Bangla Handwriting Recognition Competition

Bangla is the second most popular language and script in the Indian subcontinent and among the top ten popular languages/scripts in the world. As a script, it is used for Bangla, Ahamia and Manipuri languages. Bangla is also the national language of Bangladesh. Despite its importance, research contributions related to recognition of handwritten Bangla script are limited in the literature and in many of them recognition accuracies had been reported based on non-standard sample databases. An unbiased evaluation platform is thus necessary to rank proposed solutions in this domain. This competition on handwritten Bangla character recognition aims to serve as the first such attempt to bring together prospective researchers/groups working in this challenging area of application.

Features of Bangla script:

There are 10 numeric patterns, 11 vowels, 39 consonants and more than 200 compound characters in Bangla alphabet-set. Apart from these, there also exist 10 vowel modifiers and 2 consonant modifiers.

Scope of the competition:

This competition is restricted to recognition of off-line handwritten Bangla (i) isolated digits, (ii) isolated basic characters (vowels and consonants) and (iii) a set of frequently used isolated 150 compound (conjuncts of two or more consonants) characters. There are 10 pattern classes in the digit dataset. The basic character set comprises of 49 patterns, leaving out one consonant (chandra-bindu) that appears above a character pattern (similar to a character modifier). Finally, the list of compound characters consists of 150 frequently occurring conjunct character patterns.

The present competition is divided into four parts, viz., recognition of digits, recognition of basic characters, recognition of compound characters and recognition of all the above shapes (digit + basic character + compound character) together. More specifically, the competition is aimed at designing a 10 class digit classifier, a 49 class classifier for basic characters, a 150 class classifier for compound characters and a 209 class classifier for all.

Evaluation process:

Training samples will be provided to all registered participants for all the three character sets. All the participants will have to submit four different systems, as mentioned before. Additionally, outputs of all the classifiers on respective test datasets need to be submitted as a text file similar to the file 'Classes_labels.txt' to be provided along with the training data. Performances of all four classifiers on test datasets will be considered (weighted average) to rank the final submissions. The weighting scheme will be reported soon.

Sample dataset:

For ease of understanding, please download the file called README.pdf, another text file named 'BanglaHandwrittenCharSamples.txt' containing a few samples from all the three categories, a third text file called 'Classes_labels.txt' containing the ground truths of the samples in 'BanglaHandwrittenCharSamples.txt'.

Registration :

All researchers willing to participate in the competition are requested to register themselves by sending an email (with Cc: icfhr2010@isical.ac.in) to any or both the contacts given below.

Important Dates:

Deadline for registration (to receive training datasets):   February 28, 2010

Deadline for submission of systems: April 15, 2010

Contacts:

Utpal Garain: utpal@isical.ac.in, utpal.garain@gmail.com

Nibaran Das: nibaran@cse.jdvu.ac.in, nibaran@gmail.com

Indian Statistical Institute

www.isical.ac.in

Jadavpur University

www.jadavpur.edu