CS4400: INTERNET APPLICATIONS (SCALABLE COMPUTING), D-STREAM
NAME: SUDHANSH MEHTA (16340820)

MOBILE BIG DATA: THE FUEL FOR DATA DRIVEN WIRELESS (TUTORIAL PAPER)

Key Contributions/findings: Personalized service and public service planning with user's location information; Informed decisions based on user behavior/user behavior profiling; Usage pattern of smartphones to be utilized for learning mental status of user; Mobile cloud computing (MCC) an answer to resource limitation of mobile devices; Management and encryption of privacy sensitive data with careful data collection design.
Key Technology Insights: GPS receiver for rendering exact location vs. coarse location via network or Call Detail Records; Context Sensing for context aware applications with embedded sensors; Spacio-Temporal Study (using timestamp and geolocation information) for studying human behavior; Distinct features of mobile big data being Multi-dimensional (Multi sensory and Spacio-temporal information), personalized and real-time; Pervasive health computing/real-time health monitoring using multi-sensor data from wearable devices.
Key insights into Scalable Computing: Data Parallel Scalability using distributed computing; Mobile big data with Spacio-temporal information a proactive entity to give rise to new services; Scaling Infrastructure support for increasing Value extraction from massive Volume and high velocity and wide variety of data (3v's of Big data); Component integration ensuring all parts that deliver information grow uniformly; Scalability of Data Mining Algorithms to be kept in mind and being smart to use approximations so as to work at higher levels.

BIG SENSOR DATA SYSTEMS FOR SMART CITIES (REVIEW PAPER)

Key Contributions/findings: Big Data model faces challenges in Internet domain (due to communication infrastructure limitations); Need for approaches and models for handling cross-domain and Spacial-temporal data (time stamp and geolocation info.); Low real-time implementation of data from big sensor data systems (most research work is offline); Lack of widely accepted service architecture and communication protocols a hindrance to practical realization of services in smart cities; Challenges on how to apply knowledge from multiple source domains to a single target domain (proposal for consensus regularization to train classifier on local data as well as using already trained classifiers from already learnt domains).
Key Technology Insights: People domain i.e "people as sensors"(voluntarily generating real-time data); Smart City Framework comprising of Connectivity layer (collecting data), Data center layer (storage), Analytics (predictive analytics), Application Layer (context or industrial applications), End-user (to absorb the outcome of smart cities); Using taxi-cabs as "data mules" to cost effectively perform data collection for the smart city (as cab goes around for daily travels); Compressive Sensing (CS) using compression techniques to reduce power consumption at dense sensor layers; Using Transfer Learning in Neural Networks to reduce the calibration effort of learning a model from scratch (for eg. in sensor based activity recognition).
Key insights into Scalable Computing: Fully distributed protocols with limited knowledge of topology to be developed to overcome the problem of deployment of sensors; Routing protocols need to support in-network combination of the data from devices and identify high level data over individual data pieces from different sensors; Cloud storage (Public, Private and Hybrid) cater to different kinds of organization and except the Private type cloud are highly scalable keeping in mind client demand; Shift from traditional SQL database to NoSQL database to meet growing amount of sensor data (efficient data distribution over servers and dynamic addition of attributes to data records); Need for new addressing mechanisms/ novel routing mechanisms that do not require identification for all the devices in the dense sensor layers.