USER ANALYTICS: MODELING USER BEHAVIOR AND PREDICTING USER INTENTION

Dr. M. Omair Shafiq

School of Information Technology,

Carleton University,

Ottawa, Ontario, Canada

email: omair.shafiq@carleton.ca

IARIA

The Thirteenth International Conference on Information, Process, and Knowledge Management

eKNOW 2021

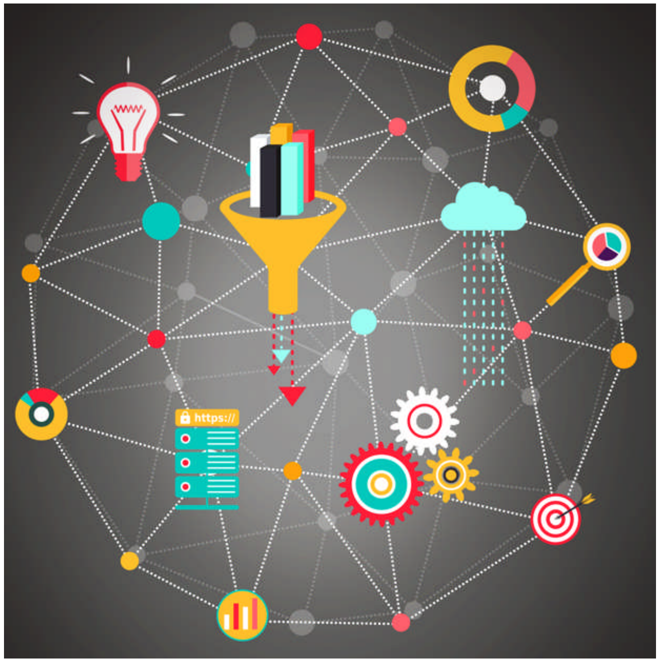
July 18, 2021 to July 22, 2021 - Nice, France

Agenda

- Introduction
- Background
- Challenges
- Motivation
- Solutions
- Open questions
- Conclusions
- References

User Analytics

- **Users**
  - Customers
    - New Customer
    - Past Customer
    - Returning Customer
    - Etc.
  - Leads
  - Potential Leads
  - Staff
  - Other stakeholders
- **User Analytics**
  - Web interaction
  - Social Media activities
  - Transaction history
  - Demographic information



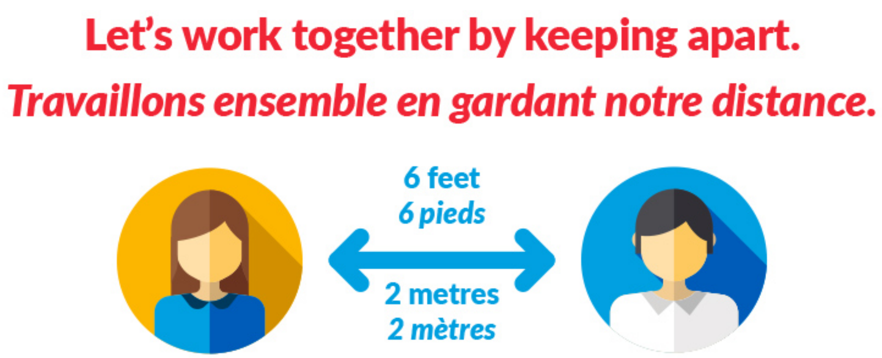
- Analyzing Patterns
- Predicting Patterns

Emerging Digital Age (Pre-COVID)

Almost anything we do these days produces data!

Effects of COVID-19 Pandemic

- COVID-19 Pandemic
- Lockdowns
- Pivoting to remote and online only options
  - Business
  - Schools, Universities
  - Government and private organizations
  - Personal communications and activities

