Prediction of Enterprise Purchases using Markov models in Procurement Analytics Applications

Adam Westerski, Rajaraman Kanagasabai, Jiayu Wong, Henry Chang

Institute for Infocomm Research
A*STAR

September 2015
**Procurement? Problems? Analytics?**

### Procurement

- **Acquisition** of variety of goods or services from an outside external sources
- **Form**: orders, transactions, tenders, quotations
- **Data**: dates, values, quantities, requesters, vendors ...

### Challenges

- **Optimise costs** by improving efficiency of purchase management (e.g. aggregating purchases)
- **Improve on supplier service** to obtain better quality goods/services.
- **Detect employees** of organisation who try to exploit the system to their own benefit
- **Detect suppliers** which try to benefit at the cost of the organisation

### Analytics

- **Inspect** the data (orders, personnel data, external organisations)
- **Detect purchasing trends** and opportunities to improve efficiency of procurement process
- **Find** anomalies and transactions that stand out from the rest
- **Highlight** suspicious activity

---

Institute for Infocomm Research (I²R)
Procurement Systems in Practice

[Introduction: what has been done so far?]

• **Procurement Management and Information Gathering**
  • procurement management systems | put/approve orders, manage suppliers
    SAP, Oracle, dedicated platforms

• **Data Analytics**
  • procurement management systems + common office tools
    MS Excel etc.
  • **Business Intelligence tools**
    Tableau, Qlikview

• **Fraud Detection**
  • commercial fraud detection frameworks
    SAS Fraud Framework, Oracle Advanced Analytics
  • fraud detection research
    Credit card fraud, insurance fraud, telecommunication fraud
Procurement System Problems

[Introduction: Common Problems of Procurement Support Systems]

- Information overflow | lots of data gathered over long time
- Noisy data | manual input, many different users, sometimes different systems
- No all information is recorded | e.g. little or no record of past fraud
  (frauds are rare but when happen cost a lot)
Research Context: Procurement Analytics System built in a number of applied research projects.
Proposed Solution

[Approach: Future purchase prediction]

Research Problem: predict future procurement orders

Applications: optimise purchase aggregation, fraud detection

APPROACH

Approach: apply Markov Chains to model reoccurring purchase sequences in time and:

• predict single next item purchase
• predict multiple future item purchases
Markov Chains

[Approach: Use of Markov Chains for time series modelling]

- **Markov chains** | many applications: physics, chemistry, finance…
  Hilgers 2006

- **Markov Chains for Data Analytics** | purchase prediction, web traffic analysis
  Bozzetto 2005; Deshpande 2004; Bertsimas 2003

- **Markov Chains in Fraud Detection** | credit card fraud, anomaly detection
  Khan 2003
Procurement Data Model

[Approach: dataset description]

Two main concepts:

- purchase order | record of single purchase placed by an employee on a given date
- purchase order item | detailed list of items/services within a purchase
<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># Purchase Order</td>
<td>141,286</td>
<td>MIN/AVG/MAX #Order PER Requester</td>
<td>1/ 12.7/ 1002</td>
</tr>
<tr>
<td># Purchase Order Item</td>
<td>316,036</td>
<td>MIN/AVG/MAX #Vendor PER Requester</td>
<td>1/ 5.8/ 163</td>
</tr>
<tr>
<td># Vendor</td>
<td>7,887</td>
<td>MIN/AVG/MAX #Item PER Requester</td>
<td>1/ 11.7/ 807</td>
</tr>
<tr>
<td># Requester</td>
<td>11,312</td>
<td>MIN/AVG/MAX Creation Date Difference PER Requester</td>
<td>0/ 25.92/ 491</td>
</tr>
<tr>
<td># Approval Officer</td>
<td>594</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Item</td>
<td>212,652</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dataset observations

[Approach: dataset preparation and cleaning]

Majority of requesters have very few orders...but a few actually make majority of orders those will be of our main interest

Small difference between total purchases and unique item count

reason: manually typed item descriptions
outcome: little repeating patterns
solution: use clustering to group similar purchases
Clustering parameters

[Approach: dataset preparation and cleaning]

- Similarity measure: q-gram distance
- Similarity threshold: number of non-matching q-grams
- Optional Criteria: exclude short strings
- Evaluation setup: 2000 manually annotated order items, 44 requesting officers

Best result → bi-gram + lower case + 24 q-gram similarity threshold
Prediction Algorithm Experiments

[Approach: purchase prediction]

Experiment setting:
• prediction of item description individually per requester
• given last purchase predict:
  • single next purchase
  • multiple next purchases

Algorithms:
• Random Sampling | pick on random from requester past items
• Probability Distribution | Cumulative Distribution Function (CDF)
• Simple Sequential Sampling | only multiple purchases experiment
• Markov Chain | first order Markov chain experiments
# Experiments Results

[Approach: purchase prediction #2]

<table>
<thead>
<tr>
<th>Setup</th>
<th>Ignored Requesters (% of all requesters)</th>
<th>AVG Precision/Recall for Requester (Requester Count / % of Dataset orders / precision / recall)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Precision &gt;= 0</td>
</tr>
<tr>
<td>0.5 train+ Markov+ 20 order set</td>
<td>98.16%</td>
<td>212 / 15.04% / 0.34 / 0.09</td>
</tr>
<tr>
<td>0.5 train+ CDF+ 20 order set</td>
<td>35.58%</td>
<td>6438 / 97.50% / 0.04 / 0.03</td>
</tr>
<tr>
<td>0.5 train+ Random Sampling+ 20 order set</td>
<td>35.58%</td>
<td>6438 / 97.50% / 0.04 / 0.03</td>
</tr>
<tr>
<td>0.5 train+ Sequence Prediction+ 20 order set</td>
<td>35.58%</td>
<td>6438 / 97.50% / 0.04 / 0.04</td>
</tr>
<tr>
<td>0.5 train+ Markov+ 20 order set+ clustering</td>
<td>72.06%</td>
<td>2356 / 78.16% / 0.32 / 0.08</td>
</tr>
<tr>
<td>0.5 train+ CDF+ 20 order set + clustering</td>
<td>37.39%</td>
<td>6214 / 97.05% / 0.15 / 0.13</td>
</tr>
<tr>
<td>0.5 train+ Random Sampling+ 20 order set + clustering</td>
<td>37.39%</td>
<td>6214 / 97.05% / 0.00 / 0.00</td>
</tr>
<tr>
<td>0.5 train+ Sequence Prediction+ 20 order set + clustering</td>
<td>37.39%</td>
<td>6214 / 97.05% / 0.00 / 0.00</td>
</tr>
</tbody>
</table>

Best result  ➔  Markov + multiple purchase prediction (20 orders) + clustering
Conclusions

[Conclusions and Future work]

CONCLUSIONS

• only possible in organization with large procurement database

• single order prediction and use of raw data for predictions gave quite bad results

• only multiple order prediction in longer time frame (1 year) gave satisfactory results

FUTURE WORK

• prediction of vendors which repeat a lot more often item descriptions across purchases

• practical evaluation of prediction results (aggregation capabilities and cost saving implications)

• experiments with different datasets
Thanks for attention!

Questions?