Development of a Smart Learning Analytics System Using Bangla Word Recognition and an Improved Document Driven DSS

Wali Mohammad Abdullah, Kazi Lutful Kabir
Department of CSE
Military Institute of Science and Technology
Dhaka, Bangladesh
shashi.mist@gmail.com,mlutful.02526@gmail.com

Nazmul Hasan, Sharmin Islam
Department of CSE
Military Institute of Science and Technology
Dhaka, Bangladesh
nazmulcse10@gmail.com,priyaislam122@gmail.com

Abstract—A smart learning analytics system is proposed here. In the field of research, education learning analytics is manipulated as an optimistic way to render learning. In this proposed system, we put forward an idea regarding the design of learning analytics based on document-driven Decision Support System (DSS) and Bangla handwritten words recognition. The proposed learning analytics system aims at classifying the documents via syntactic analysis as well as categorizing them by semantic analysis of their contents. But the documents may be presented as printed format or hand written format. To choose the optimal feature vector for hand written words recognition (especially for Bangla basic characters) projection-based features are mostly efficient. Decision makers use DSS for decision making which is actually a computerized information system. We propose a route through the classified observations for providing the explanation which may improve the acceptance of the decision maker with a better correction rate using kNN algorithm. This system can certainly assist researchers for learning as well as decision making.

Keywords- Learning analytics; artificial neural network; document-driven DSS; data mining; word recognition.

I. INTRODUCTION

The usage of "Big data" can be considered as s growing phenomenon in various fields like computer science, political science, medicine and economics, physics and social sciences. The process of examining these large amounts of data to uncover hidden patterns, unknown correlations and other useful information is referred to as "Big data" analytics.

Analytics is also perceived as a reliable tool for decision-making in the education sector. Analytics in education borrow techniques from different fields. Among these fields Educational Data Mining (EDM), Social Network Analysis, web analytics and Information Visualization facilitate the exploration of data coming from educational contexts [1].

In learning analytics system, we have a huge amount of data in the database. Now, from this database the system should provide the exact data or file which is asked by the users. So, we need a decision support system to make clusters of these documents and to categorize them as required. Now-a-days, data grow too fast. It has exceeded human capacity to retrieve and utilize the content of the documents from the database which has become a problem for the decision makers. Decision makers often use

decision support system in order to get rid of the plight of this information explosion. A successful DSS depends on the accuracy of the results as well as on acceptance of these results by its users [2]–[6].

In Bangladesh, either of the two languages (i.e. English and Bengali) is used for learning. Besides, the document may be preserved in printed form or in hand written form. Among them, segmenting the Bangla (i.e. Bengali) handwritten words from a document and then identify the exact contents from the database is a challenge. Inclusion of projection-based features yields a greater positive impact on recognition rate rather than of inclusion of other features [7]

A. Our Contribution

In our system we propose to use DSS for categorizing the documents. Another contribution is the consequential identification of the exact documents by using projection based features. The rest of the paper has been organized as follows. In Section II, we discussed the present system for learning analytics and associated preliminary concepts. Our proposed system is described in Section III. Finally, Section IV concludes the paper.

II. PRESENT SYSTEM

The primary objective of the Learning Analytics (LA) is to improve the efficiency of the distance learning. In the present LA system, the major activities are the collection and analysis of the materials as well as reporting the data to the learners [8]. Most of the cases the LA system is used to identify the necessity and measurement of the improvements of the study materials, evaluate the performance of the students in a particular course and provide the necessary assistance regarding any kind of suggestion, sometimes it also evaluate the performance of the teacher and provide necessary suggestions. Finally LA makes a very efficient and effective relationship among the study materials, learners, and instructors [1].