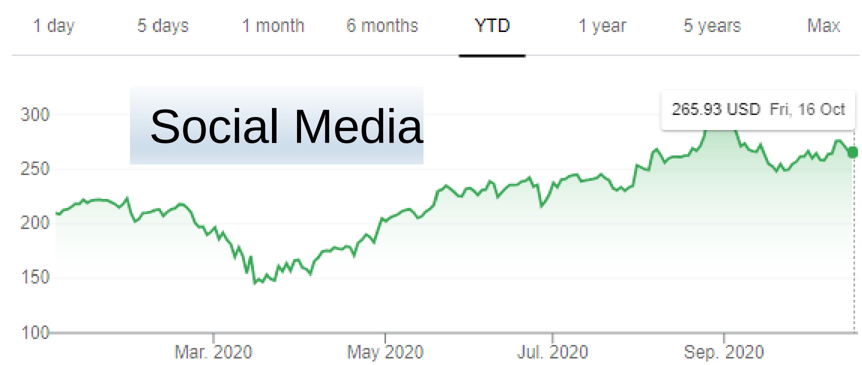


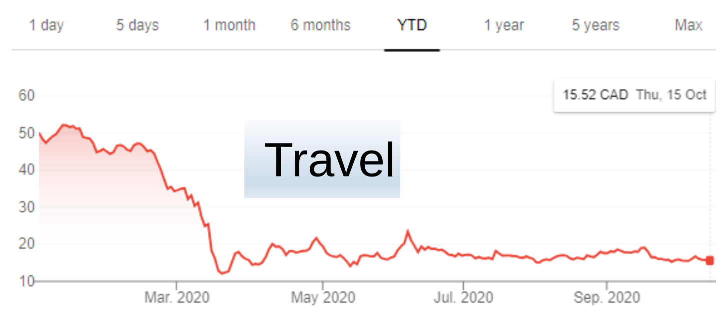
Challenges due to COVID-19 Pandemic

- Online presence
- Web traffic monitoring and tracking
- Still limited solutions available for
  - Customer care
  - Business protection
  - Student support
  - Government

Potential implications of COVID-19 Pandemic

- General Communication
  - Zoom
  - MS Teams
  - and more …
- Online delivery of educatio
  - Blackboard
  - Moodle
  - and more …
- Online shopping
  - Amazon
  - Daraz
  - and more …
- Food delivery
  - Zamato
  - Food Panda
  - Uber Eats
  - and more …





Data overload

- Data is generated by users interacting with online
  ap
- W

data overload
has intensified

Motivation

- Let machines process machine generated data, not humans !

Enhanced and Improved Analytics

Formalizing Machine Data

Employing Analytical Solutions to automatically process formalized machine data

Let machines process machine

User interaction online

- Users interacting with online web-based systems, websites
- User interaction data
  - Clickstream data
  - Text-based content

User online activity – data generation

Clickstream data: a closer look

- Users interacting with online systems generating Clickstream data
  - Logs from online systems contains info such as:
    - Pages in a website
    - Timestamps
    - Keyboard and Mouse/trackpad actions
    - Activities/Assignment/Resources accessed

Example of Clickstream data from user interaction

Activity	start_time	end_time	idle_time	mWheel	mWheel_C	mClick_L	mClick_R	m_mov	KyS
Other	2.10.2014 11:25:33	2.10.2014 11:25:34	0	0	0	0	0	84	0
Aulaweb	2.10.2014 11:25:35	2.10.2014 11:25:42	218	0	0	4	0	397	0
Blank	2.10.2014 11:25:43	2.10.2014 11:25:43	0	0	0	0	0	59	0
Deeds	2.10.2014 11:25:44	2.10.2014 11:26:17	154117	6	0	8	0	1581	4
Other	2.10.2014 11:26:18	2.10.2014 11:26:18	0	0	0	2	0	103	0
Other	2.10.2014 11:26:19	2.10.2014 11:26:27	460	0	0	4	0	424	8
Blank	2.10.2014 11:26:28	2.10.2014 11:26:28	0	0	0	1	0	93	0

Other data (text-based content)

- Users interacting with online systems generating text-based contents
  - Comments
  - Reviews
  - Blogs
  - Etc.
- Sentiment analysis
- Beyond sentiment analysis

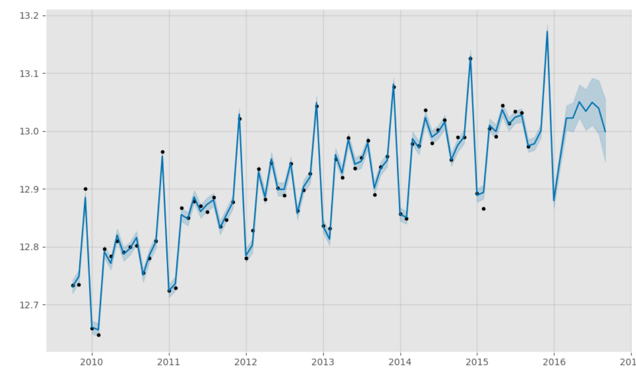
Comment	Sentiment
Excellent service. Food was fresh! Will buy again!	Positive?
The grocery items were delivered past their expiration date and had to be thrown out because they were not edible any more	Negative?
Exactly what I needed for my meal. Very fresh.	Positive?
A few bad or moldy raspberries in the box. But rest I enjoyed them immensely and will buy them again!	???
Received my limes at home in Pandemic. I appreciate the delivery, but I can't afford their price.	???
Not happy with the strawberries I received this time, but I do not have time to go to the store. Will probably end up buying them again from Amazon.	???
Very fresh berries. I will order them again this month before my prime membership expires. Might switch to Instacart then.	???

