

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 ...

environment

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

behaviour

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

goal

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)

Context

[Approach: A*STAR Procurement Analytics System]

The screenshot displays the ProcurementTracker web application. The browser address bar shows 'localhost:8080/tyr/'. The application header includes the 'ProcurementTracker' logo and 'SEMANTIC LAB' branding. The main navigation menu includes 'Continuous Scan', 'Spending Analytics', 'Pattern Analytics', and 'Visual Analytics'. The primary data view is a table with columns for Date, Item, Vendor, Requester, Value, and SuspicionRank. The table lists several procurement entries with their respective dates, values, and suspicion ranks. Below the table is a pagination control showing '(1 of 14701)' and a page size selector set to '10'. A detailed view for 'PURCHASE ORDER DETAILS (No. 4516023312)' is shown below the table, including fields for ITEM, REQUESTER, VENDOR, VALUE, APPROVAL DATE, PROJECT CODE, VENDOR CODE, and ORG. CURR. Below the details are several alert indicators with their respective suspicion ranks and probabilities, such as 'ORDER ALERT' (Suspicion Rank: 0.831), 'ORDER SPLIT' (Probability: 0.85), and 'BORDER VALUE' (Probability: 0.937).

Date	Item	Vendor	Requester	Value	SuspicionRank
15-02-2012	*****	*****	*****	74408.0	Border Value (0.83)
18-10-2011	*****	*****	****	69550.0	Border Value (0.55)
21-03-2012	*****	*****	*****	75915.4	Border Value (0.52)
21-03-2012	*****	*****	****	61427.9	Border Value (0.51)
16-07-2012	*****	*****	****	38272.0	Border Value (0.5)
07-07-2011	*****	*****	*****	58278.4	Price Deviation (0.5)
12-03-2012	*****	*****	*****	69015.0	Border Value (0.5)
29-03-2011	*****	*****	*****	68582.0	Price Deviation (0.49)
10-03-2011	*****	*****	****	68582.0	Price Deviation (0.49)
11-07-2013	*****	*****	****	62700.0	Price Deviation (0.47)

PURCHASE ORDER DETAILS (No. 4516023312) [Show Full Details]

ITEM: *****	REQUESTER: *****	VENDOR: *****	VALUE: 74408.0
APPROVAL DATE: 05-05-2012	PROJECT CODE: *****	VENDOR CODE: *****	ORG. CURR.: SGD
CREATION DATE: 15-02-2012	PROJECT DESC: *****		ORG. VAL.: 74408.0

