Investigating Swarm Intelligence for Performance Prediction*

Mohammad Majid al-Rifaie†
Goldsmiths, University of London
Department of Computing
London SE14 6NW, UK
m.majid@gold.ac.uk

Matthew Yee-King
Goldsmiths, Uni of London
Department of Computing
London SE14 6NW, UK
m.yee-king@gold.ac.uk

Mark d'Inverno
Goldsmiths, Uni of London
Department of Computing
London SE14 6NW, UK
dinverno@gold.ac.uk

ABSTRACT
This paper proposes a new technique for analysing the behaviour of students on an online course. This work considers a range of social learning behaviours supported in our recently designed and implemented collaborative learning system which supports students giving and receiving feedback on each other's developing work and practice. The course was delivered to several thousand students on Coursera during which students were directed onto our social learning environment to take part in group work and assessment activities. This work introduces a swarm intelligence technique, Stochastic Diffusion Search (SDS), and shows how it can be adapted and applied to our data in order to perform classification tasks. The novelty of the approach is not only in using this technique, but also applying it to data linked to *social behaviour* (i.e. how students interact with each other) which differentiates the work apart from many clickstream analysis studies. This paper investigates what combined activity is the best predictor of success or failure in the course. The aim is to argues that the results obtained using the proposed approach indicate the promising potential of predicting students performance through applying swarm intelligence technique to social behaviours. This work has a number of potential benefits including designing better social learning systems, designing more effective social learning and assessment exercises, and encouraging disengaged students. In addition, this work is an important step in addressing our long term goal of evidencing how critical student learning takes place as they give and receive feedback to and from each other on work in progress.

Keywords
Social learning, swarm intelligence, education system modelling, MOOC

1. INTRODUCTION

Increasingly researchers are focusing on the significance of social learning and investigating its impact within the various online learning environments. Acknowledging the importance of collaboration and 'teamwork', as an embedded element in the Massive Open Online Courses (MOOCs), this method of learning is desirable for many employers who rely on highly collaborative and online-based works. Our programme of work is concerned with designing a novel learning technology, online courses and assessments, which provide us with a range of data we can use to understand how learning takes place through online social interaction. Our pedagogy is influenced by our home institution's "art-school" pedagogy across practice-based subjects (such as art, music and design) where students learn by sharing "work in progress" within tutor groups and giving and receiving feedback to each other. The aim of this work is to use learning analytics to build strong arguments for the adoption of social learning pedagogies supported by innovative technology. Therefore this paper focuses on extracting information from *social learning activity logs*, not the full range of more traditional courseware access and activity logs. The objective is to gain a better understanding if these activities have any measurable relation to learning, and if so which are the most important activities and in which combinations. The analysis presented here is a first step in that direction, where the attempt is to predict if students will pass or fail a course, using only low level user interface telemetry data gathered from our social learning platform. Given the undeniable significance of data classification in different and diverse scientific domains (e.g. computer science, psychology, medicine), various techniques have been proposed over the years. Nature-inspired metaheuristic algorithms are among one of the categories which aimed at providing solutions to this problem.

In this paper a novel method in addressing data classification in the context of educational data is used where a swarm intelligence algorithm is adapted for this purpose. A recent review [2] details the extensive applications of this algorithm in the last two decades in various fields (e.g. discrete and global optimisation, pattern recognition, resource allocation, medical imagining, etc).

*(Does NOT produce the permission block, copyright information nor page numbering). For use with ACM_PROC_ARTICLE-SP.CLS. Supported by ACM.

†Corresponding author

The research questions which drive this paper are as follows:

1. How can the proposed swarm intelligence technique (SDS) be applied to educational data?
2. What kinds of social learning activities, and what combinations of social learning activities are the best predictors?
3. Does social interaction data contain strong predictive potential of student success?
4. How does an SDS analysis of social learning data help us in designing and delivering learning activities, in improving social/group learning activities, and in building better social learning systems?

