Extracting Adaptation Strategies for E-Learning Programs with XCS

Anke Endler
Department Psychology III
University of Würzburg, Germany
anke.endler@uni-wuerzburg.de

Martin V. Butz
Department Psychology III
University of Würzburg, Germany
butz@psychologie.uni-wuerzburg.de

Günter Daniel Rey
Department Psychology IV
University of Würzburg, Germany
rey@psychologie.uni-wuerzburg.de

ABSTRACT

This paper investigates XCS performance on a scarce and noisy artificial and a real-world data set. The real-world data set is derived from an E-Learning study, in which motivation was correlated with the adaptation of difficulty. The artificial data set was generated to evaluate if XCS can be expected to mine information from the real-world data set. By adding sparsity and noise to the artificial data set, mimicking the properties of the real-world data set, we show that XCS can handle scarce and noisy data well. We furthermore show that the extracted structure contains problem-relevant information, and that revealed structures in the real-world data correspond to actual psychological learning theories. Thus, the contributions of the paper are twofold: (1) We show that XCS can mine highly scarce and noisy data; and (2) the results suggest that the current motivational state of the user may be utilized to adapt an E-Learning program for improving learning progress.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Experimentation

Keywords

Learning Classifier Systems, Datamining, Knowledge Extraction

1. XCS

XCS is a learning classifier system, which was introduced by Stewart Wilson [9]. As such, it evolves rules, using reinforcement learning techniques for rule evaluation and a genetic algorithm for rule evolution. The system has been successfully applied to various classification and datamining problems, as well as reinforcement learning problems [1, 2, 3]. Moreover, due to its rule-based representation, knowledge extraction is easy to accomplish [10]. That is, the system is not only suitable to yield good classification accuracy, but it is also suitable to extract particular feature dependencies hidden in the analyzed data.

We utilize an XCS version that processes integer-valued inputs, similar to Wilson's XCSI setup [11]. Moreover, we do not use any action encoding, similar to the XCSF setup. Thus, our XCS setup specifies no action or classification, and the system predicts one reward value. To avoid further name confusions, we will refer to our setup as an XCS system.

We analyze XCS performance in a case where only very scarce and noisy data is available for learning. The data set was extracted from an E-Learning study, in which the current user motivation was correlated with task difficulty adjustments.

2. DATA ENCODING AND GENERATION

The focus of this work was the extraction of adaptation strategies for E-Learning programs based on the motivational state of the user. Each participant of the conducted study worked through two successive learning phases, which consisted of a learning block, during which 10 tasks were presented, and two test blocks, during which performance was assessed. A motivational questionnaire, based on the one introduced in [6], was used to detect the learner's current motivational state, which was subdivided into anxiety, probability of success, interest, and challenge. Each factor was encoded by three possible values: low, medium, or high. At the beginning of each learning block, adaptation of task difficulty took place, increasing, decreasing, or maintaining the previous level of difficulty. This adaptation of difficulty was randomly applied by the E-Learning program during the study to gain a broad sample of data. Within a learning block, difficulty was further adapted according to the user's performance, increasing (decreasing) the difficulty after two successive correct (incorrect) answers.

The study was conducted with 37 participants, yielding a data set of 74 data entries. Further details on the study, the participant distribution, and prior data analyzes can be found in [4].

The features of each data entry consisted of 5 nominal values, each of which could take on three actual values. This yields a problem input space size of $3^5 = 243$ possible input

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GECCO'11, July 12–16, 2011, Dublin, Ireland.
Copyright 2011 ACM 978-1-4503-0690-4/11/07 ...$10.00.

value combinations. Since the study provided us with 74 data entries, maximally 30.5% of the problem space could be covered.

3. EVALUATION ON ARTIFICIAL DATA

We now proceed with analyzing if XCS is able to identify the hidden systematics in the artificial data set, which was based on a decision tree structure. This structure simulated an intuitively logical and easily verifiable artificial scenario, with comparable characteristics to the real-world scenario. To imitate the scarcity and inaccuracy of real data, we limit the number of different input values to the number of values derived from the study, and add Gaussian noise.

We evaluate XCS performance dependency on three crucial parameters: its maximum population size N, the number of learning iterations T, and the start of the compaction mechanism C. We, furthermore, vary the noise ratio. Intuitively speaking, parameter N fosters competition in the population: give a large population, competition is low and learning is delayed, however, given a very small value, competition may be too strong. Parameter T specifies the time during which XCS can evolve appropriate rules. Enough time needs to be provided for XCS to converge - however, overly long learning may also result in overfitting. The compaction mechanism simply stops mutation and crossover operators from being applied, thus condensating the population to the most dominant, accurate classifiers at that time. All reported evaluations are done using ten independent learning runs, reporting the average values for mean absolute error and number of distinct classifiers.