ORDER ALERT SUSPICION RANK: 0.831

REQUESTER SUSPICION INDICATOR SUSPICION RANK: 0

VENDOR SUSPICION INDICATOR SUSPICION RANK: 0

ORDER SPLIT PROBABILITY: 0.85

VENDOR SPENDING PROBABILITY: -

INTENSE ACTIVITY PROBABILITY: -

PRICE DEVIATION PROBABILITY: 0

DUPLICATE PAY PROBABILITY: -

BORDER VALUE PROBABILITY: 0.937

UNUSUAL VENDOR PROBABILITY: -

ROUND VALUES PROBABILITY: -

NAME PATTERNS PROBABILITY: -

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

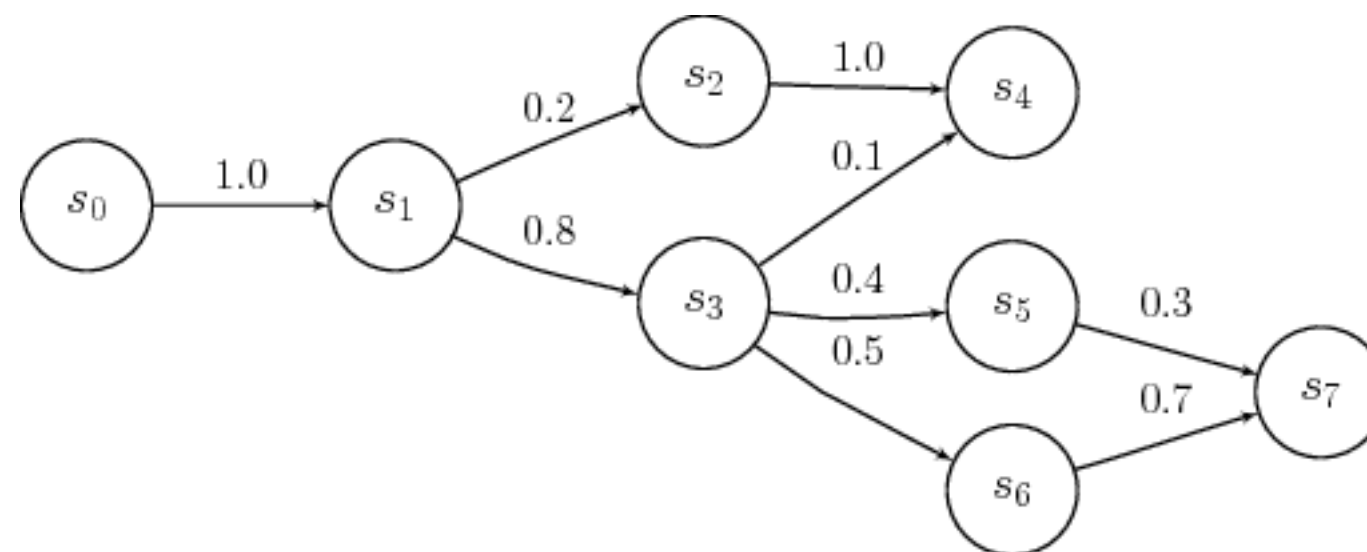
Approach: apply Markov Chains to model reoccurring purchase sequences in time and:

- predict single next item purchase
- predict multiple future item purchases

Markov Chains

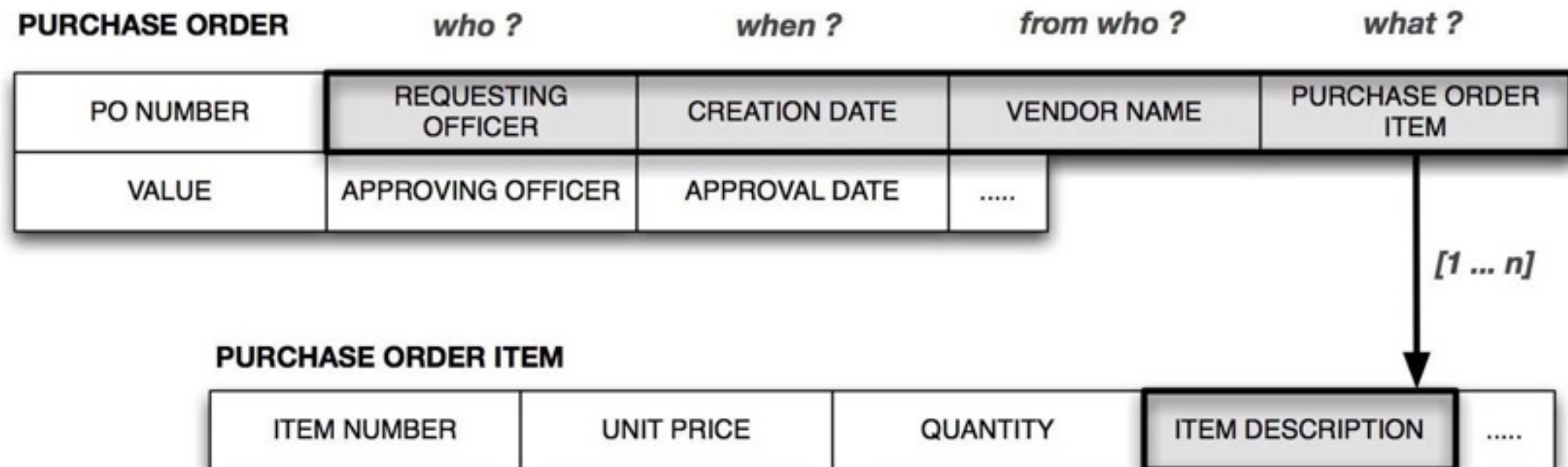
[Approach: Use of Markov Chains for time series modelling]

- **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 a employee on a given date
- purchase order item | detailed list of items/services within a purchase

