

Comparing the application of FCA and LDA to short-text classification

Noel Rogers – D13128921

Background & literature

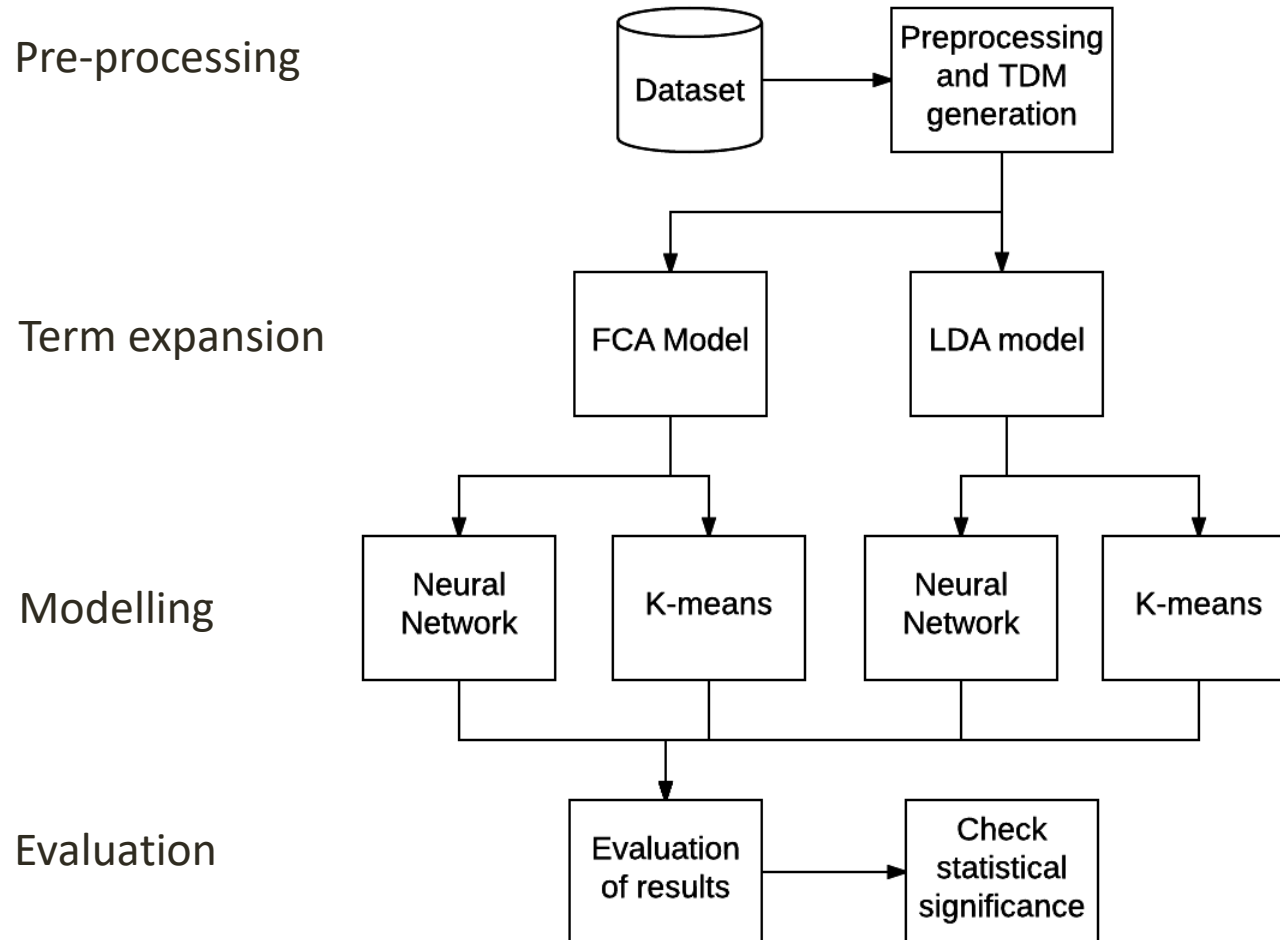
- Text classification: Bag-Of-Words -> Term-Document-Matrix -> machine learning classifier
- Popular classifiers: Naïve-Bayes, SVM and Neural Networks
- Short-Text: too sparse and noisy
- To alleviate this add more non-zero weights
 - Using an external knowledge base (e.g. Wikipedia)
 - Using the patterns within the corpus (FCA and LDA)

Motivation & research question

- **LDA** – Probability based approach to modelling topics in documents
- **FCA** – Uses set and lattice theory to understand concepts
- No comparison between their application to short text has been performed.