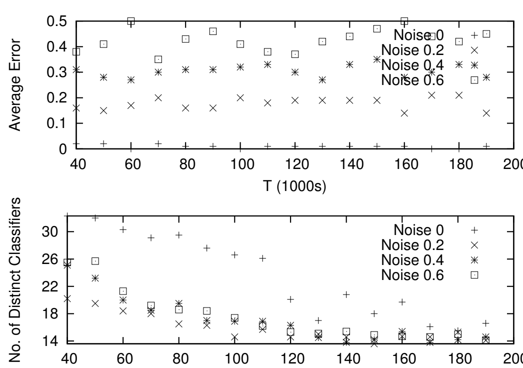

Figure 1: Size-limited, noisy artificial data set with $N = 400$ and $C = 75\%$ of T

Fig. 1 shows XCS performance on scarce and noisy artificial data with various noise ratios. The figure shows that, as can be expected, noise has a major influence on the average error. The higher the noise, the higher the average error, irrespective of T. The number of distinct classifiers, on the other hand, drops when noise is added. More noise, however, does not have any significant influence on the number of distinct classifiers, which decreases for a higher T, irrespective of noise. The results suggest that $T = 100,000$ or higher is appropriate. For this setting, we can expect a number of distinct classifiers lower than 20, which guarantees some generalization over the data.

Table 1 shows one batch of classifiers detected by XCS on an artificial size-limited and noisy data set, with $T = 100,000$ and noise of 0.2. This particular batch reached an average error of 0.178 with 13 classifiers. The last two columns of the table indicate the correctly predicted percentage of the problem space that is covered by a rule, with reference to the entire problem space and the problem space covered by the data set, respectively.
All classifiers are able to predict the part of the problem space covered by the data set correctly, i.e. they can predict if learning success is rather high or low. 7 of the 13 classifiers can predict their entire coverage correctly and all 6 of the remaining classifiers are able to predict at least 50% of their coverage correctly.

To ensure further that we gain reliable rules, we test them on unseen data using cross-validation.

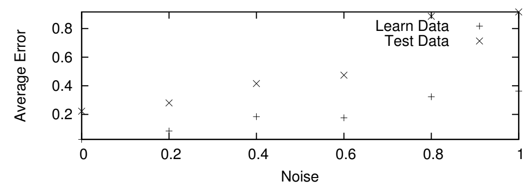

Figure 2: Cross-validation on size-limited, noisy artificial data set with $N = 400$, $T = 100,000$ and $C = 75\%$ of T

Fig. 2 shows the average learning and test error for the artificial data with several levels of Gaussian noise and $T = 100,000$. Both, learning and test error increase significantly with increasing noise, so that for a high noise ratio the rules derived from XCS cannot feasibly predict unknown data, but will make virtually random predictions, which allow no conclusions to be drawn about the actual reward. For a noise of 0.2 or smaller, however, the test data indicates that the system is able to reliably learn an acceptable number of rules, which can predict unknown data feasibly well.

4. PERFORMANCE ON REAL-WORLD DATA

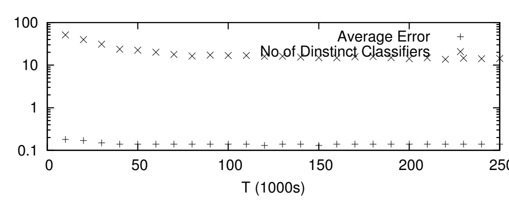

Figure 3: Real-world data set with $N = 400$ and $C = 75\%$ of T, log-scaled

Fig. 3 shows the average error and the number of distinct classifiers for the real-world data. The average error remains stable around a value of 0.139 for $T \geq 40,000$. This is comparable to, and even slightly lower than, the average error achieved in the artificial data set for a noise of 0.2, which is an acceptable noise ratio. The number of distinct classifiers decreases significantly for up to $T = 80,000$. For higher T, it only decreases slightly, which again coincides with a noise of 0.2 in the artificial data.

As XCS with the real-world data produces similar results to the artificial data with noise of 0.2, we can hope to gain

Anx	Succ	Int	Chall	Adapt	Reward	Err	Fit	CorTot	CorSet
2	1-2	1-2	0-2	1	0.597	0.35	0.98	1.0	1.0
1-2	0-2	0-2	1	2	0.153	0.28	0.21	0.66	1.0
1-2	0	1	0-2	0	0.815	0.001	0.99	0.5	1.0
1	2	0	1-2	1	1.210	0.00	1.0	1.0	1.0
0-2	1-2	0-2	0-2	0	0.007	0.06	1.0	1.0	1.0
0-1	0-1	0-2	0-2	2	0.747	0.08	1.0	0.78	1.0
1-2	0	1-2	1-2	0-2	0.021	0.03	1.0	0.67	1.0
2	0-1	0	0-1	0-1	0.606	0.13	1.0	0.5	1.0
1	1	0	1-2	1	0.863	0.13	1.0	1.0	1.0
2	0-2	0-2	2	2	0.214	0.09	1.0	1.0	1.0
0-1	0-2	0-2	0-2	0	-0.054	0.16	1.0	1.0	1.0
0	0-2	0	0-2	1	0.824	0.18	1.0	0.67	1.0
0-1	0-2	1-2	0-2	1	0.156	0.18	1.0	1.0	1.0

