to increase the degree of individualization (e.g., putting more emphasis on the training of key number ranges). Prediction of external test results is especially important for model validation. The prediction of short-term performance can be used to improve adaptation (e.g., minimizing frustration).

## 3 Results and Discussion

**Clustering.** The best BIC score was reached for k = 6 clusters. This result is supported by the clear separability of the transformed data in three dimensions (Fig. 3, left) and the clearly visible clusters on the diagonal of the similarity matrix (Fig. 3, right). Furthermore, the six clusters (C1-C6) can be interpreted

**Table 3.** Data per cluster (**C1 - C6**): Number of children **NC** (%), mean age **AG** (SD), number of passed skills **NPS**, probability of having problems **PP** in different number ranges of the training. + denotes the number ranges passed during training.

| | | **C1** | **C2** | **C3** | **C4** | **C5** | **C6** |
|---|---|---|---|---|---|---|---|
| **NC** | all | 13 (14.77) | 5 (5.68) | 16 (18.18) | 9 (10.23) | 30 (34.09) | 15 (17.05) |
| | CC | 0 (0.00) | 2 (40.00) | 5 (31.25) | 4 (44.40) | 16 (53.30) | 11 (73.30) |
| | DD | 13 (100.0) | 3 (60.00) | 11 (68.75) | 5 (55.60) | 14 (46.70) | 4 (26.70) |
| **AG** | all | 9.26 (0.87) | 8.18 (0.42) | 8.60 (0.67) | 8.52 (1.29) | 8.78 (0.93) | 8.53 (0.87) |
| | CC | - | 8.06 (0.03) | 8.10 (0.49) | 7.52 (0.27) | 8.16 (0.53) | 8.11 (0.44) |
| | DD | 9.26 (0.87) | 8.26 (0.58) | 8.82 (0.64) | 9.32 (1.21) | 9.49 (0.78) | 9.67 (0.71) |
| **NPS** | A, B | 12, 9 | 12, 14 | 15, 12 | 19, 22 | 22, 25 | 22, 30 |
| **PP** | $A_{10}$ | $0.80^{+}$ | $0.95^{+}$ | $0.79^{+}$ | $0.31^{+}$ | $0.39^{+}$ | $0.19^{+}$ |
| | $B_{10}$ | $0.68^{+}$ | $0.20^{+}$ | $0.57^{+}$ | $0.11^{+}$ | $0.14^{+}$ | $0.14^{+}$ |
| | $A_{100}$ | 1.00 | 1.00 | $0.94^{+}$ | $0.91^{+}$ | $0.89^{+}$ | $0.49^{+}$ |
| | $B_{100}$ | 0.99 | $0.98^{+}$ | 0.99 | $0.96^{+}$ | $0.87^{+}$ | $0.30^{+}$ |
| | $A_{1000}$ | x | x | x | 0.98 | $0.72^{+}$ | $0.56^{+}$ |
| | $B_{1000}$ | x | x | x | 0.98 | 0.99 | $1.00^{+}$ |

regarding the characteristics and distinct learning patterns of the samples (Tab. 3), which are reflected in the training trajectories (Fig. 4). The children assigned to C1 have only passed the number range from 0-10. The difficulties with number representation (part A) as well as procedural knowledge (part B) imply an early disorder of numerical functions. All children of this group were diagnosed with DD. Children in C2 have passed the number range 0-100 for part B, but exhibit difficulties in part A. This learning pattern suggests problems with domain-specific functions such as quantity comparison and symbolic representation. In contrast to C2, children in C3 passed the number range 0-100 for part A, but not for B. This observation indicates intact number processing, but difficulties in understanding and executing procedures. The clusters C4 and C5 have passed the number range 0-100 for both parts and the number range 0-1000 for part A, respectively. C6 is the best performing cluster, with children having passed all number ranges and thus finished the training. The performance differences between clusters C4-C6 are probably due to differences in capacity and availability of domain-general functions (attention, working memory, processing speed). Notably, C4-C6 contain DD children (26.7% in C6). This fact can be attributed to age differences: DD children in C6 attend the 4th or 5th grade of elementary school. The interpretation of learning patterns confirms the usefulness of trajectory information for clustering.

