Individual events and collective phenomena in large-scale retail: spatio-temporal analysis of sales data

Mirco Nanni
KDD Lab, ISTI-CNR, Pisa
mirco.nanni@isti.cnr.it
Summary

• Individual events in retail sales are usually considered separately
  – E.g. churning customers

• Yet, analyzing the distribution of such events in SPACE and TIME can add significant knowledge
The process

• Define and Extract the events for each customer, separately
• Apply spatio-temporal clustering to find groups of customers with co-located events
• Interpret results to infer causes of the mass-events found
Context

• COOP retail sales (Unicoop Tirreno)
• Data spanning 2 years
• Individual events considered:
  – Churn
  – Fidelity to bio products
  – Fidelity pasta with COOP brand
  – Switch to CFL light bulbs
Two clustering approaches

**SaTScan:** Statistical tool, designed for epidemic studies. Finds areas with density of events larger than outside.

**DBScan:** Data mining tool. Finds areas with high density of events.
Results

Churn

SaTScan

DBScan
Results

Fidelity to bio products

SaTScan

DBScan
Results

Fidelity to pasta with COOP brand

SaTScan

DBScan
Results

Switch to CFL light bulbs

SaTScan

DBScan
The shop closed, and no nearby COOP alternatives were available in that period.
Interpretation of results

SaTScan - churn 2/2

A competitor (a mall) opened a branch nearby, easily accessible from the city through main roads.
Interpretation of results

SaTScan - bio products

Significant clusters in the city of Rome, probably influenced by the presence of national bio products association (AIAB)
Interpretation of results
SaTScan - pasta COOP

Large cluster in Livorno (Leghorn), probably due to the large impact of COOP in general and as a brand in the area.
Interpretation of results
DBScan - CFL light bulbs

Several clusters appear in Rome, in different periods, probably representing various “waves” of advertising.
What next?

- Traditional churn (based on individual) and collective churn analyses can be combined
  - The first looks for models of defection
  - The second helps spotting abrupt defections, due to external causes and not to customer evolution