In this paper, first Stochastic Diffusion Search (SDS) algorithm is explained, detailing its behaviour and highlighting one of its main features (i.e. partial function evaluation). Then, an introduction is given to the classification problem in general followed by a brief section on the nature of the educational dataset used in this paper and the features available from the dataset. After elaborating on the data in the datasets in the context of the work, the swarm intelligence algorithm used is adapted for the purpose of the experiments conducted in this paper and the results are reported. A discussion on the behaviour of the proposed algorithm is presented showing its potential in using all the available features as well as identifying the most significant features. Finally, the paper is concluded with the summary of the research reported in the paper along with directions for future research.

2. RELATED WORK

With the increasing use of online learning platforms, a large number of researchers have been working on predicting grades from students performance over the course of the studies. This topic of research is of importance because, for example, only in the United States several hundred thousand students drop out of high school every year and perhaps interventions can provide the means to reduce the number of those falling behind in their studies [1, 7]. With the growing interest in MOOCs as alternative or adjunct learning platforms, behaviour prediction has attracted the attention of many educational data analyst, such as Brady et al. [15] who used higher granularity temporal information for their analytics work; in another work, Macfadyen et al. [8] explain the concept of "an early warning system" for educator, aiming to provide the means for the educators to intervene with an appropriate set of actions to improve the performance of the weaker students; a similar work was presented by Rogers et al. [11] which aims to identify students at rist of failure. The predictive power of demographics versus activity patterns in MOOCs are discussed by Brooks et al. [3] focusing on whether it is possible to find a link between performance and demographics. Other researchers, such as Coleman et al. [4] or Elbadrawy et al. [6], have also been exploring whether it is feasible to identify behavioural patterns for prediction. In addition to attempting to improve students performance, Yang et al. [14] have been focusing on the concept of dropouts which is a critical challenge for online courses. Considering the above recent work, it is evident that extracting useful knowledge from education data should ultimately be incorporated in the design of the online systems. In a recent work by researchers from Harvard University and MIT, Whitehill et al. [13] emphasised on the importance of intervention and especially automatic intervention in MOOCs in order to take measures to reduce the number of students quitting; they claim that their proposed system might encourage students to return into the course. In another work, by Rollinson and Brunskill, [12] emphasis has been put on the importance of coupling predictive models with an alternative student model and policy (which constitute the core of the Intelligent Tutoring Systems), focusing again on the importance of using predictive models along with other tools. Having mentioned the above research, it is important to state that arguably one of the important features in MOOCs is enabling learners to discuss their work with their peers and receive feedback. In a recent research, Olsen et al. [9] direct the prediction power towards collaborative learning environment; in their work, they argue that by adding collaborative learning features they were able to enhance their understanding on the impact of collaborative learning. Tightly related to the mentioned work, the importance of social centrality in the context of MOOCs is discussed by Dowell et al. [5] where they adopt an approach, which uses language and discourse as a tool to explore the association with the existing and established measures related to learning (i.e. traditional academic performance and social centrality). While this work does not endorse or reject the impact of social learning, it clearly shows an increasing interest in exploring the impact of collaborative learning.

3. STOCHASTIC DIFFUSION SEARCH

Stochastic Diffusion Search (SDS) [2] which was first proposed in 1989 is a probabilistic approach for solving best-fit pattern recognition and matching problems. SDS, as a multi-agent population-based global search and optimisation algorithm, is a distributed mode of computation utilising interaction between simple agents. Its computational roots stem from Geoff Hinton's interest in 3D object classification and mapping and its applications span from continuous optimisation to medical imagining. The SDS algorithm commences a search or optimisation by initialising its population and then iterating through two phases: the test and diffusion phases. In the test phase, SDS checks whether the agent hypothesis is successful or not by performing a hypothesis evaluation which returns a boolean value. Once the activity (i.e their status as being 'true' or 'false') of all the agents are determined, successful hypotheses diffuse across the population and in this way information on potentially good solutions spreads throughout the entire population of agents. In other words, each agent recruits another agent for interaction and potential communication of hypothesis. The spreading of information occurs during the diffusion phase.

In standard SDS (which is used in this paper), *passive recruitment mode* is employed. In this mode, if the agent is inactive, a second agent is randomly selected for diffusion; if the second agent is active, its hypothesis is communicated (*diffused*) to the inactive one. Otherwise there is no flow of information between agents; instead a completely new hypothesis is generated for the first inactive agent at random. Therefore, recruitment is not the responsibility of the active agents. In this work, activity of each agent is determined when its fitness is compared against a random agent (which is different from the selecting one); if the selecting agent has a better fitness (smaller value in minimisation problems)

Table 1: The list of features logged, along with examples of the total figures for a single student. The last column represents the grade correlation of each individual figure.

