

SAS Text Miner – from an end-user perspective

Presented by:

Alex Filimon – Dalhousie University / Novus Consulting Group

Agenda

- Introduction
- Text mining introduction
- The use of SAS Text Miner for research and business purposes
- Discussion

Introduction

- Introduction
- Text Mining Introduction
- The use of SAS Text Miner for research and business purposes
- Discussion

■ Alex Filimon

- Part-time professor with Dalhousie University
- Teaching excellence award for 2004-2005
- ABD with University of East Anglia, UK
 - *Knowledge discovery in medical databases*
- Senior partner with the Novus Consulting Group

Text mining introduction

- Introduction
- **Text Mining Introduction**
- The use of SAS Text Miner for research and business purposes
- Discussion

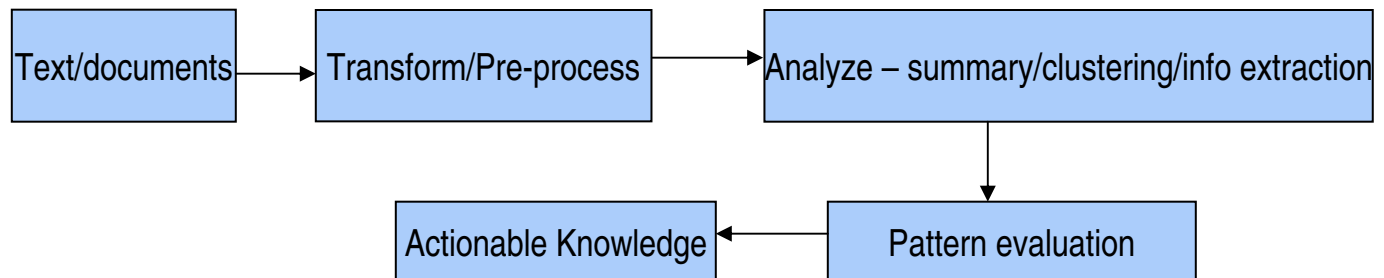
■ Text mining definition

- “the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources”.(Hearst, M.)

■ Key issue

- Create technology that combines a human’s linguistic capabilities with the what modern computers can do.

■ Text mining process



Text mining introduction (2)

- Introduction
- **Text Mining Introduction**
- The use of SAS Text Miner for research and business purposes
- Discussion

■ Extract information

- Useful with large volumes of text
- Identifies people, places and time
- Excellent for security issues, bank fraud and medical papers

■ Summarize

- Reduces the level of detail in a document and retains its main points
- It's difficult to interpret meaning
- It looks for headings, “in conclusion”, other markers of subtopics
- Extremely useful for researchers in the medical field and legal workers

Text mining introduction (3)

- Introduction
- **Text Mining Introduction**
- The use of SAS Text Miner for research and business purposes
- Discussion

■ Clustering

- Used to group documents “on the fly”
- You can narrow down the documents by identifying which topics are relevant
- Very useful for MIS

■ Concept linkage

- Browsing the info, instead of searching
- Useful for identifying links between treatments and diseases
- Dan Swanson – magnesium deficiency => migraines

■ Other techniques

- Categorization (good for businesses)
- Information visualization (academic area)
- Question answering (education and medical areas)
- Topic tracking (competitive analysis)

SAS Text Miner for business and research

- Introduction
- The use of SAS EM for CRM classes
- **The use of SAS Text Miner for research and business purposes**
- Discussion

- Medical field
 - FAQ's
 - Drug design
 - New treatment
- Business
 - Competitive analysis
 - Intellectual property
 - Customer support
 - Content personalization
- Research and education
 - Research on a topic
 - Citation analysis
 - FAQ's

SAS Text Miner for business and research

- Introduction
- The use of SAS EM for CRM classes
- **The use of SAS Text Miner for research and business purposes**
- Discussion

■ Observations

- The “bag of words” approach is problematic
 - “man drives car”; “car drives man”
 - It should take into account a high-level context
 - Therefore, you would need “hand –crafted” rules

■ For medical research we have an “ontology problem”

- “Scientists would rather share each other’s underwear, than use each other’s nomenclature” (Keith Yamamoto)
- Ex.: “heart attack”; “myocardial infarct”; “coronary thrombosis”
- B.S; S.O.B

- *"SAS Enterprise Miner™ and SAS Text Miner are worth their weight in gold," said Dr. Patricia Cerrito of the University of Louisville."*

SAS Text Miner case study

- Introduction
- The use of SAS EM for CRM classes
- **The use of SAS Text Miner for research and business purposes**
- Discussion

- Precision BioLogic
 - Specialty coagulation market
 - FDA filings and patents
 - Research papers about diagnostic kits
- Techniques
 - Concept linkage
 - Categorization
 - Information extraction
 - Topic tracking

- Introduction
- The use of SAS EM for CRM classes
- The use of SAS Text Miner for research and business purposes
- Discussion**

Discussion

Presented by Alex Filimon

Dalhousie University / Novus Consulting Group