**Classification.** During training, we classify the children to a particular subgroup depending on their current training status. As expected, classification accuracy increases with the number of training sessions (Fig. 5, left). Five sessions are already sufficient for the introduced method (blue) to cluster 50% of the

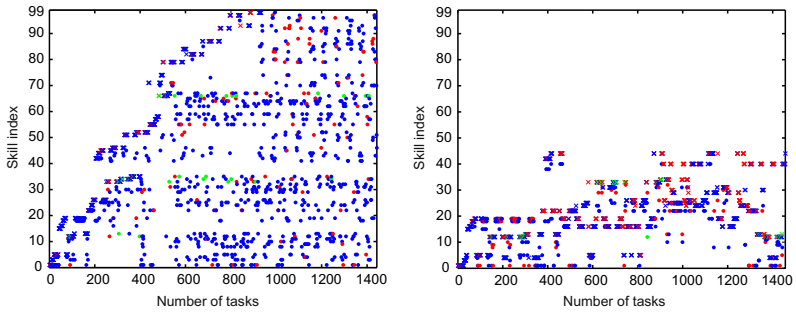



**Fig. 4.** Example trajectories of two children from clusters C6 (left) and C1 (right). A cross denotes a task played at the actual difficulty level while a dot denotes a random repetition. Red stands for a wrong answer, blue for correct, green for neutral.

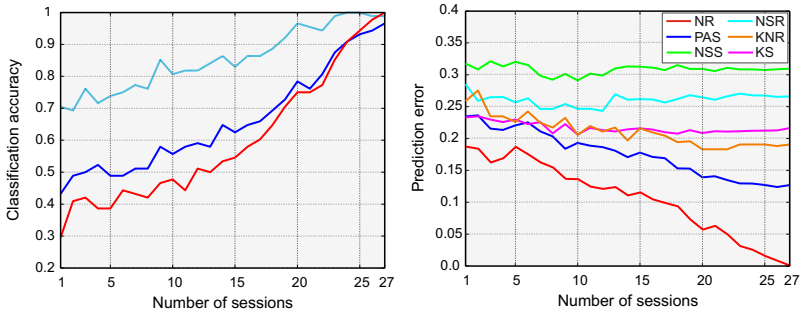
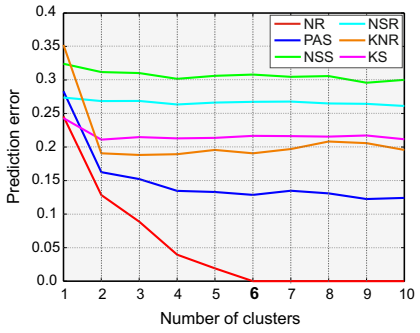



**Fig. 5.** Classification accuracy (left) and performance prediction for selected features (right) over time. Accuracy using offline features (red), the introduced method (blue) and portion of children classified correctly or to a direct neighbour cluster (light blue).

children correctly (chance: $16.\bar{6}\%$). Considering that some neighbouring clusters are close to each other (for instance, C1 and C2 are statistically distinguishable but similar), the assignment of a child to a direct neighbour of the correct cluster will not significantly deteriorate prediction quality. The estimation of the percentage of children assigned to the correct cluster or its direct neighbour (light blue) yields a success rate higher than 70% already after five sessions. The classification with the global features used for clustering (red) performs worse for small numbers of sessions, and equally well after 20 sessions. This behaviour highlights the importance of using local features for classification at an early stage in the training.

