Cyber Frauds and Fakes

June 2018 @ UCAS

Dongwon Lee

Penn State / IST
Where Am I From?
Penn State University

- State College, PA
  - Out of nowhere, but close to everywhere
- West: 2.5 hours to Pittsburgh
- East: 4 hours to New York
- South: 3 hours to Washington DC
- North: 3 hours to Niagara Fall
Penn State i-School

- College of Information Sciences and Technology (IST)
  - http://ist.psu.edu/
- 50+ tenure-track faculty on diverse areas
  - CompSci & EE
  - Design
  - Law
  - Psychology
  - Social Science
  - MIS & Business
  - Medical Informatics
  - Library Info Science
Research Group

- PIKE: Penn state Information, Knowledge, and wEb research group
  - http://pike.psu.edu/

- Member: 7
  - 6 Ph.D. students
  - 1 M.S. student

- Data Science / Machine Learning
- Cybersecurity / Privacy
- Social Computing / HCI
Summer Class @ UCAS

- Class materials available at:
  - http://tinyurl.com/y7pbcz64
  - http://pike.psu.edu/classes/ucas/2018-summer/

- Lecture notes to be uploaded after each class in the evening

- Grading
  - Attendance: 10%
  - Exam: 30%
    - Open book
  - Assignment: 60%
What is “Fraud”?

- Oxford dictionary
  - “wrongful or criminal deception intended to result in financial or personal gain”

- Van Vlasselaer et al. (2015)
  - “Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving, and often carefully organized crime which appears in many types of forms”
  - 5 characteristics
Examples of Frauds

- Credit card fraud
- Insurance fraud
- Product warranty fraud
- Healthcare fraud
- Telecommunications fraud
- Money laundering
- Click fraud
- Identity theft
- Plagiarism
Social Phenomenon

- Fraud is social phenomenon
  - Potential benefits for the fraudsters come at the expense of the victims
  - Individuals, enterprises, governments, society

- Serious problem
  - Typical org loses 5% of its revenues to fraud
  - Insurance fraud in US is ~$40B per year
  - Credit card companies lose ~7 cents per $100 transactions due to fraud
Detection vs. Prevention

- Detection
  - Ability to recognize or discover fraudulent activities
  - An *ex post* approach

- Prevention
  - Measures that can be taken to avoid or reduce fraud
  - An *ex ante* approach

- Detection power vs. prevention power
  - Complementary, not independent
Detection Approaches

- Expert-based approach
  - “If-Then-Else” rules
  - Challenges to create & update rules

- Data-driven approach
  - Better accuracy -- Statistically based
  - Operational efficiency
  - Cost efficiency
  - Chance of errors vs. power of accuracy
Data-Driven Approaches

- Supervised vs. Unsupervised
  - Labeled vs. unlabeled samples during learning
- Eg, outlier detection
  - Unsupervised
  - To detect abnormal behavior/characteristics

Table 1.3  Example Credit Card Transaction Data Fields

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Transaction type</th>
<th>Date of transaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time of transaction</td>
<td>Amount</td>
<td>Currency</td>
</tr>
<tr>
<td>Local currency amount</td>
<td>Merchant ID</td>
<td>Merchant category</td>
</tr>
<tr>
<td>Card issuer ID</td>
<td>ATM ID</td>
<td>Cheque account prefix</td>
</tr>
</tbody>
</table>
Figure 1.4 Outlier Detection at the Data Item Level

Figure 1.5 Outlier Detection at the Data Set Level
Reference