Tracking users: What to track?

- Pattern detection
  - Based on the current and past interaction data, (how) did the customer end up purchasing?
- Pattern prediction
  - Based on the current and past interaction data, what are the chances that the customer will purchase now or in future?
- Intention detection
  - Based on current and past interaction data, is the customer currently intending to be leaving or returning?
- Intention prediction
  - Based on current and past interaction data, what will be the intention of the customer to leave or return?

Beyond simple tracking

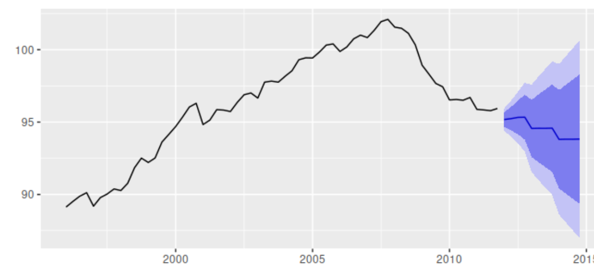
- Detection to Prediction
  - Classifying unseen data (e.g., current user behavior) based on past or historical data
  - Forecasting user behavior for future based on past or historical data



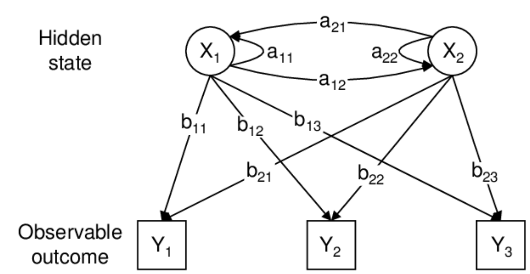
- Pattern to Intention
  - Repeating or recurring events of users , sequential, seasonal, periodic, etc.
  - Objective, intent, plan of action of users
    - may not be visible or obvious in events

State of the art techniques

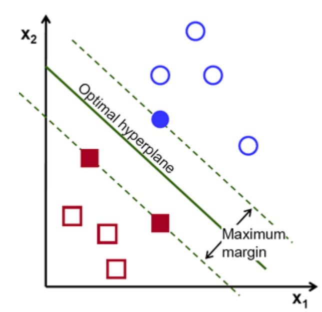
- Researchers have focused on the advantages of Machine Learning (ML) models for interaction data



- Some of the techniques are
  - Classical time series analysis
    - Auto Regressive Integrated Moving Average (ARIMA)

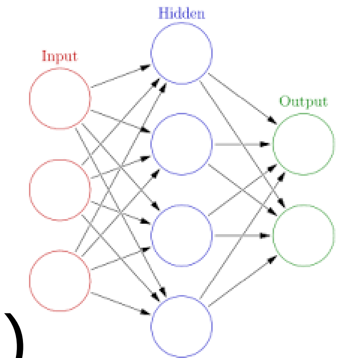


  - Hidden Markov Models (HMM)
  - Support Vector Machine (SVM)

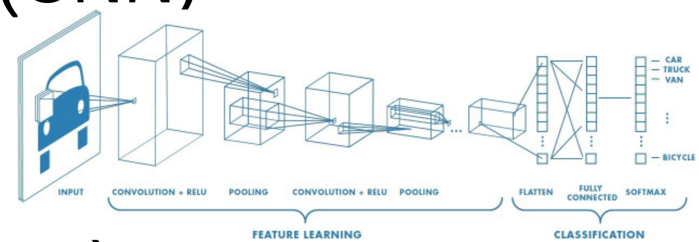


State of the art techniques

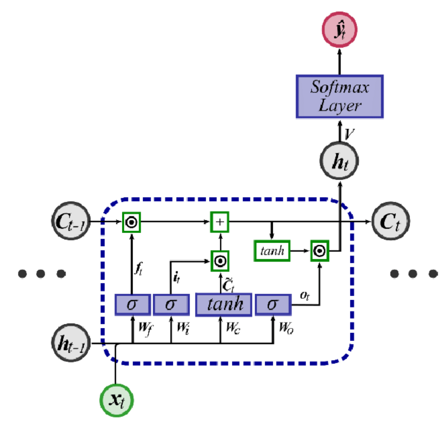
- Artificial Neural Network (ANN)