**Performance Prediction.** Student's performance in the four selected areas was predicted as described in Tab. 2. Figure 6 (left) shows the prediction errors after 27 sessions (offline prediction) on one to ten clusters. Most errors were significantly reduced (indicated by a two-sided t-test corrected for multiple comparisons with Bonferroni-Holm) by using the cluster information (Fig. 6, right). **NSS** and **NSR** do not show a high cluster dependency. However, as




| Feat. | $\text{Error}_1$ | $\text{Error}_6$ |
|---|---|---|
| **PAS** | 0.28 | 0.13* |
| **NR** | 0.25 | 0.00* |
| **HS** | 2.69, 5.72 | 0.34*, 1.34* |
| **NSS** | 0.32 | 0.31 |
| **NSR** | 0.27 | 0.26 |
| **HRT+** | 4.70 (0.12) | 3.69* (0.09*) |
| **HRT-** | 5.67 (0.14) | 4.50* (0.11*) |
| **AT+** | 3.26 (0.16) | 2.61* (0.13*) |
| **AT-** | 2.98 (0.15) | 2.33* (0.12*) |
| **KS** | 0.24, 0.10, 0.95 | 0.22*, 0.33*, 0.73* |
| **KNR** | 0.35, 0.90, 0.55 | 0.19*, 0.82*, 0.74* |

* $p - value < 0.01$

**Fig. 6.** Offline prediction errors (error measures from Tab. 2) plotted by the number of clusters (left) and listed for one and six clusters (right). For **EPT** features, absolute and relative errors (in brackets) are given and the numbers for **KS** and **KNR** denote classification error, recall and precision. The **HS** error is given for part A and B.

these features are predicted for skills (number ranges) passed by the cluster majority, the number of skills (number ranges) for which we can predict **NSS** (**NSR**) depends on **PAS** (**NR**). The high prediction accuracy of the long-term training performance (**PAS**, **NR**, **HS**) shows that clustering the children based on trajectory features is indeed meaningful. Furthermore, the accurate prediction of post-test results **EPT** demonstrates the correlation between achievement in external assessments and in-tutor performance and thus proves the validity of the student model. The promising results in the identification of knowledge gaps (**KS, KNR**) provide a valuable tool in the analysis of learning patterns and allow experts to elaborate individualized learning strategies. The accurate predictions of knowledge gaps together with the good prediction of short-term training performance (**NSS, NSR**) enable a tutoring system to better adapt the training to individual children. This, however, requires online performance prediction. Online prediction errors for the relevant features were computed after each session. As expected, the prediction errors depend on the classification accuracy (Fig. 5, right), i.e. prediction accuracy increases over the course of the training (due to their cluster independency, this does not hold for **NSS** and **NSR**). A good prediction accuracy is reached already after few trainings and allows to draw conclusions about short-term performance and knowledge gaps.

**Conclusion.** In this work, clustering was applied to learning trajectories of students to determine subgroups in a data set obtained from 88 children (50 children with DD and 38 controls). The computed BIC score suggested that six clusters are optimal. Moreover, the different clusters could be interpreted according to theory about mathematical development and DD. The online classification of the

children to a particular subgroup has shown to be an inherent problem in the beginning of the training, but by using local features the classification accuracy was notably improved, enabling accurate prediction of student's future performance. Student's performance was predicted in four important areas. The results have demonstrated that the prediction accuracy can be significantly increased by taking subgroup information into account. The usefulness of clustering for the analysis of learning pattern and further training individualization contribute to a better support for children with learning difficulties.