III. PROPOSED SYSTEM

In this section we have presented our proposed system for learning analytics. Our system uses an improved DSS for categorizing the documents and to identify the exact contents from

TABLE I
WORST CASE PERFORMANCE OF kNN ALGORITHM

ClassLabels	{3x1 cell}
GroundTruth	[150x1 double]
NumberOfObservations	150
ControlClasses	[2x1 double]
TargetClasses	1
ValidationCounter	1
SampleDistribution	[150x1 double]
ErrorDistribution	[150x1 double]
SampleDistributionByClass	[3x1 double]
ErrorDistributionByClass	[3x1 double]
CountingMatrix	[4x3 double]
CorrectRate	0.9333
ErrorRate	0.0667
LastCorrectRate	0.9333
LastErrorRate	0.0667
InconclusiveRate	0
ClassifiedRate	1
Sensitivity	1
Specificity	1
PositivePredictiveValue	1
NegativePredictiveValue	1
PositiveLikelihood	NaN
NegativeLikelihood	0
Prevalence	0.3333
DiagnosticTable	[2x2 double]

TABLE II
AVERAGE CASE PERFORMANCE OF kNN ALGORITHM

ClassLabels	{3x1 cell}
GroundTruth	[150x1 double]
NumberOfObservations	150
ControlClasses	[2x1 double]
TargetClasses	1
ValidationCounter	10
SampleDistribution	[150x1 double]
ErrorDistribution	[150x1 double]
SampleDistributionByClass	[3x1 double]
ErrorDistributionByClass	[3x1 double]
CountingMatrix	[4x3 double]
CorrectRate	0.9400
ErrorRate	0.0600
LastCorrectRate	0.8667
LastErrorRate	0.1333
InconclusiveRate	0
ClassifiedRate	1
Sensitivity	1
Specificity	1
PositivePredictiveValue	1
NegativePredictiveValue	1
PositiveLikelihood	NaN
NegativeLikelihood	0
Prevalence	0.3333
DiagnosticTable	[2x2 double]

TABLE III
BEST CASE PERFORMANCE OF kNN ALGORITHM

ClassLabels	{3x1 cell}
GroundTruth	[150x1 double]
NumberOfObservations	150
ControlClasses	[2x1 double]
TargetClasses	1
ValidationCounter	1
SampleDistribution	[150x1 double]
ErrorDistribution	[150x1 double]
SampleDistributionByClass	[3x1 double]
ErrorDistributionByClass	[3x1 double]
CountingMatrix	[4x3 double]
CorrectRate	0.9800
ErrorRate	0.0200
LastCorrectRate	0.9800
LastErrorRate	0.0200
InconclusiveRate	0
ClassifiedRate	1
Sensitivity	1
Specificity	1
PositivePredictiveValue	1
NegativePredictiveValue	1
PositiveLikelihood	NaN
NegativeLikelihood	0
Prevalence	0.3333
DiagnosticTable	[2x2 double]

documents, we have utilized projection based features. In the following sections we have presented them elaborately.

A. Improved Document-Driven DSS

The enhancement of the effective communication between the users and DSS is the key element in DSS design. Computer systems are used to solve semi-structured or unstructured problems which are characterized by a complex decision-making process. We have emphasized on one popular algorithm (*kNN* classifier algorithm) to solve this problem. We have used a route through the classified observations for providing the explanation which has improved the acceptance of the decision maker. We can classify all observations into user specified clusters. After implementing the algorithm and then performing accuracy test we have observed that *kNN* algorithm can be used for better correction rate. The worst case, average case and the best case performances of *kNN* algorithm is presented in Table I, II and III respectively.

B. Document Identification System

The documents stored can be either printed or handwritten. Recognition of printed documents won't be a tough task as already developed OCR(for printed) is commercially available for not only Bangla but also many other languages with acceptable recognition rate.

The most challenging job is to recognize the contents from a handwritten Bangla document. In order to accomplish this objective, the first line from the document is segmented as document title. Then, next task is to identify the words of the first line. To recognize the words, step by step character recognition is the suitable approach in this case. Feature extraction is the most important step of character recognition. According to Kabir

et. al. for handwritten Bangla character recognition, projection-based features plays a vital role to constitute optimal feature vector [7].

Three projection-based feature sets have been considered to constitute the feature vector. These are: projection profile features (both left and right) and shadow features.