- Convolutional Neural Network (CNN)



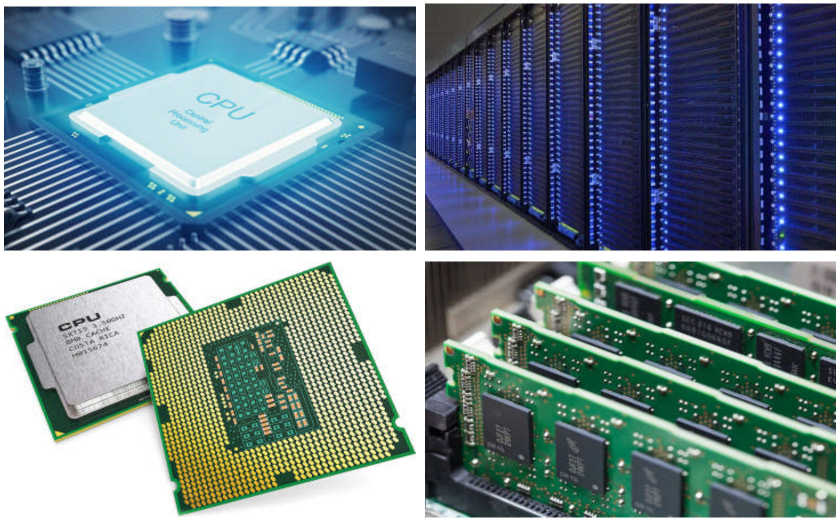
- Recurrent Neural Networks (RNN)
  - Long Short-Term Memory-NN (LSTM)



- and etc…

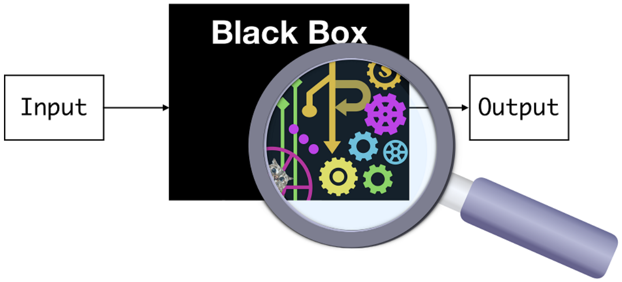
Limitations of state of the art solutions

- Computation intensive

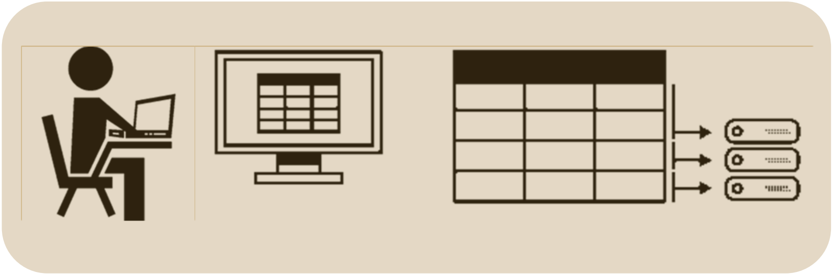


- Memory intensive

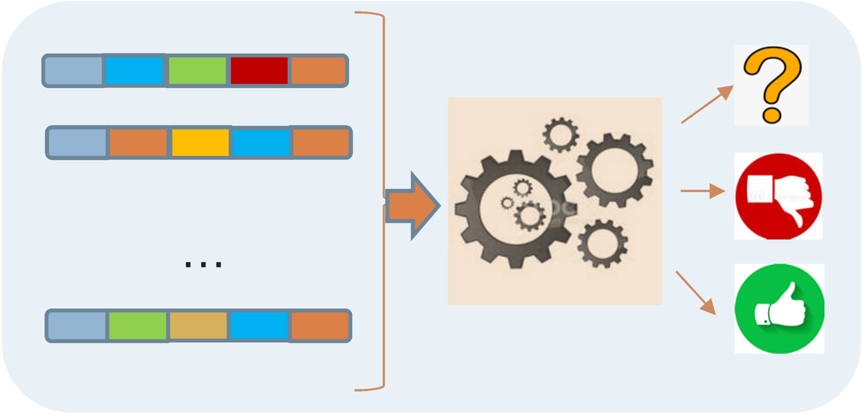
- Limited explainability
  - Work like a black-box



