Classification in E-Procurement

P.J. Roberts
@UK plc

R.J. Mitchell and V.F. Ruiz
University of Reading

J.M. Bishop
Goldsmiths College, University of London

Presented by r.j.mitchell@reading.ac.uk
Overview

Paper reports work done in a 3-year Knowledge Transfer Partnership Between @UK plc, University of Reading and Goldsmiths College

In fact three linked KTPs

Produced e-procurement system SpendInsight

National Audit Office says could save NHS £500m p.a.

System extended to GreenInsight

Allows procurers to assess environmental as well as economic cost

Key to the systems: classifying products from different sources

This paper focuses on methods used to analyse the product data

Normal best method, SVM, outperformed by KNN and Naïve Bayes
The three KTPs

Three linked projects

- Spidering the web for suppliers of products
  - to build a catalog of web pages

- Classification - to automatically classify data
  - standards eClass, NSV, UNSPCC

- Ranking user search queries
  - return ordered list of matches, most relevant first

During project, opportunities arose to get data on NHS procurement

Project methods focused on such data (though applicable elsewhere)

Led to SpendInsight system
SpendInsight

data collection

saving
spend
data packs

invoice (AP)
purchase orders (PO)
contracts

data collection

classify
benchmark
product and company matching

Classification in E-Procurement- 4
Presented at CIS 2012  © Dr Richard Mitchell 2012
Matching Products

Companies

Products

Unit of measure. Re-sellers.

Item level detail

allows like-for-like comparison, which means that opportunities for savings can be detected such as:

price variance,
price benchmark, and
contract opportunities and contract leakage.

Key is to classify ...
Classification Examples

- **eClass**
  - **F**: Medical and Surgical Equipment
    - **FC**: Surgical Instruments
      - **FCB**: Surgical Instruments
      - **FCC**: Disposable Surgical Instruments
      - **FCD**: Repair of Surgical Instruments
    - **FQ**: Medical Prostheses
      - **FQD**: Pacemakers
      - **FQP**: Joint Replacement Hips
      - **FQN**: Joint Replacement Knees
  - **W**: Office Equipment Telecomms
    - Computers & Stationery
    - **WP**: Paper Items & General Stationery Sundries
      - **WPC**: Copier & Printer Paper

Classification in E-Procurement- 6
Presented at CIS 2012  © Dr Richard Mitchell 2012
Extension for Carbon Footprint

Spend analysis

• eClass classification
• map to CenSA categories
• CenSA carbon analysis

Carbon analysis

E-procurers can assess both economic & environmental cost
Also possible to assess financial cost of being green
Product Classification

Work from Purchase Order (PO) lines

87 NHS trusts ... 2,179,122 PO lines
909 distinct labels
Each line has short description, may be mislabelled

More difficult than standard classification
very many classes
short textual descriptions
often not employing correct grammar
with irrelevant / subsidiary information

Need to automatically classify
Methods Tried

K-nearest Neighbour (preliminary tests show K best at 5)
Rocchio - equal balance of negative and positive prototypes
Naïve Bayes - Bernoulli model
Support Vector Machine - linear models
Two Null hypotheses - as control (random or most often used)

Tested on Reuters data set and on PO data

Performance assessed by F measure - mean of precision / recall
  Macro averaged (across all classes)
  Micro averaged (sum of each class)

$$p_c = \frac{TP_C}{TP_C + FP_C} \quad r_c = \frac{TP_C}{TP_C + FN_C}$$
Macro-Average F Measure

**PO Data**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>C4.5</th>
<th>k-NN</th>
<th>NB</th>
<th>Roc</th>
<th>SVM</th>
<th>NH1</th>
<th>NH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro F</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

**Reuters Data**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>C4.5</th>
<th>k-NN</th>
<th>NB</th>
<th>Roc</th>
<th>SVM</th>
<th>NH1</th>
<th>NH2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro F</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

SVM best on standard text, but not on PO
Micro-Average F Measure

SVM best on standard text, but not on PO

SVM best on standard text, but not on PO

Classification in E-Procurement- 11
Presented at CIS 2012  © Dr Richard Mitchell 2012
Why SVMs do worse

Consider key differences

PO has 2,179,122 documents, Reuters has 9,495
PO has 909 classes, Reuters has 66
PO ~ 8.04 features per doc, Reuters ~62.78

Each feature in the PO data appears in an average of 325.59 documents: in Reuters the figure is 19.38

PO data contains appreciable label noise (where classes are misclassified), the Reuters data does not.

To evaluate significances of these

Project PO data into Reuters, so share characteristics.
Projecting PO Data into Reuters

Suggest

SVM good as retained performance from basic Reuters data
C4.5, KNN, NB retained performance from PO data
Conclusions and Further Work

Classification of the PO Data has been achieved
And the results integrated into SpendInsight and GreenInsight
Savings are being made in NHS and elsewhere
SVM is not the best method for the classification
May be because of class distribution or noise
Further work needed to investigate
C4-5, KNN and Naïve Bayes work well
Further work done by Roberts on pre-processing [in PhD thesis]
And on identifying problem classes (see CIS2010 paper)
Thanks to @UK, rest of KTP team and UK Govt.