1) Projection-Based Features:

- Left Projection Profile Features: It represents the projection from left side of a particular character. For all rows, the
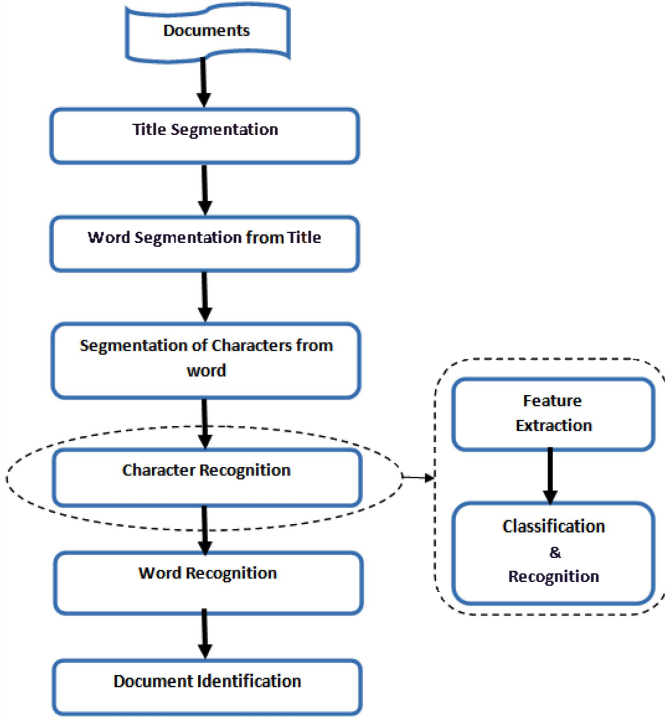
Fig. 1.   Document Identification System Architecture

Fig. 2.   Illustration of right projection profile features to recognize individual characters and word recognition

total number of non-black pixels until reaching the starting left side portion of the character is counted [7].

- Right Projection Profile: It represents the projection from right side of a particular character. For all rows, the total number of non-black pixels until reaching the starting right side portion of the character is counted [7].

- Shadow Features: Lengths of projections of the character images on the four sides and eight octant dividing sides of the minimal bounding boxes enclosing the character are computed. Total 24 shadow features are extracted from each character as lengths of projections on three sides of each octant are considered [9].

Finally an artificial neural network fulfills the tasks of classifier to complete the recognition process. This character recognition step leads to word recognition. After that, the recognized words are checked against the learned contents. If found, then the whole document is included in the cluster of the matched content(s). If not found, then the database will be updated for new content.

We have tested total 15 documents (Printed-10, Handwritten-05) with different header styles (Some samples are presented in Figure 3 and 4). Our system exhibited flawless performance in title segmentation and word segmentation from a document (Table IV). After that, the segmented words of the titles are checked for matching of contents.
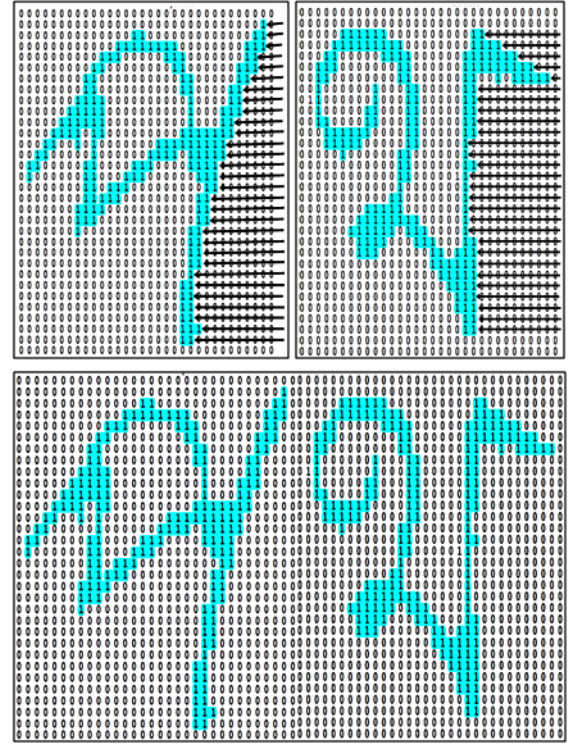
TABLE IV
PERFORMANCE ANALYSIS

Number of Documents Tested	Title Segmentation	Word Segmentation from title	Detection and Classification Accuracy
15 (Printed-10, Handwritten-05)	100% (First line as title)	100%	11 documents (Printed-08, Handwritten-03) has been accurately identified and classified. Accuracy in this particular test set- 73.33%

IV. CONCLUSION

In this paper, learning analytics is recognized as a dominant tool for helping people to reflect on their learning activities and a smart learning analytics system has been proposed. An strategy about the design of learning analytics based on document-driven Decision Support System (DSS) and Bangla handwritten words recognition have been used in our approach. The documents have been classified as well as the contents of the documents have been categorized meaningfully.   For learning purpose, the documents may be preserved in printed format or handwritten. The optimal feature vector (consisting of projection-based features) have been constituted for handwritten Bangla words recognition. DSS, which is actually a computerized information system, has also been used for the purpose of decision making.