Table 1: A set of rules for the test function with limited input and noise of 0.2, with $N = 400$, $T = 100,000$ and $C = 75\%$ of T

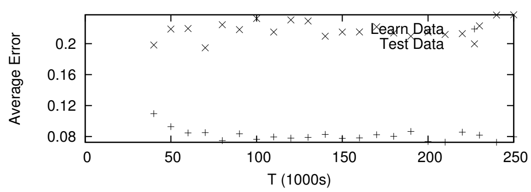

Figure 4: Cross validation with real-world data set with $N = 400$ and $C = 75\%$ of T

acceptable results using cross validation. Fig. 4 shows the average error for the learn and test data derived from cross-validation with the real-world data for various choices of T. The learning error stays around 0.08, while the test error never exceeds 0.24. These results show a slightly higher error than the results for artificial data without noise but are, again, comparable to, and even slightly lower than, the results for artificial data with noise of 0.2. We, therefore, conclude that XCS is able to extract feasible information from the real-world data.

5. KNOWLEDGE EXTRACTION AND ANALYSIS

Our evaluations, using artificial data as well as cross-validation, suggest that XCS is able to derive reasonably reliable rules from scarce and noisy data.

To support these results further, we extract a number of stable rules from ten runs with $T = 160,000$, $N = 400$ and $C = 75\%$ of T. We only consider those rules with fitness greater than 0.9 and experience greater than 5000, which appeared in at least two different runs. This final set of 12 rules is shown in Tab. 2. The first column gives a rule number to every rule. The next five columns show the rule. The last column shows the reward prediction from every instance of this rule. Predictions for the same rule always show a similar value with the highest deviation being 0.218 for Rule 2. Rule 4 subsumes Rule 3, Rule 7 subsumes Rule 5 and Rule 10 subsumes Rule 11 as well as Rule 12.

To analyze the validity and utility of the extracted knowledge, we now interpret the extracted rules according to two well-known, widely accepted psychological learning theories that include motivational aspects.

5.1 Zone of Proximal Development (ZPD)

The zone of proximal development (see [7]) predicts that the highest learning success can be expected if expertise and difficulty are on a similar level as depicted in Fig. 5.

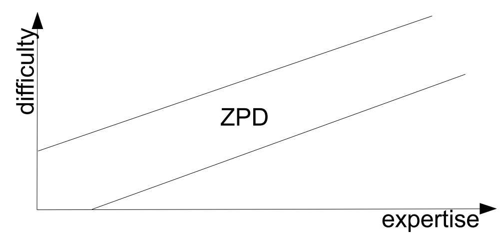

Figure 5: Zone of Proximal Development (cf. [5])

We interpret challenge as an indication for the interaction between task difficulty and learner's expertise. More precisely, this interpretation assumes that high challenge indicates that the difficulty exceeds the learner's expertise while low challenge indicates, vice versa, that the learner's expertise exceeds difficulty. A balance of difficulty and expertise will, in this scenario, result in medium challenge. High challenge should then indicate that difficulty is too high and the learner will leave the ZPD. In this case, reducing the level of difficulty will return the learner into the ZPD and therefore, as a tendency, increase learning success. This is supported by Rule 1. If, on the other hand, difficulty is decreased further when a learner reports low to medium challenge, they leave the ZPD because difficulty is too low for their expertise, resulting in low learning success, like Rule 2 predicts. The same holds if the difficulty is not decreased for high challenge (see Rule 11). All other rules give no indication for either challenge or adaptation.

5.2 Yerkes-Dodson Law

The Yerkes-Dodson Law (see e.g. [8]) assumes an interrelation between arousal, task difficulty, and performance. The law postulates that a certain amount of arousal, i.e. motivation, is necessary to activate learning. For easy tasks, higher activating motivation is expected to result in higher performance. For difficult tasks, however, too high activating motivation may result in a decrease of performance again.

