10 papers

types

importance

Real-world knowledge discovery
Twitter <-
      #Sentiment Knowledge Discovery in Twitter Streaming Data
      Albert Bifet and Eibe Frank

Paper	Type	Importance From Quotes
City Forensics: Using Visual Elements to Predict Non-Visual City Attributes Sean M. Arietta Alexei A. Efros Ravi Ramamoorthi Maneesh Agrawala IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 12, DECEMBER 2014 Forecasting Fine-Grained Air Quality Based on Big Data Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, Tianrui Li KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	City/Urban planning City/Urban analysis Parameters example: 1. Population 2. Real-time nature parameters	From *City Forensics* paper, Correlation mining between attributes from the big data, quoted: "*can then be used **to predict the value of one attribute** given another when that attribute is not readily available.*" An example of value prediction, is by doing city images analysis with "*Broken Windows Theory*". Broken Windows Theory suggests that visual evidence of neighborhood decay (e.g broken glass, graffiti, trash, etc.) correlates with increased levels of crime). Big data of streets images can be found and mined from Google StreetView, Bing Streetside, etc) From *Air Quality Forecasting* paper, **to predict** "...*future air quality, inform people's decision making(eg: whether to go for picnic or jogging in a park) and **government's policy making**...*"
A Scalable Data Science Workflow Approach for Big Data Bayesian Network Learning Jianwu Wang, Yan Tang, Mai Nguyen, Ilkay Altintas BDC '14 Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing	Machine Learning technique	Quoted: "For the United States health care sector alone, the creative and effective use of Big Data could **bring more than $300 billion potential annual value each year** [35]. **Making the best use of Big Data and releasing its value are the core problems** that researchers all over the world long to solve since the New Millennium. " With the right performance tuning of the right data mining techniques (ie: machine learning, bayesian network), we can save budget

		expenditures by predicting outcomes and analysing current values. "Bayesian Network is an ideal tool for causal relationship modeling and probabilistic reasoning. BN is a widely used in modeling, prediction, and risk analysis."
Sentiment Knowledge Discovery in Twitter Streaming Data		Twitter is a **"what's-happening-right-now" tool** that enables interested parties to follow individual users' thoughts and commentary on events in their lives| in almost real-time. It is a potentially valuable source of data that can be used to delve into the thoughts of millions of people as they are uttering them. Twitter makes these utterances immediately available in a data stream, which can be mined using appropriate stream mining techniques. In principle, this could make it possible to infer people's opinions, both at an individual level as well as in aggregate, regarding potentially any subject or event. Twitter is a what's-happening-right-now" tool that enables interested parties to follow individual users' thoughts and commentary on events in their lives| in almost real-time. Twitter data follows the data stream model. In this model, data arrive at high speed, and data mining algorithms must be able to predict in real time and under strict constraints of space and time. Data streams present serious challenges for algorithm design 　　　# Sentiment analysis can be cast as a classication problem where the task is to classify messages into two categories depending on whether they convey positive or negative feelings.
Finding the Missing Link to Industry: LinkedIn Professional Groups as Facilitators of Empirical Research		

Consumer Involvement Model of Fan Page: Mining from Facebook Data of a Real Celebrity Fashion Brand Yung-Tzu Joyce Lin, Meng-Yen Tom Lin, Kuo-Chen Li		

Examples of Real-world Knowledge Discovery

Sentiment Knowledge Discovery in Twitter Streaming Data