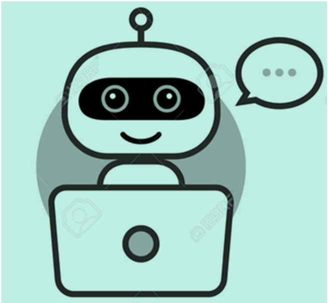
Solutions



User online activity – data generation

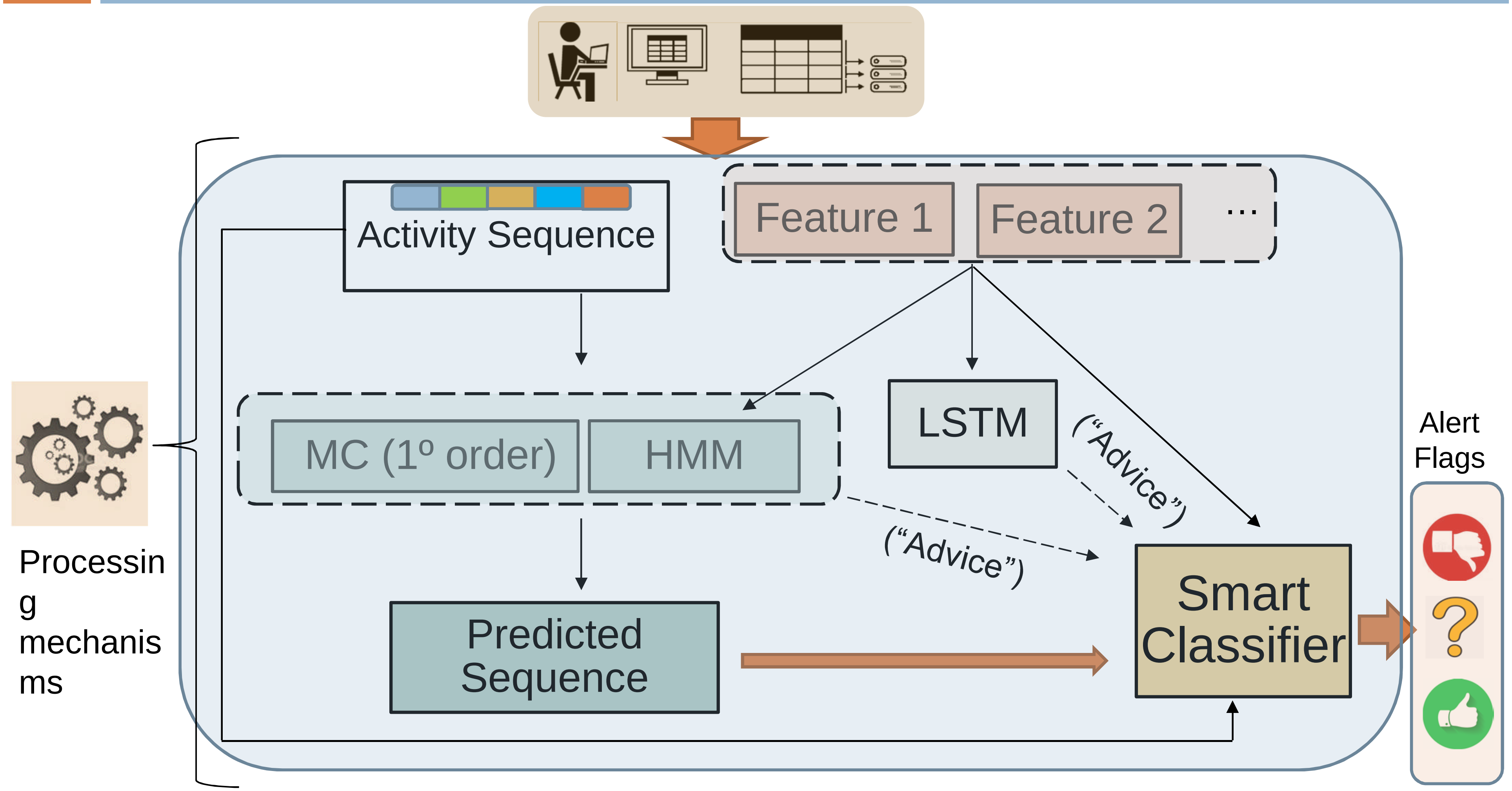


Processing, analyzing, classifying, prediction



Alert mechanisms

Solutions: an example



Interdisciplinary Research and Industrial Use-cases





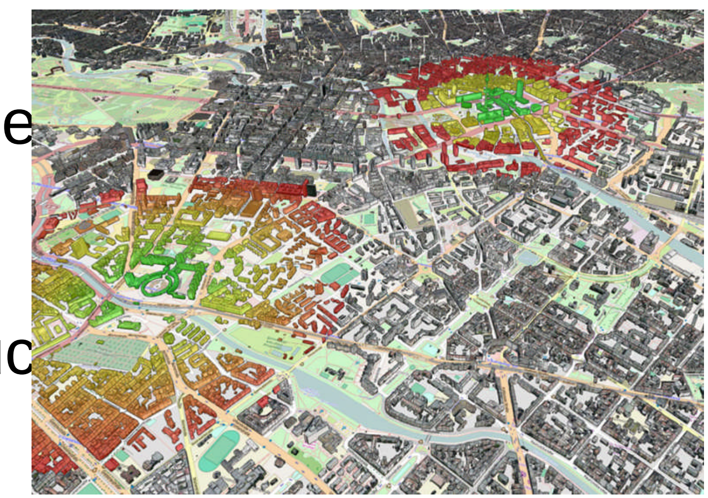


Social Media Analytics

- Collecting, organizing, analyze data from Social Media
- Facebook
- Twitter
- Blogs
- Instagram etc.
- Analyzing and identify patterns of activities and latest trends
- Predicting future trends

Urban Analytics

- Exploratory analysis of different types of complaints reported by citizens for different departments, at different times
- Multiple Linear Regression of predictin time taken to resolve tickets
- Correlation with public data sources an social media (i.e., Twitter)
- Measuring and predicting performance of different organizations or departments within suc organizations



Conclusions

- The use of hybrid systems combining classic machine learning models with deep learning models
  - allows early detection of patterns and intentions in monitoring online user activity
  - enables better explainability of predictions
- The quality of data produced and captured from online systems is important and often neglected
- Quality is measured by how useful and effective the data and the system could be when assisting users in a timely fashion
- Respect privacy of users, follow all the legal, ethical and any other laws, regulations and guidelines
- Proper and explicit consent of users to collect user activity data
- Pattern to Intention of Users

References

- Yasuko Matsubara and Yasushi Sakurai. 2019. Dynamic Modeling and Forecasting of Time-evolving Data Streams. In The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2019), August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA, 11 pages. https://doi.org/ 10.1145/3292500.3330947
