Research and Work Experience

Saatvik Shah

Malaviya National Institute of Technology, Jaipur

As of December,2015

Table of Contents

Mixed Script Information Retrieval

- **Where** : Zine Research Lab, NIT Jaipur
- **Objective** : Language Identification(LI), Named Entity Recognition(NER) and subclassification in corpus having english + 8 indic languages
- **Methodology** : Hierarchical classification model, combining distinct supervised classifiers for LI and NER with semi-supervised search engine ranking based correction and Wikipedia-based keyword scoring for named-entity subclassification
- **Result** :
  1. Weighed F-Score of **0.8082**.
  2. 2nd amongst 10 participating teams.
  3. Django application implementing the above model
- **Links** : *Publication[1]*, *Presentation*, *Code*, *Django-App*

P300 Recognition in EEG

- **Where** : Zine Research Lab, NIT Jaipur
- **Objective** : Recognition of P300 potential in EEG(recordings of electrical activity in the brain)
- **Methodology** : Created a framework to preprocess and evaluate multiple stacked feature vectors extracted via signal processing techniques(statistical, time-series, frequency, wavelets, etc) with a variety of classifiers
- **Result** :
  1. Accuracy of **97%** at ideal training interval.
  2. Complete framework for future research usage.
  3. GUI + Backend for using P300 with the Emotiv EEG headset.
- **Links** : *Publication(under review)[2]*, *Code*

Recognition of Motor-Imagery EEG

- **Where** : Zine Research Lab, NIT Jaipur
- **Objective** : Stochastic Algorithms for recognition of Motor-Imagery potential from EEG
- **Methodology** : Experimented with multiple stochastic optimization algorithms on a neural-network architecture for detecting multiclass Motor-Imagery potential. Found 2 good techniques.
- **Result** :
  1. BSA-NN used a randomized backtracking swarm on a multilayer neural-net
  2. GSEA used a group based evolution and mutation scheme
  3. Both achieved an average accuracy of $\tilde{\mathbf{69\%}}$ on 3 subjects, giving results better than 21 previous algorithms.
- **Links** : *Publication #1[3]*, *Publication #2[4]*, *Code*

Educational Data Analytics Workbench

- **Where** : Fundamental Research Group, IIT Bombay
- **Objective** : Research Workbench for complete data analysis/visualization of user logs generated on EdX
- **Methodology** : Developed an application, which sat on a Hadoop stack ensuring that logs are cleaned and information mined. Multiple inferences such as usage patterns, question-answer difficulty, etc were visulized as well.
- **Result** :
  1. Django Application running on a multinode Hadoop cluster
  2. Our techniques were taken forward by their research group to improve personalized tutoring as well as EdX, IIT Bombay
- **Links** : *Report*, *Presentation*, *Code*

Optimization of L3 Multicast

- **Where** : Arista Networks, Bangalore
- **Objective** : Optimized Port-Channel setup in L3 Multicast
- **Methodology** : Designed and implemented improvements for port channel handlers and reduced reprogramming of data structures in L3 Mulicast, so as to achieve significant speedup in networks heavily involving port-channels
- **Result** :
  1. Designed and developed the complete scheme to meet requirements
  2. Wrote and ran multiple unit, product and stress tests to ensure a robust system
- **Links** : None, Proprietary work

Centralized Gear Logging

- **Where** : Remote(North East India Frontier Railways)
- **Objective** : Complete system for gear-maintenance logging and centralization on a server
- **Methodology** : Developed an Android App for location sensitive gear search with customized GUIs. Central SQL Database to receive records and website to organize them into usable reports.
- **Result** :
  1. Android Application + Website to achieve a complete system
  2. Under deployment in **20** railway stations, and completely deployed at **2**.
  3. Currently preparing a publication
- **Links** : *Patent*

Trademark Search Engine

- **Where** : Remote(Pixvera - a startup)
- **Objective** : Fast and scalable unsupervised trademark search
- **Methodology** : Developed a complete framework using multiple color, textural, shape and keypoint features with unsupervised algorithms
- **Result** :
  1. Framework for image search
  2. KMeans with stacked feature vectors show best results
  3. Currently preparing a publication
- **Links** : None, Proprietary work

Miscellaneous

- Connecting the Dots, an Image Processing + Robotics competition at Tryst, IIT Delhi, where my team stood **3rd**
- Relevant Class Projects
  1. Music recommendation system from unrated user listening logs
  2. Movie management system which auto-detects movies in a parent folder and scrapes relevant information from IMDb/RottenTomatoes
- Developed an Android client for a startup(AnalyzeNControl), to display relevant regulatory information
- Developed a regulatory content aggregation system which scrapes and manages data from **10** different websites for AnalyzeNControl
- Participated on multiple competitions on Kaggle, a platform to solve problems in data science

References

[1] **Saatvik Shah, Vaibhav Jain, Sarthak Jain, Anshul Mittal, Jatin Verma, Shubham Tripathi, and Rajesh Kumar.**
Hierarchical classification for multilingual language identification and named entity recognition.
In *Proceedings of the Forum for Information Retrieval Evaluation*, 2015.

[2] **Saatvik Shah, Anirudha Kumar, and Rajesh Kumar.**
A robust framework for optimum feature extraction and recognition of p300 from raw eeg.
Neural Computing and Applications, Under Review.

[3] **Saurabh Kumar Agarwal, Saatvik Shah, and Rajesh Kumar.**
Classification of mental tasks from eeg data using backtracking search optimization based neural classifier.
Neurocomputing, 2015.

[4] **Saurabh Kumar Agarwal, Saatvik Shah, and Rajesh Kumar.**
Group based swarm evolution algorithm (gsea) driven mental task classifier.
Memetic Computing, 7(1):19–27, 2015.