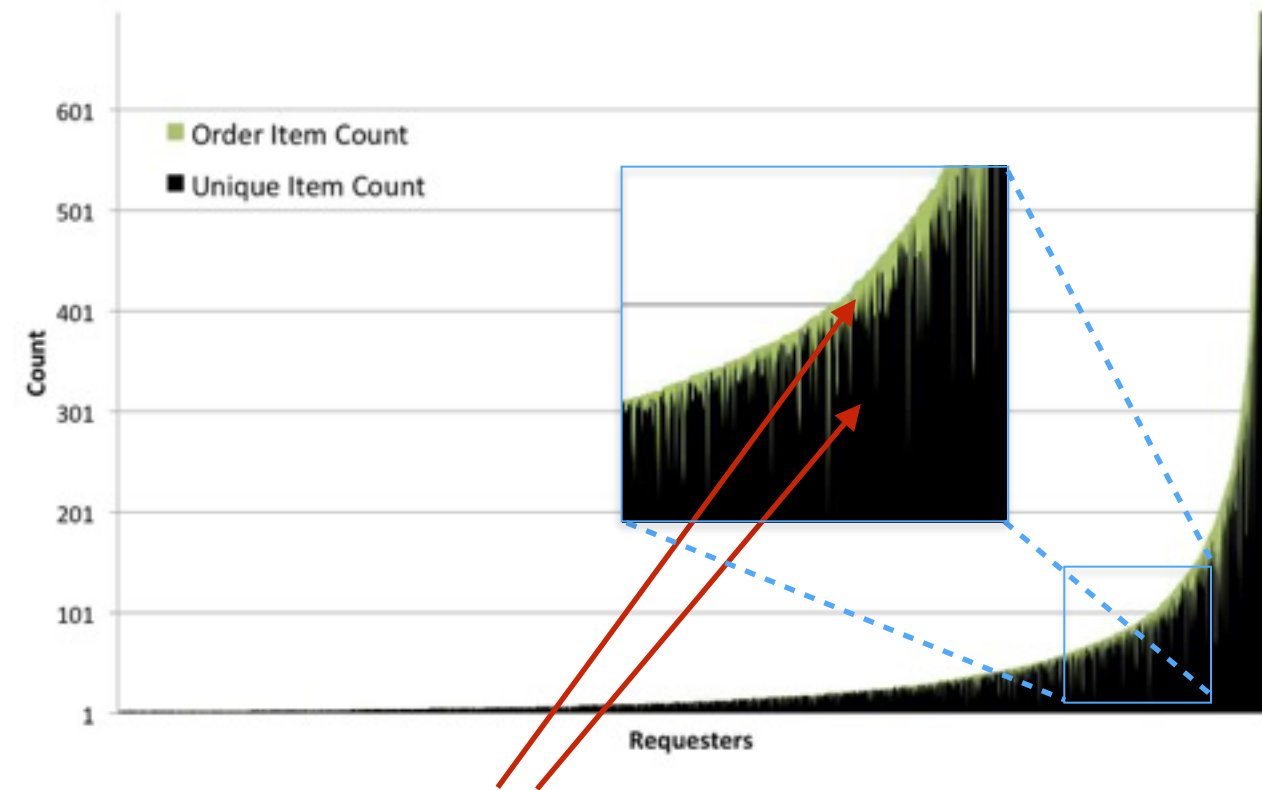