	Description	Example	Corr
F1	Play video	199	0.41
F2	Delete a reply	16	0.12
F3	Open item in search result list	0	0.15
F4	Report problem with media	22	0.48
F5	Load media	7580	0.41
F6	Report problem with reply	24	0.26
F7	Delete an annotation	0	0.19
F8	Save after edit	0	0.15
F9	View my files	954	0.40
F10	View set of shared files	8865	0.41
F11	Save after edit	0	0.11
F12	Delete video	0	0.18
F13	Periodically log and comment when video is playing	1928	0.30
F14	Play region and view thread	1313	0.53
F15	Save user profile	32	0.23
	Course final grade	100	1.00

than the randomly selected agent, it will be flagged as active, otherwise inactive. A higher rate of inactivity boosts exploration, whereas a lower rate biases the performance towards exploitation.

4. CASE STUDY AND DATASET

The analysis presented in this paper is based on a dataset gathered during a seven week creative programming course on Coursera which ran in Summer 2014. The course presented students with a series of worked example programs written using Processing [10] that were either musical, graphical or game based. It was assessed using weekly quizzes and three, biweekly peer assessments. The peer assessments required the students to select one of the tutor-supplied worked examples and extend it in some way of their choosing. They then had to create a five minute screencast video wherein they explained the changes they had made from the example code and demonstrated the running program. This video was uploaded to our social learning system and then a link to this was submitted to the main MOOC LMS. Our system allowed them to place comments along the timeline of the video and to view a range of suggested content from other students, such as highly commented and uncommented videos. Our system collects detailed logs of certain interface elements that the user clicked on or moused over, including a user id and a timestamp. The data set used in this paper consists of these clickstream logs plus final grades achieved on the course. There were a total of 993 students who created logs on our system and gained a final grade on the Coursera platform. The dataset spanned a period of about seven weeks. Each student's log data and final grade was converted into a feature vector containing totals for all of the observed log types taken over the entire time period of the study. Table 1 shows an example of such a vector. The research began by attempting to correlate individual elements of the vector to $final\_grade$ but individual correlations were statistically insignificant to predict grades so instead a multivariate classification approach is attempted, the results of which form the remainder of this paper. The main aim was to label students as pass (≥ 50) or fail (< 50).

5. APPLY THE SDS ALGORITHM

Here the process through which the SDS algorithm was adapted to perform the classification tasks is detailed and the steps taken during the *test* and *diffusion* phases are explained. In order to apply this swarm intelligence algorithm to the dataset the following are considered:

- **Search space** is the entire dataset
- **SDS hypothesis** refers to a student record
- **Student attributes:** Each student record has fifteen attributes or features (i.e. play, report_media, region_block, etc; see Table 1).
- **Micro-features:** The fifteen features of each student record are considered the micro-features of the hypothesis. Therefore each SDS hypothesis has fifteen micro-features referring to the attributes of the student.

Next, the phases used in SDS algorithm are highlighted and each phase is described briefly in the context of the dataset presented.

During the *initialisation phase*, one student is chosen randomly from the dataset and is set as a model. Then each agent is randomly associated with a student record from the search space. During the *test phase*, each agent (which is already allocated to a student) randomly picks one of the fifteen micro-features and compares its value against that of the model. If the difference between the two corresponding micro-features is within a specific threshold, τ_d (where τ is the threshold and d is the dimension) the agent becomes active, otherwise inactive. The process in the *diffusion phase* is the same as the one detailed in the algorithm description: each inactive agent picks an agent randomly from the population; if the randomly selected agent is active, the inactive agent adopts the hypothesis of the active agent (i.e. they refer to the same student as their hypothesis), otherwise the inactive agent picks a random student from the dataset.