- Chaochao Chen, Kevin Chen-Chuan Chang, Qibing Li, and Xiaolin Zheng. 2018. Semi-supervised Learning Meets Factorization: Learning to Recommend with Chain Graph Model. ACM Trans. Knowl. Discov. Data 12, 6, Article 73 (October 2018), 24 pages. https://doi.org/10.1145/3264745
- Gene P. K. Wu and Keith C. C. Chan. 2020. Discovery of Spatio-Temporal Patterns in Multivariate Spatial Time Series. ACM/IMS Trans. Data Sci. 1, 2, Article 11 (May 2020), 22 pages. https://doi.org/10.1145/3374748
- Joshua C. Chang. 2019 Predictive Bayesian selection of multistep Markov chains, applied to the detection of the hot hand and other statistical dependencies in free throws. R. Soc. open sci. 6: 182174. http://dx.doi.org/10.1098/rsos.182174
- Azzedine Boukerche, Lining Zheng, and Omar Alfandi. 2020. Outlier Detection: Methods, Models, and Classification. ACM Computing Surveys 53, 3, Article 55 (June 2020), 37 pages. https://doi.org/10.1145/3381028

References

- Min Du, Feifei Li, Guineng Zheng, and Vivek Srikumar. 2017. DeepLog: Anomaly Detection and Diagnosis
- Alexis Amezaga, M. Omair Shafiq, “Monitoring and Predicting Online Activities of Students: An HMM-LSTM based Hybrid Approach”, Technical Report, Carleton University, Canada, September 2020.
- Archika Sharma, M. Omair Shafiq, “Retail Customer and Market Proclivity Assessment using Historical data and Social Media Analytics”, Technical Report, Carleton University, Canada, August 2020.
- Tobias Hatt and Stefan Feuerriegel. 2020. Early Detection of User Exits from Clickstream Data: A Markov Modulated Marked Point Process Model. In Proceedings of The Web Conference 2020 (WWW ’20), April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3366423.3380238

Questions

?