**Acknowledgments.** The work was funded by the CTI-grant 11006.1 and the BMBF-grant 01GJ1011.

## References

1. Amershi, S., Conati, C.: Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments. Journal of Educational Data Mining, 18–71 (2009)
2. Baker, R.S.J.d., Pardos, Z.A., Gowda, S.M., Nooraei, B.B., Heffernan, N.T.: Ensembling predictions of student knowledge within intelligent tutoring systems. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 13–24. Springer, Heidelberg (2011)
3. Baschera, G.M., Gross, M.: Poisson-Based Inference for Perturbation Models in Adaptive Spelling Training. International Journal of AIED 20(4), 333–360 (2010)
4. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. UMUAI 4, 253–278 (1994)
5. Gong, Y., Beck, J.E., Ruiz, C.: Modeling multiple distributions of student performances to improve predictive accuracy. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (eds.) UMAP 2012. LNCS, vol. 7379, pp. 102–113. Springer, Heidelberg (2012)
6. Gross, M., Vögeli, C.: A Multimedia Framework for Effective Language Training. Computer & Graphics 31, 761–777 (2007)
7. Haffner, J., Baro, K., Parzer, P., Resch, F.: Heidelberger Rechentest (HRT): Erfassung mathematischer Basiskomptenzen im Grundschulalter (2005)
8. Haghir Chehreghani, M., Busetto, A.G., Buhmann, J.M.: Information theoretic model validation for spectral clustering. In: Proc. AISTATS, pp. 495–503 (2012)
9. Hofmann, T., Buhmann, J.M.: Pairwise data clustering by deterministic annealing. IEEE Trans. Pattern Anal. Mach. Intell. 19(1), 1–14 (1997)
10. Kardan, S., Conati, C.: A framework for capturing distinguishing user interaction behaviours in novel interfaces. In: Proc. EDM, pp. 159–168 (2011)
11. Käser, T., Kucian, K., Ringwald, M., Baschera, G.M., von Aster, M., Gross, M.: Therapy software for enhancing numerical cognition. In: Interdisciplinary Perspectives on Cognition, Education and the Brain, vol. 7, pp. 219–228 (2011)
12. Käser, T., Busetto, A.G., Baschera, G.-M., Kohn, J., Kucian, K., von Aster, M., Gross, M.: Modelling and optimizing the process of learning mathematics. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 389–398. Springer, Heidelberg (2012)
13. Kast, M., Meyer, M., Vögeli, C., Gross, M., Jäncke, L.: Computer-based Multisensory Learning in Children with Developmental Dyslexia. Restorative Neurology and Neuroscience 25(3-4), 355–369 (2007)

14. King, W., Giess, S., Lombardino, L.: Subtyping of children with developmental dyslexia via bootstrap aggregated clustering and the gap statistic: comparison with the double-deficit hypothesis. Int. J. Lang. Comm. Dis. 42(1), 77–95 (2007)
15. Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., Gälli, M., Martin, E., von Aster, M.: Mental Number Line Training in Children with Developmental Dyscalculia. NeuroImage 57(3), 782–795 (2011)
16. Mislevy, R.J., Almond, R.G., Yan, D., Steinberg, L.S.: Bayes nets in educational assessment: Where the numbers come from. In: Proc. UAI, p. 518 (1999)
17. Pardos, Z.A., Gowda, S.M., Baker, R.S., Heffernan, N.T.: The sum is greater than the parts: ensembling models of student knowledge in educational software. SIGKDD Explor. Newsl. 13(2), 37–44 (2012)
18. Pardos, Z.A., Trivedi, S., Heffernan, N.T., Sárközy, G.N.: Clustered knowledge tracing. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 405–410. Springer, Heidelberg (2012)
19. Pavlik, P.I., Cen, H., Koedinger, K.R.: Performance factors analysis - a new alternative to knowledge tracing. In: Proc. AIED, pp. 531–538 (2009)
20. Pelleg, D., Moore, A.: X-means: Extending k-means with efficient estimation of the number of clusters. In: Proc. ICML, pp. 727–734 (2000)
21. Roth, V., Laub, J., Kawanabe, M., Buhmann, J.M.: Optimal cluster preserving embedding of non-metric proximity data. IEEE Trans. Pattern Anal. Mach. Intell. 25(12), 1540–1551 (2003)
22. Trivedi, S., Pardos, Z.A., Heffernan, N.T.: Clustering students to generate an ensemble to improve standard test score predictions. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 377–384. Springer, Heidelberg (2011)
23. Trivedi, S., Pardos, Z.A., Sárközy, G.N., Heffernan, N.T.: Co-clustering by bipartite spectral graph partitioning for out-of-tutor prediction. In: EDM, pp. 33–40 (2012)