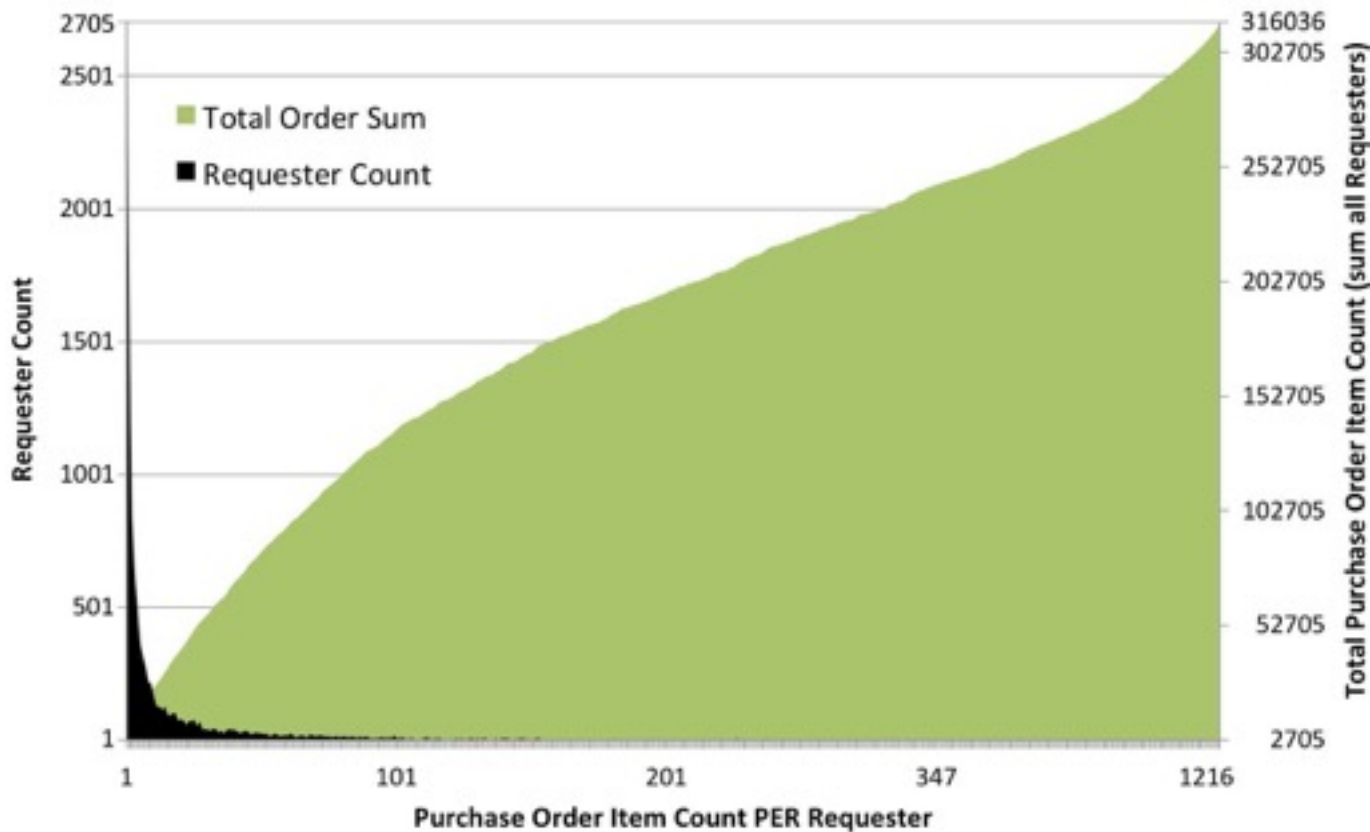
Dataset statistics