No.	Anx	Suc	Int	Chal	Adapt	Reward
1	0-2	0-2	0-1	1-2	0	0.727, 0.833
2	0-2	2	0-1	0-1	0-1	0.379, 0.396, 0.423, 0.597
3	0-2	1-2	1	0-2	0-1	0.596, 0.633
4	0-2	0-2	1	0-2	0-1	0.620, 0.833
5	1-2	0-2	0-1	0	0-2	0.479, 0.538
6	1-2	1-2	0-1	1-2	0-2	0.670, 0.625, 0.634
7	1-2	0-2	0-2	0	0-2	0.454, 0.473, 0.535
8	2	0-2	0-2	0-2	1	0.704, 0.709, 0.741, 0.749, 0.763
9	0	0-1	0-2	0-2	1-2	0.462, 0.473, 0.497, 0.502
10	0-2	2	1-2	0-2	1-2	0.454, 0.517, 0.594
11	0-2	2	1-2	1-2	1-2	0.466, 0.513
12	0-2	2	2	0-2	1-2	0.437, 0.481

Table 2: Rules derived from the study's data

Anxiety may be such an activating motivational factor. In the study, self-assessed anxiety has to be seen in the context of performance having no consequences for the participants. Therefore, we assume that even reported high anxiety does not leave the range where it is activating rather than blocking learning. Consequently high (low) anxiety should lead to a high (low) learning success. This is confirmed by Rules 8 and 9, respectively. Medium to high anxiety, which is specified in Rules 5, 6 and 7, shows a wider range in learning success but mainly within the boundaries of the learning success of Rules 8 and 9.

Interest can be analyzed in the light of the same law. Interest, however, may take the full range in the scope of the study, so that we expect medium interest to result in high learning success, and low or high interest in a lower learning success. Rules 3 and 4 show an acceptable learning success for medium interest. Rules 2 and 5 show a low learning success for low to medium interest and Rules 10, 11 and 12 show a low learning success for medium to high interest. Rules 1 and 6, however, contradict this theory. Rule 1 shows a very high learning success for low to medium interest. This might be due to other factors, such as ZPD. Rule 6 also shows a rather high learning success for low to medium interest. Rule 6 differs from Rule 5 mainly in challenge, suggesting that high challenge may compensate for the lack in interest.

6. CONCLUSIONS

In this paper, we utilized XCS to extract information from scarce and noisy data. With only 74 data sets, the learning material did not cover the entire problem space. Similarly, only a small number of sets were available for each covered condition, inevitably yielding noisy data samples.

We used data from an intuitively logical and easily verifiable artificial scenario to evaluate XCS performance. With this artificial data set, we were able to show that XCS is able to extract rules from scarce and, to some extent, noisy data. We could therefore use XCS to extract adaptation strategies, which are based on the motivational state of the user, for an E-Learning program. This real-world data showed a similar behavior to the artificial data with a noise ratio of 0.2, which indicated that XCS could extract knowledge from the data. 10-fold cross-validation, as well as consistency with psychological theories on learning, supported the reliability of the extracted rules.

We conclude that, when using adequate parameter settings, XCS can handle scarce and noisy data well. Moreover, general and reliable rules can be extracted from the data. Further studies have to confirm the reliability of the extracted rules, as well as their applicability in E-learning scenarios.

7. REFERENCES

[1] E. Bernadó-Mansilla and J. M. Garrell-Guiu. Accuracy-based learning classifier systems: Models, analysis, and applications to classification tasks. *Evolutionary Computation*, 11:209–238, 2003.

[2] L. Bull, editor. *Applications of Learning Classifier Systems*. Springer-Verlag, 2004.

[3] M. V. Butz. *Rule-Based Evolutionary Online Learning Systems: A Principled Approach to LCS Analysis and Design*. Springer-Verlag, Berlin Heidelberg, 2006.

[4] A. Endler. Towards Motivation-Based Adaptation Strategies for Difficulty in E-Learning Systems. Diplomarbeit, Julius-Maximilans-Universität Würzburg, 2010.

[5] G. D. Rey. *E-Learning. Theorien, Gestaltungsempfehlungen und Forschung*. Huber, 2009.

[6] F. Rheinberg, R. Vollmeyer, and B. D. Burns. FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen. Technical report, Universität Potsdam, 2001.

[7] L. Vygotski. Learning and mental development at school age. *Educational psychology in the U.S.S.R.*, pages 21 – 34, 1963.

[8] B. Weiner. *Human Motivation*. Springer, 1985.

[9] S. W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2):149–175, 1995.

[10] S. W. Wilson. Get real! XCS with continuous-valued inputs. In P. L. Lanzi, W. Stolzmann, and S. W. Wilson, editors, *Learning classifier systems: From foundations to applications (LNAI 1813)*, pages 209–219. Springer-Verlag, Berlin Heidelberg, 2000.

[11] S. W. Wilson. Mining oblique data with XCS. In P. L. Lanzi, W. Stolzmann, and S. W. Wilson, editors, *Advances in learning classifier systems: Third international workshop, IWLCS 2000 (LNAI 1996)*, pages 158–174. Springer-Verlag, Berlin Heidelberg, 2001.