Kishaloy Halder

AS6-05-21, COM 1, 13 Computing Drive,
School of Computing, NUS
Singapore - 117417

D.O.B : November 13, 1988
Email : kishaloy@comp.nus.edu.sg
Phone : +6596142654

ACADEMIC QUALIFICATION

Qualification	University	Year	CPI obtained/Max CPI
PhD (pursuing)	National University of Singapore	2018(expected)	4.5/5
M.Tech in CSE	IIT Bombay	2012	8.65/10
B.Tech in CSE	WBUT	2010	8.91/10

RESEARCH INTERESTS

Recommendation Systems, Information Retrieval, Information Mining on the Web.

PHD RESEARCH PROJECT

Personalized Recommendation in Healthcare Forums
Supervised by: Prof. Kan Min-Yen **[Jan '15 - till date]**

- A two-stage probabilistic graphical model based framework is introduced in which partially observed higher-level interests on topics expressed by each user can be modeled to provide Personalized Recommendation.
- The model is specialized for use in consumer health forum thread recommendation by equating user reported clinical conditions as interests. The condition-specific latent topics of threads and user-interests discovered in the first stage are further optimized in the second stage using a topic regression model. Experiments with two real-world consumer health forums show that the model significantly outperforms competitive state-of-the-art baselines.
- Importantly, the framework imbues the recommendation process with added transparency, allowing a resultant recommendation system to justify its recommendation with respect to each user's interest in certain health conditions.

RESEARCH PUBLICATIONS

- Kishaloy Halder, Min-Yen Kan, Kazunari Sugiyama, "Health Forum Recommendation using an Interest Aware Topic Model", WWW 2017. [under review]
- Aobo Wang, Kishaloy Halder, Min-Yen Kan,"Named Entity Recognition from Chinese Microblog Text", TACL. [under review]
- Kishaloy Halder, Umesh Bellur, Purushottam Kulkarni, "Risk Aware Provisioning and Resource Aggregation based Consolidation of Virtual Machines", Fifth International Conference on Cloud Computing, IEEE Cloud 2012
- Soham Das, Kishaloy Halder, Sanjoy Pratihar, Partha Bhowmick, "Properties of Farey Sequence and their Applications to Digital Image Processing", International Conference on Information Processing, 2010 (**Best Paper Award**)

OTHER RECENT RESEARCH PROJECTS

Named Entity Recognition in Chinese Microblog-texts
Supervised by: Prof. Kan Min-Yen **[Nov '15 - till date]**

- Ambiguity and time-sensitive nature of microblog text make it difficult for conventional Named Entity Recognition techniques to perform well in microblog corpora. Another key problem is the lack of training data in microblog domain. Working on efficient methods to overcome these in informal Chinese microblog domain.
- Developing a novel method that acquires the training data automatically through bootstrapping.
- Identifying time sensitive nature of Named Entities mentions to minimize amount of training data required.

Finding Popular Travel Destination From Tweets
Guided by: Prof. Chua Tat-Seng **[Jan '15 - Apr '15]**

- Developed a web application that uses millions of tweets from Twitter users all across the globe to generate a dynamic list of trending destinations emerging from live reactions of people on social media.
- Even though there are numerous travel sites that provide ranking of destinations, they do not reflect the temporal nature of popularity of Points of Interests (POI) and may also be biased due to the commercial interests of travel agencies.
- Used a range of NLP tools and machine learning models along with efficient data processing framework to build a scalable, efficient and real time system.

Information Extraction from Social Media
Supervised by: Prof. Ooi Beng Chin **[Autumn '14]**

- Developed a system to present a meaningful subset of tweets given a query. Modified off-the-shelf "river-twitter" library to receive tweets in batches, extracted the named entities (using twitter-NER) from them, and finally storing them in a distributed inverted index (ElasticSearch).
- Retrieving all the tweets given a query, performing selection and ranking based on temporal and contextual information present in the tweet and present it to the user
- Segregating computation-tasks into offline and online components optimally to incur minimal retrieval latency

INDUSTRY EXPERIENCE

Flipkart Internet Private Limited, Bangalore
Software Development Engineer, Web-Engagement **[Jan '14 - June '14]**

- Worked in the User Generated Content(UGC) team. Responsibilities included
  - Developing cross-platform apis for website, mobile site, and mobile app serving thousands of requests per second. The Apis present a useful subset of all the product-reviews to the users
  - Automated analysis of aspect wise user sentiment from user-reviews and ratings
  - Building the website notification system to identify and attract potential customers

Symantec Corporation, Pune
Associate Software Engineer, Information Availability Group **[Aug '12 - Jan '14]**

- Worked on **DRAzure**, a collaboration between **Microsoft** and **Symantec** to provide **Disaster Recovery as a Service** with **Windows Azure** as the cloud platform - it involves detailed understanding of cluster computing environment, volume replication as well as the cloud
- Was responsible for designing, developing back-end components to run seamlessly in a distributed environment.

PAST RESEARCH PROJECTS

Risk Aware Provisioning and Placement of Virtual Machines **M.Tech Project**
Sponsored by IBM, Guided by Prof. Purushottam Kulkarni and Prof. Umesh Bellur **[May'11-July '11]**

- **Goal:** Developing a **Risk scoring model** that captures the risk associated with sizing and placement of Virtual Machines (VMs). The number of servers required to host a certain numbers of VMs has to be minimized ensuring that the performance degradation of the individual VMs is within some acceptable threshold
- **Major Challenges:** Traditional bin-packing algorithms don't take into account the time varying resource demands of VMs. Other approaches don't exploit the correlation existing between resource demands of VMs
- **Contribution:** Designed a risk scoring model which depicts the overall chance of getting resource when it is required. Our placement algorithm leads to 32% reduction in the number of servers required to host a set of VMs compared to other state-of-the-art placement algorithms

SCHOLASTIC ACHIEVEMENTS

- Secured **All India Rank 60** out of 107086 candidates (percentile: 99.94) in GATE 2010
- Was among **top 30** qualifying students from all over India in JEST 2010 conducted by Chennai Mathematical Institute
- Obtained **CAS**(Center for Advanced Studies) grant from **IBM** to attend the IEEE Cloud 2012 Conference
- **Winner** of '**Information on Demand**', a technical contest organized by **IBM University Relations** in 2010
- Was awarded **full-scholarship** to attend **Workshop on Research Frontiers in Computing** organized by NUS in 2012

COURSES TAKEN (PG ELECTIVES)

- Knowledge Discovery and Data Mining, Social Media Computing, Big Data Analytics, Advanced DBMS
- Artificial Intelligence, Introduction to Probability and Linear algebra, Foundations of Machine Learning
- New Trends in IT (Virtualization and Cloud Computing), Computer Networks, Mobile Computing
- Implementation Techniques of Relational Database Management Systems

SOFTWARE SKILLS

- Programming Languages : C, C++, C#, JAVA
- Scripting Languages : Python, Perl, Bash
- Databases : MySQL, Couchbase
- Big Data Frameworks : Map-Reduce(Hadoop), ElasticSearch
- Version Control, Build Tools: svn, Git, Maven