“Can Formal Concept Analysis yield better classification accuracy of short-text documents than Latent Dirichlet Allocation, as measured by precision, recall and F-score?”

Design & methodology



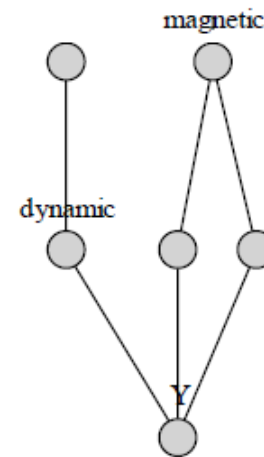
Design & methodology

LDA

- Find two distributions – topics per document (θ) and words per topic (ϕ)
- 3 parameters – α , β and t . Topic number found using perplexity values.
- New weights can be obtained from ϕ times θ

FCA

- Set of concepts \rightarrow lattice
- Proximity of concepts in lattice \rightarrow similarity matrix S
- New weights given by T times S

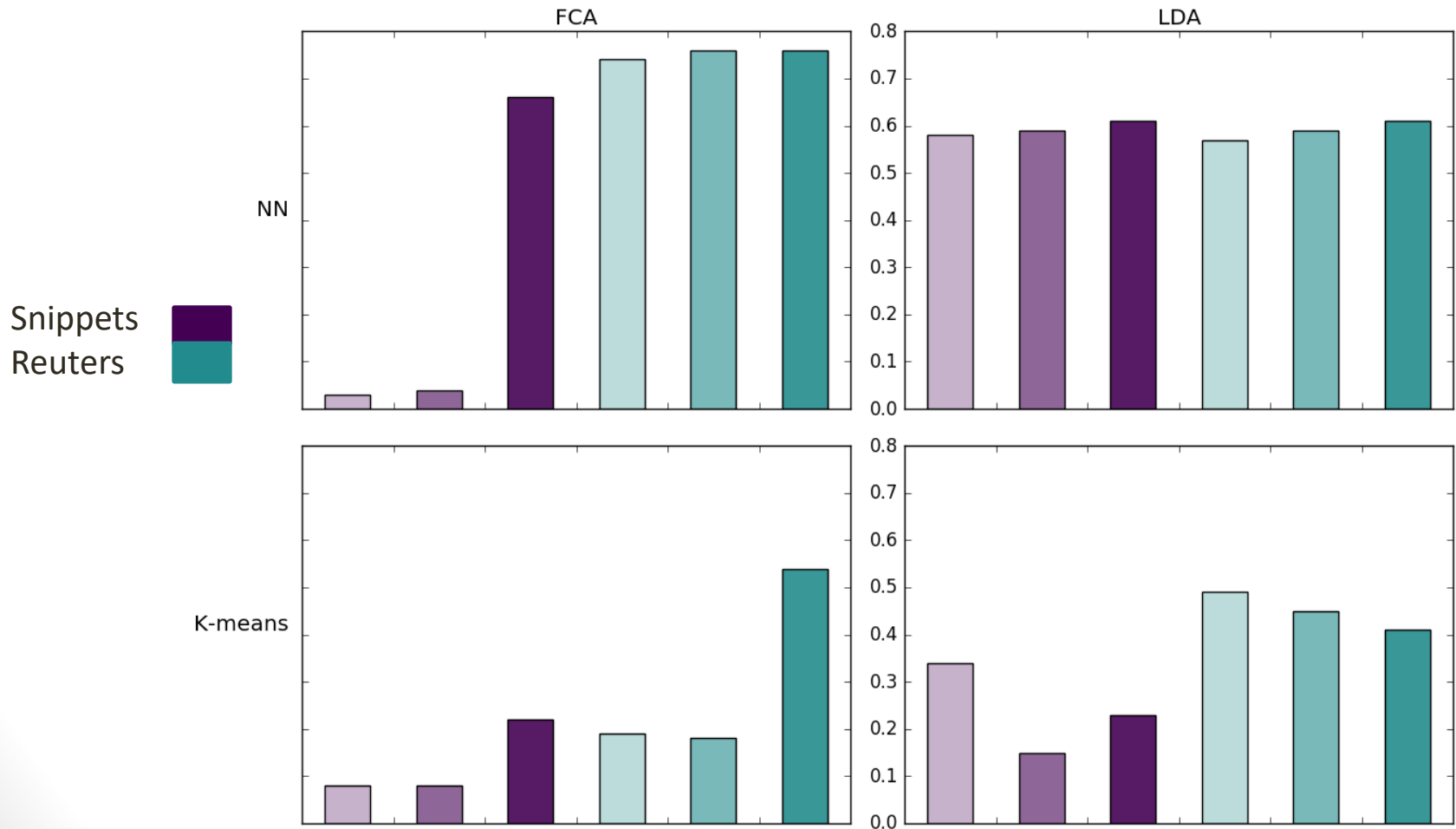


Implementation

- Python used for all steps
- **LDA:** Optimal topic numbers: 181 and 161
- **FCA:** Concept distances ranging between 0 and 12
- **K-means:** K selected using elbow method
- **Neural Networks:** > 85% accuracy on all training sets
- High correlations were found so two additional runs were performed:
 - Remove features with > 0.8 correlation
 - Select top 10% of features using ANOVA

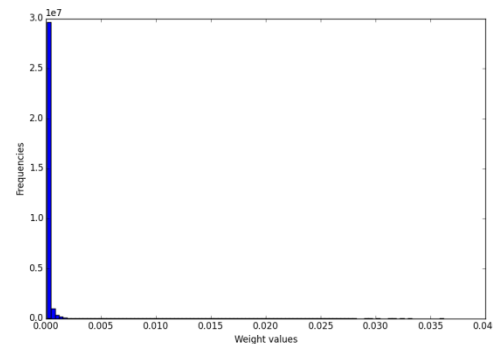