[Approach: dataset description]

Metric	Value	Metric	Value
# Purchase Order	141,286	MIN/AVG/MAX #Order PER Requester	1/ 12.7/ 1002
# Purchase Order Item	316,036	MIN/AVG/MAX #Vendor PER Requester	1/ 5.8/ 163
# Vendor	7,887	MIN/AVG/MAX #Item PER Requester	1/ 11.7/ 807
# Requester	11,312	MIN/AVG/MAX Creation Date Difference PER Requester	0/ 25.92/ 491
# Approval Officer	594		
# Item	212,652		

Dataset observations

[Approach: dataset preparation and cleaning]



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

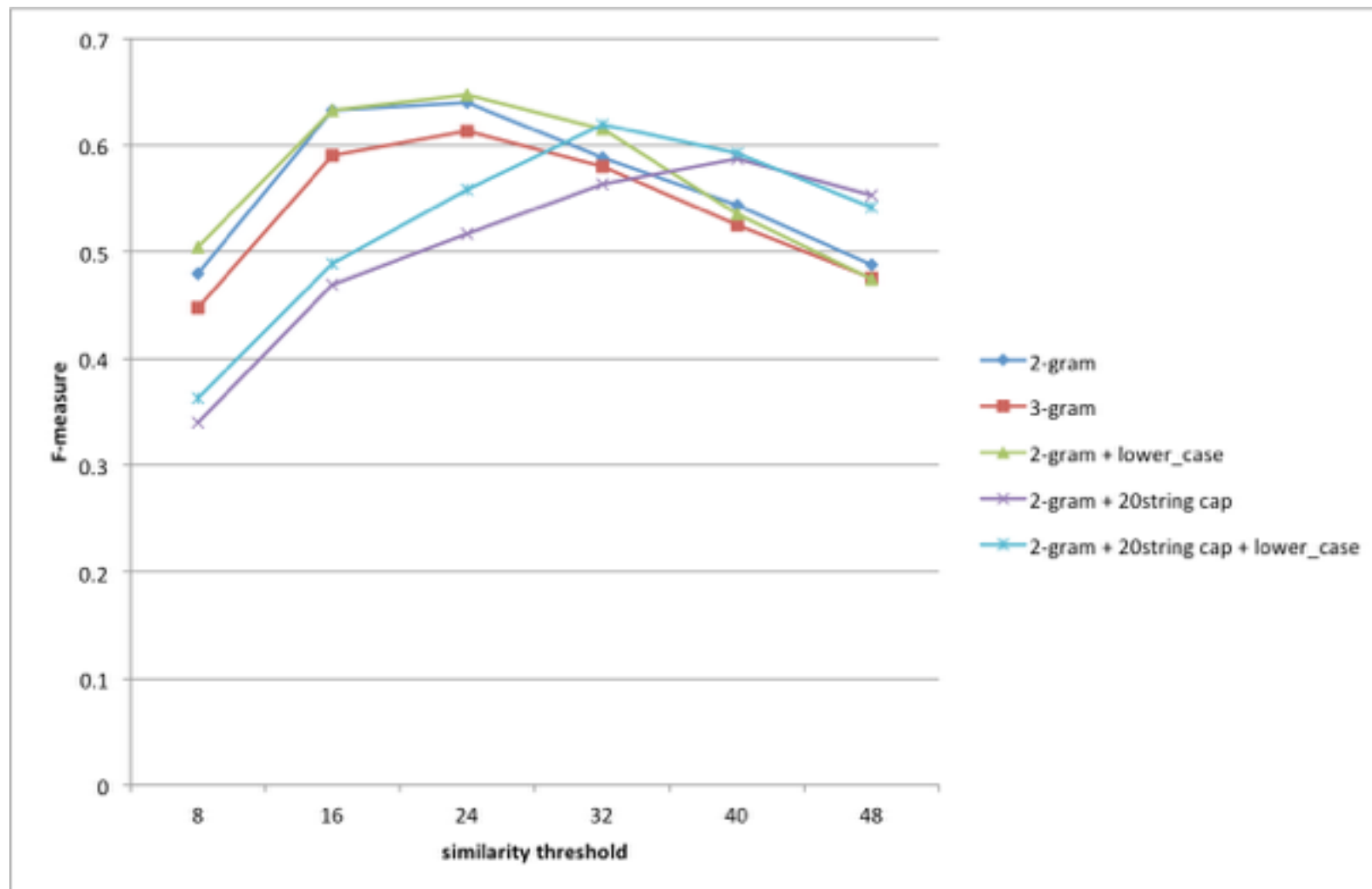
Majority of requesters have very few orders

...but a few actually make majority of orders

→ those will be of our main interest

Clustering parameters

[Approach: dataset preparation and cleaning]



Hierarchical Clustering

- **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]

Setup	Ignored Requesters (% of all requesters)	AVG Precision/Recall for Requester (Requester Count / % of Dataset orders / precision / recall)		
		Precision ≥ 0	Precision > 0	Precision > 0.5
		0.5 train+ Markov+ 20 order set	98.16%	212 / 15.04% / 0.34 / 0.09
0.5 train+ CDF+ 20 order set	35.58%	6438 / 97.50% / 0.04 / 0.03	1319 / 49.41% / 0.20 / 0.17	120 / 0.88% / 0.95 / 0.80
0.5 train+ Random Sampling+ 20 order set	35.58%	6438 / 97.50% / 0.04 / 0.03	1258 / 46.28% / 0.19 / 0.17	110 / 0.58% / 0.94 / 0.84
0.5 train+ Sequence Prediction+ 20 order set	35.58%	6438 / 97.50% / 0.04 / 0.04	1367 / 40.10% / 0.17 / 0.19	98 / 0.54% / 0.93 / 0.91
0.5 train+ Markov+ 20 order set+ clustering	72.06%	2356 / 78.16% / 0.32 / 0.08	1235 / 52.47% / 0.61 / 0.15	600 / 12.88% / 0.96 / 0.24
0.5 train+ CDF+ 20 order set + clustering	37.39%	6214 / 97.05% / 0.15 / 0.13	3134 / 83.76% / 0.29 / 0.25	382 / 2.59% / 0.91 / 0.75
0.5 train+ Random Sampling+ 20 order set + clustering	37.39%	6214 / 97.05% / 0.00 / 0.00	14 / 1.05% / 0.01 / 0.02	0 / 0.00% / 0.00 / 0.00
0.5 train+ Sequence Prediction+ 20 order set + clustering	37.39%	6214 / 97.05% / 0.00 / 0.00	6 / 0.31% / 0.02 / 0.03	0 / 0.00% / 0.00 / 0.00

no clustering

clustering

different requirement thresholds related to precision per requester

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?