Categories, Classes and Termination The agents iterate through the test and diffusion phases again until all agents are active. At this stage, the students referred to by all the active agents are assigned to a category. Additionally, the number of active agents on each student is logged. Once a category is determined, the process is repeated from the initialisation phase where agents are initialised throughout the search space and the first student which has not yet been assigned to any categories is set as the new model. Then the algorithm iterates through the test and diffusion phases until all students are allocated to a category. Finally, categories form the classes, and when there exist students that belong to more than one class, they will be allocated to the one which has attracted a larger number of active agents. The only tunable parameters for SDS is the swarm size, N which is empirically set to $N = 10,000$. Threshold, $\vec{\tau}$, which is the acceptable distance between the model and other samples for each dimension, d, is calculated using the following formula:

$$\vec{\tau}_d = \sum_{t=1}^{c} \left| \frac{\text{MAX}\left(\vec{I}_{id}^{t}\right) - \text{MIN}\left(\vec{I}_{id}^{t}\right)}{c} \right| \quad d = [1, 2, ..., 15] \quad (1)$$

where c is the number of student types or classes in the dataset (i.e. pass and fail); $\vec{I}_{id}^{t}$ represents the value of i^{th} student with type t and dimension d. There are 2 student types ($c = 2$) and the dimensionality of the problem is 15 (see Table 1). Therefore the difference between the minimum and maximum values in each band (e.g. pass and fail) is calculated, then the sum of the differences in each dimension is averaged and used to calculate the threshold. Using the formula above the threshold $\vec{\tau}$ is calculated using the

Table 2: Weekly breakdown of and fail/pass rate

	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7
Active students	245	974	629	683	488	528	265
Ratio	25%	98%	63%	69%	49%	53%	27%
% of fails	28%	39%	28%	16%	10%	5%	2%
% of passes	72%	61%	72%	84%	90%	95%	98%

training dataset. Using the threshold vector presented, if the randomly picked model falls on the first class (e.g. the fail class), it is likely that the active agents have a bigger presence in this class. It is worth noting that while in some iterations there is a high presence of active agents for some students, in some other iterations there is a high number of inactive agents on the same students. The reason why a student record could make an agent active in one iteration and inactive in another can be explained through SDS's random micro-features selection: each record consists of fifteen micro-features (the same as the number of attributes for each student), therefore if an agent picks one of the micro-features that are within the threshold, the agent becomes active, but if it randomly picks one of the other micro-features, the agent becomes inactive. Deducing from this, it is evident that having more micro-features within the range of the model results in more agents becoming (and staying) active, and as a result forming a stable category.

6. EXPERIMENTS AND RESULTS

In this section, the results of several experiments are reported along with a discussion on the relevance of the experiments to the research questions. The total number of students who used the online learning platform and obtained a final grade was 993. The number of active students each week and the fail/pass rate of students are detailed in Table 2, and the SDS algorithm is used as the classifier.

6.1 Experiment I: Weekly data analysis

The logged actions of all students who have participated in the previous and current weeks are cumulated and fed into the system for analysis.

One of the important elements in the cumulative data is the distribution of fail and pass in each of the training and test datasets. Fig. 1 shows this distribution in the test dataset. Note that the training datasets will have the same distribution as the test dataset. As illustrated in the figure, other than the first week, in the rest of the week, the cumulative data shows 39% and 61% of the data belonging to the fail and pass categories respectively. The classifier is trained and the prediction accuracy of the classifier is evaluated on the test datasets.

Table 3 and Fig. 2 show the weekly prediction-accuracy on the test datasets. As expected, and due to the presence of more data as students progress to the next weeks, there is a gradual increase in the prediction accuracy of the swarm

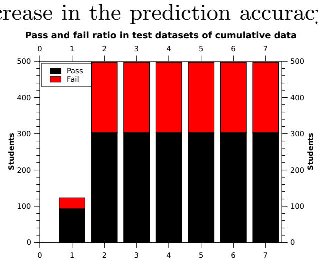

Figure 1: Pass / fail ratio in test datasets of cumulative data

Table 3: Weekly accuracy percentages

	Mean	Median	StDev	Min	Max
Week1	38.40	39	3.67	32	46
Week2	46.97	47	1.59	45	53
Week3	59.93	60	2.83	49	64
Week4	72.07	74	6.44	54	80
Week5	74.37	78	8.83	47	83
Week6	82.30	84.50	5.67	59	87
Week7	80.67	84	9.47	50	88

intelligence classifier. Looking at the maximum value in Table 3, the prediction accuracy rises to 88% on week 7. The notable increase in the accuracy starts in week 4 (i.e. with median accuracy of 74% and the maximum accuracy of 80%, allowing the teachers to have a rough estimate about the students who are likely to pass or fail. The results reported in this paper are based on 30 runs for each experiment.