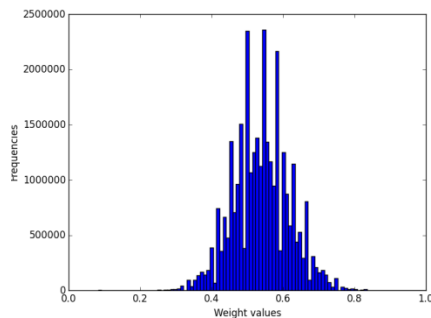
Results & analysis

F-measures from each experiment



Results & analysis

- FCA outperformed LDA in most cases once a small feature set was selected
 - These were statistically significant differences
- Much higher degree of correlations between FCA features than LDA features
- Differences in weight distributions for FCA and LDA:



Contribution to body of knowledge

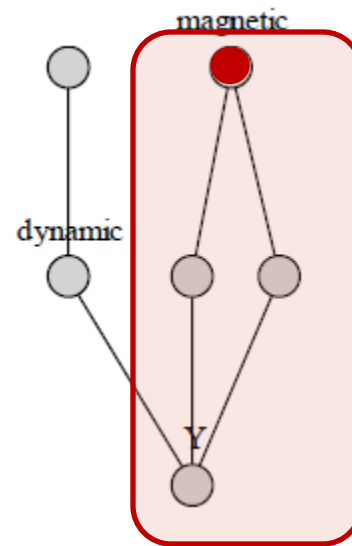
- Directly compared the evaluation performance of LDA and FCA enhanced supervised and unsupervised algorithms.
- No feature engineering: LDA > FCA
- Selection of small number of features: FCA > LDA
- Analysed the differences in the TDMs enhanced by each technique
 - FCA gives high degree of correlation
 - Proximity calculation proposed as likely cause

Future work

- Use fuzzy rather than standard FCA

	y_1	y_2	y_3	...
x_1	1	1	0.7	
x_2	0.8	0.6	0.1	\vdots
x_3	0	0.9	0.9	
\vdots		...		\ddots

- Compute proximities using a local neighbourhood around concepts
- Generate proximities from an iceberg lattice



Thank you

Questions?