যুক্তরাষ্ট্রে বিজ্ঞান ও প্রকৌশলে উচ্চশিক্ষা - পিএইচডি নাকি মাস্টার্স?

By Ragib Hasan

মার্কিন অনেক বিশ্ববিদ্যালয়ে ডাইরেক্ট পিএইচডি করার সুযোগ আছে। অর্থাৎ, বিএসসি ডিগ্রিধারীরা সরাসরি পিএইচডি প্রোগ্রামে ভর্তির সুযোগ পান। বাংলাদেশে অনেকের মাঝে একটা ভুল ধারণা দেখেছি -- পিএইচডি করতে গেলে আগে মাস্টার্স থাকা প্রয়োজন। অন্তত মার্কিন বিশ্ববিদ্যালয়ে সেটা ঠিক না -- সুযোগ্য প্রার্থীদের সরাসরি পিএইচডিতে ভর্তি করা হয়। আর পিএইচডি করতে করতে মাস্টার্স ডিগ্রিটা নেয়া বা না নেয়া অনেক জায়গাতেই ছাত্রের ইচ্ছার উপরে নির্ভর করে। যেমন, আমার গ্র্যাড স্কুল ইউনিভার্সিটি অফ ইলিনয়ের কম্পিউটার বিজ্ঞান বিভাগে মাস্টার্স করতে হলে ৬টা কোর্স আর থিসিস লিখতে হতো। অনেক জায়গায় আবার ৬/৭টা কোর্স করলেই মাস্টার্স নেয়ার সুযোগ আছে। আমার কর্মস্থল **UAB**তে ৩৬ ক্রেডিটের কোর্স করলে কম্পিউটার বিজ্ঞানে মাস্টার্স পাওয়া যায়।

Fig. 3.   Sample Document (Title left-aligned)

Fig. 4.   Sample Document (From Newspaper)

A route through the classified observations has been used for providing the explanations which may improve the acceptance of the decision maker with a better correction rate using *kNN* algorithm. Finally, this system can certainly assist researchers for learning analytics as well as decision making.

REFERENCES

[1] E. Durall and B. Gros, "Learning Analytics as a Metacognitive Tool."

[2] D. A. Savić, J. Bicik, and M. S. Morley, "A DSS generator for multiobjective optimisation of spreadsheet-based models," *Environmental Modelling & Software*, vol. 26, no. 5, pp. 551–561, 2011.

[3] H. Kopackova, J. Komarkova, and P. Sedlak, "Decision making with textual and spatial information," *WSEAS Transactions on Information Science and Applications*, vol. 5, no. 3, pp. 258–266, 2008.

[4] M. S. Gönül, D. Önkal, and M. Lawrence, "The effects of structural characteristics of explanations on use of a DSS," *Decision Support Systems*, vol. 42, no. 3, pp. 1481–1493, 2006.

[5] P.-N. Tan, M. Steinbach, V. Kumar *et al.*, *Introduction to data mining*. Pearson Addison Wesley Boston, 2006, vol. 1.

[6] D. J. Power, "A brief history of decision support systems," *DSSResources. COM, World Wide Web, http://DSSResources. COM/history/dsshistory. html, version*, vol. 4, 2007.

[7] K. L. Kabir, M. K. Shafin, T. T. Anannya, D. Debnath, M. R. Kabir, M. A. Islam, and H. Sarwar, "Projection-Based Features: A Superior Domain for Handwritten Bangla Basic Characters Recognition," in *Proceedings of the 9th International Conference on Intelligent Systems and Control(ISCO)*, vol. 2.   IEEE, 2015, pp. 1054–1060.

[8] Y. Tabaa and A. Medouri, "LASyM: A Learning Analytics System for MOOCs," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no. 5, 2013.

[9] M. Karim, M. Kaykobad, and M. Murshed, *Technical Challenges and Design Issues in Bangla Language Processing*.   IGI Global, 2013.