6.2 Experiment II: Analysis of feature vector

As highlighted before, one of the main purposes of analysing the presented data is identifying weaker students as early as possible and therefore finding ways of improving their performance. However, there are many features collected from the online learning platform and identifying the "more relevant" features from the entire feature vector (of size 15) is of importance. Therefore, each of the features, have been singled out and used both for training the swarm intelligence classifier as well as the evaluation phase. The summary of the solo performance of these features are reported in Fig. 3 and Table 4. For instance, feature 13 (F13 or 'playing') in all weeks (except week 1, 2 and 3) is the most influential feature and has returned the highest prediction accuracy. While the grade correlation of this feature is only 0.41, this finding highlights the role of watching videos in the learning process. Knowing what the feature represents, its value is evident and the algorithm proved capable of identifying this important feature. Identifying the most influential features would entail that the analysis could be focused on the n most important features, instead of stretching the computational power to consider all the input features for predication analysis. The results in this section demonstrate that there could exist some individual features which would provide stronger prediction power when used individually than along with the other features.

6.3 Experiment III: Feature combinations

As shown in Table 4, in order to identify the important features, the three most influential features in each week are labelled 1-3 in brackets. The impact of each feature is calculated by giving the weights of 6 to the most influential feature (shown as (1) in the table), and 3 and 1 to the second two influential features (shown as (2) and (3) in the table). The impact of each feature is then calculated using the aforementioned weights. The six most important features are listed below in the order of importance:

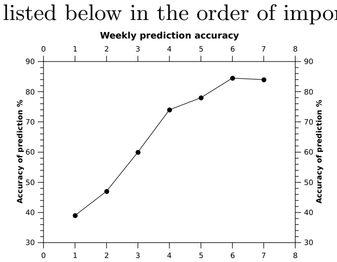

Figure 2: Prediction accuracy of the weekly cumulative data

Table 4: Analysing the impact of individual features (1-15). Prediction accuracies are shown in percentages. The three most influential features in each week are labelled 1-3. The impact of each feature is calculated by giving the weights of 6 to the most influential feature (shown as (1)), and 3 and 1 to the second two influential features (i.e (2) and (3)). The impact of each feature is calculated using the weights.

	Wk1	Wk2	Wk3	Wk4	Wk5	Wk6	Wk7	Impact
F1	32	39	49	74**(2)**	76**(2)**	84**(1)**	83**(2)**	**15**
F2	32	39	39	39	39	39	39	
F3	34	45	42	44	45	46	47	
F4	32	39	39	54	34	41	74	
F5	45**(3)**	59	65**(1)**	71**(3)**	75**(3)**	73**(3)**	74	**9**
F6	32	39	39	41	39	39	39	
F7	32	39	39	39	39	39	39	
F8	32	39	39	39	39	46	45	
F9	50**(2)**	61 **(2)**	57	68	70	78**(2)**	77	**9**
F10	58**(1)**	62**(1)**	65 **(1)**	69	71	72	73	**18**
F11	32	39	39	39	39	39	39	
F12	32	39	39	39	39	39	39	
F13	32	39	58**(3)**	82**(1)**	83**(1)**	84**(1)**	85**(1)**	**25**
F14	38	52**(3)**	60**(2)**	71**(3)**	74	78**(2)**	82**(3)**	**9**
F15	32	39	40	40	40	40	40	

1. F13: Periodically log when video is playing
2. F10: View set of shared files
3. F01: Play video
4. F05: Load media
5. F09: View my files
6. F14: Play region and view thread.

The top six features include a combination of *individual* learning activities (e.g. playing a video to watch, as well as viewing the files saved by the student themselves) and *social* learning activities (e.g. periodically making notes and logging information while watching a video, which could be uploaded by the student themselves or their classmates, knowing that the logged items are visible to the rest of the students) all contributing to the learning process. Investigating the above list, one of the interesting observations is that the social learning activity (of interacting with the posted video) has had the largest score (i.e. 25 as shown in Table 4) and is identified as the most important feature.

In the first part of this experiment, the six highest impact features shown before are selected as input to the system and results are demonstrated in Table 5. While the results are comparable to the previous experiment when all the features where used, the outcome exhibits a slight reduction in the prediction accuracy which could be due to some of the conflicting nature of the features (e.g. combining features which are as diverse as having the impact of 25 and 9). Please note that this hypothesis should be treated with caution as a more in-depth analysis is required to verify this thought. In the second experiment of this section (and in an attempt to explore the previous hypothesis), only two of most significant features (which are the social learning features) are used; the two features used are F13 (periodically log when video is playing) and F10 (view set of shared files). As shown in Table 6, the results demonstrate the highest prediction

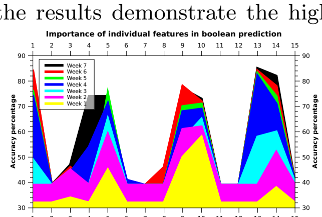

Figure 3: Impact of using individual features. Layers in this diagram represents accuracy of features in each week.

accuracy found on this dataset from week 4 of the term. The median prediction accuracy for week 4 is 83% which is 10% and 9% higher than when six most important features and all features are used respectively (see Tables 3 and 5). Comparing the prediction accuracy reported in Tables 3, 5 and 6 shows that while using the two most important social features, does not improve the prediction accuracy at the very early stages of the term (week 1, 2 and 3), it does enable a stronger prediction from the middle week (week 4) onwards. While this may or may not be extendible to other case studies, this finding highlights the usefulness of investigating the positive or negative nature of social features in online learning environment.

6.4 Discussion

Here, the key research questions raised in Section 1 are discussed next and various aspects of the findings are analysed. As stated in the first research question, this paper applies the Stochastic Diffusion Search (SDS) to classify educational data. The potential and strength of the this algorithm is demonstrated in the results and the flexibility of the algorithm to deal with various feature vector is also highlighted. Given SDS's existing 'partial function evaluation' feature (i.e. each micro-feature, or attribute, is used independently of the others in the test phase), and the resulting low computational cost of comparing samples, this algorithm is likely to be particularly useful when applied to problems with huge dimensionality, which is usually the case in educational data analysis. In this context, the link between cheap computational cost and scalability is the subject of an ongoing research. To address the second research question, three experiments are run (see Fig. 4). Neither of the three experiments (using all features, 6 best features, and 2 best social features) are able to provide a reliable prediction in the first three weeks (e.g. less than 60%) of this seven-week course analysed in this paper; it is worth noting that in the first three weeks, when the social features are solely used in the analysis, the algorithm exhibits the worst outcome, possibly due to the lack or reduced social interactions among the students in the very first a few weeks. However, looking at the performance of the algorithm in weeks 4-7, it can be seen that while using all features or the six most significant features are not causing a huge difference in week 4, the gap widens from week 5-7, showing that the use of all features could prove better than the top six features. On the other hand, having picked the two top features (which are inherently social in nature and involve interactions with other students), the algorithm outperforms the other configurations and provides the prediction accuracy as high as 83% in week 4, and up to nearly 90% in week 7. To address the third research question, the role of social features reflecting the social learning activities are investigated. These features are shown to have played a significant role and as highlighted in the fourth research question, identifying the link between the *social* learning activities and the *student success* in this dataset could give insight to course developers and educators with regards to designing and delivering

Table 5: Combining the most influential six features.

	Mean	Median	StDev	Min	Max
Week 1	45.2	45.5	4.41	32	52
Week 2	52.5	52	2.21	48	57
Week 3	59.57	60	2.75	46	63
Week 4	72.67	74	6.22	62	82
Week 5	72.67	75	7.84	57	83
Week 6	78.43	82	8.03	55	86
Week 7	79.77	80.5	4.85	68	87

Table 6: Combining two of the most influential features.

	Mean	Median	StDev	Min	Max
Week 1	32	32	0	32	32
Week 2	39	39	0	39	39
Week 3	54.37	54	1.03	52	57
Week 4	81.4	83	4.00	66	84
Week 5	81.77	82.5	2.42	75	85
Week 6	87.4	88	1.00	85	89
Week 7	87.8	88	0.76	86	89

course activities. Having established a link between social learning and student success, the results highlight the possibility of providing a more surgical feedback (based on the important features verses all features) to the students who are picked as likely to fail by the system. This study has also shown the importance of the social features used which could be of help when providing feedback to students.

7. CONCLUSIONS

The paper demonstrates the ability of the proposed swarm intelligence classifier in dealing with the existing educational data. The simplicity of this algorithm with one tunable parameter (i.e. agent size) makes it an attractive technique to use. One of the key contribution of the paper is to provide evidence that the data collected on our social learning platform (delivered to several thousand students on Coursera), which records the way in which students share, view and comment on each other's work, is related to performance. Specifically, whilst predicting the final fail/pass of students might be difficult on the first few weeks, the prediction accuracy rises to 83% in week 4 and as high as 89% on week 7. Given two of the social features are demonstrated to have played an important role in the prediction accuracy of the algorithm, as the work progresses, the authors will start to look at questions such as what social behaviours are the best predictors of performance? When can such predictions be made? What kinds of social behaviour impact upon the predicted grades of students? Is it possible to help design interventions for students and tutors to help each other? Finally, after several years of building a system through participatory design and concentrating on the user experience, we are now in a position to use a data driven approach to build systems to support communities of learners.

8. REFERENCES

[1] E. Aguiar, H. Lakkaraju, N. Bhanpuri, D. Miller, B. Yuhas, and K. L. Addison. Who, when, and why: a machine learning approach to prioritizing students at risk of not graduating high school on time. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, pages 93–102, 2015.

[2] M. M. al-Rifaie and M. Bishop. Stochastic diffusion search review. In *Paladyn, Journal of Behavioral Robotics*, volume 4(3), pages 155–173. Paladyn, Journal of Behavioral Robotics, 2013.

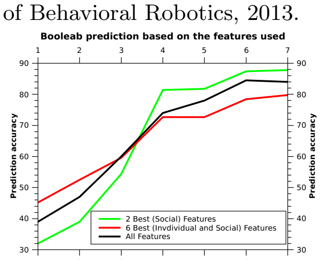

Figure 4: Impact of using various features in the accuracy of the prediction

[3] C. Brooks, C. Thompson, and S. Teasley. Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses (moocs). In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 245–248. ACM, 2015.

[4] C. a. Coleman, D. T. Seaton, and I. Chuang. Probabilistic Use Cases: Discovering Behavioral Patterns for Predicting Certification. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*, pages 141–148, 2015.

[5] N. M. Dowell, O. Skrypnyk, S. Joksimović, A. Graesser, S. Dawson, D. Gašević, P. de Vries, T. Hennis, and V. Kovanović. Modeling learnersâĂŹ social centrality and performance through language and discourse. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.

[6] A. Elbadrawy, R. S. Studham, and G. Karypis. Collaborative multi-regression models for predicting students' performance in course activities. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, pages 103–107, 2015.

[7] Y. Gong and J. E. Beck. Towards detecting wheel-spinning: Future failure in mastery learning. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 67–74. ACM, 2015.

[8] L. P. Macfadyen and S. Dawson. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education*, 54(2):588–599, 2010.

[9] J. K. Olsen, V. Aleven, and N. Rummel. Predicting student performance in a collaborative learning environment. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.

[10] C. Reas and B. Fry. *Processing: A Programming Handbook for Visual Designers and Artists*. The MIT Press, aug 2007.

[11] T. Rogers, C. Colvin, and B. Chiera. Modest analytics: using the index method to identify students at risk of failure. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, pages 118–122. ACM, 2014.

[12] J. Rollinson and E. Brunskill. From predictive models to instructional policies. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.

[13] J. Whitehill, J. J. Williams, G. Lopez, C. A. Coleman, and J. Reich. Beyond prediction: First steps toward automatic intervention in mooc student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining*, 2015.

[14] D. Yang, T. Sinha, D. Adamson, and C. P. Rosé. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-Driven Education Workshop*, 2013.

[15] C. Ye, J. S. Kinnebrew, G. Biswas, B. J. Evans, D. H. Fisher, G. Narasimham, and K. A. Brady. Behavior prediction in moocs using higher granularity temporal information. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 335–338. ACM, 2015.