Customer Intelligence on Social Media

Rosaria Silipo
KNIME.com AG
KNIME
The Open Analytics Platform
The KNIME Platform: Open for Innovation

- **Powerful:** Legacy ↔ Future Tools
- **Collaborative:** Scientists ↔ Analysts
- **Integrative:** Legacy ↔ Future Data
- **Transparent:** Existing ↔ Future Expertise
- **Agile:** Internal ↔ External Wisdom
The KNIME Analytics Platform
Over 1000 native and embedded nodes included:

MySQL, Oracle, etc.  ETL  Statistics  via BIRT
SAS, SPSS, etc.  Row,  Data Mining  PMML
Excel, Flat, etc.  Column  Machine Learning  XML
Hive etc.  Matrix  Web Analytics  Databases
XML, PMML  Text, Image  Text Mining  Excel, Flat, etc.
Text, Doc, Image  Time Series  Network Analysis  Hive etc.
Web Crawlers  Java  Social Media Analysis  Spark
Industry Specific  Python  WEKA  Text, Doc, Image
Community / 3rd  Community / 3rd  R  Industry Specific
Community / 3rd  JFreeChart  WEKA  Community / 3rd

MySQL, Oracle, etc.  ETL  Statistics  via BIRT
SAS, SPSS, etc.  Row,  Data Mining  PMML
Excel, Flat, etc.  Column  Machine Learning  XML
Hive etc.  Matrix  Web Analytics  Databases
XML, PMML  Text, Image  Text Mining  Excel, Flat, etc.
Text, Doc, Image  Time Series  Network Analysis  Hive etc.
Web Crawlers  Java  Social Media Analysis  Spark
Industry Specific  Python  WEKA  Text, Doc, Image
Community / 3rd  Community / 3rd  R  Community / 3rd
The Problem
The Problem

A major European Telco | Its Forum Site

- Can you tell us what people say about our new product?
- Can you tell us who is supporting the product and who trashing it?
- Of those, can you tell us who is an influencer?
The Data

• The Data Set unfortunately cannot be shared
• Slashdot Forum Data are!
• Slashdot was a public forum built in 1997 and hosting a number of discussions: from software to philosophy, from science fiction to politics.
• Politics was the biggest discussion group
• So, politics is what we analyzed to find out:
  – What users were thinking about a political issue
  – Who was pro and who was con
  – Who was an influencer
The Data
The Politics Group in the Slashdot DataSet

- 24,000 non-anonymous users
- 496 posts
- 140,000 comments
- Most posts have around 200 comments
Text Analytics
Text Analytics: Options

- Tag (Word) Clouds
- Topic Detection
- Topic Shift
- Sentiment Analysis

I find PRODUCT X to be very good and useful, but it is a bit too expensive.
Text Analytics: Workflow

Data loading, preprocessing, and tagging
- Table Reader
- Slapdot
- Strings To Document
- Column Filter
- Dictionary tagger
- Dictionary tagger

Read Data

Levels of attitude
- Document Scoring
- User Scoring

Visualization
- Color Manager
- Renaming and Filtering
- Scatter Plot

Sum of frequencies of positive and negative words per post/comment

Sum of sum of frequencies of positive and negative words per user

Document type is required

Loading MPQA Stanford dictionary for sentiment attribute

Copyright © 2014 KNIME.com AG
Text Analytics: Results

Most negative user: pNutz

Most positive and most talkative user: dada21
Text Analytics: Open Questions

• Is dada21 an influencer?
• Is pNutz an influencer?
• Shall we take marketing actions about any of them?
Network Mining
Network Mining: Options

- User Interaction Graph
- Influencers vs. Followers
- User Network Investigation

[Graph of network with scatter plot and network visualization]
Network Mining: Workflow

- **Read Data**
  - File Reader
  - Anonymous filter
  - Create edge table
  - Filter anonymous node
  - Link post user with referenced post/thread user

- **Create Empty Network**
  - Create Network Content
  - Extract largest sub-graph

- **Centrality Index for authority score**

- **Visualization**
  - Extract largest component
  - Authority & hub weight
  - Feature Table Column Rename
  - Compute authority & hub weight with R

- **Scatter Plots**
  - Scatter Plot
    - Hub vs. authority
    - Hub vs. authority

---

Copyright © 2014 KNIME.com AG
Network Mining: Results

- Dada21
- Carl Bialik from the WSJ
- Doc Ruby
Network Mining: Open Questions

- Is Carl Bialik from WSJ positive or negative about the topic?
- Is dada21 positive or negative about the topic?
- What shall we do marketing-wise with non-influencers such as doc Ruby?
Combining Text Analytics and Network Mining
Text Analytics and Network Mining: Workflow

- **Read Data**
  - Table Reader
  - Row Filter
  - Snippet (Local)

- **Network Creation and Analysis**
  - Network Creator
  - Create edge table
  - Edge table weighted by count of posts
  - Link post user with referenced user/retweet user
  - Filter anonymous node
  - Extract largest component
  - Authority & Hub weight

- **Text Processing and Sentiment Analysis**
  - StackDoc
  - Strings To Document
  - Column Filter
  - Dictionary tagger
  - Document scaling
  - User Scoring

- **Sentiment Analysis**
  - Subjectivity Corpus
  - Preprocessing
  - Create documents
  - Only topic-related ID and Document
  - Tag positive words
  - Tag negative words
  - Accumulation of positive and negative word frequencies
  - Document scaling

- **Joiner**
  - Join results from text mining and network mining

- **Network Mining**
  - Preprocessing
  - Create documents
  - Only topic-related ID and Document
  - Tag positive words
  - Tag negative words
  - Accumulation of positive and negative word frequencies
  - Document scaling

- **Scatter Plots**
  - Scatter Plot (FreeChart)
  - Image To Table
  - Concatenate (Optional)

 visualization of combined results from text mining and network mining

- **Join**
  - Node 201
  - For attitude class
  - Follower score vs leader score
  - Scatter Plot (FreeChart)
  - Image To Table
  - Level of attitude vs follower score
  - Concatenate (Optional)
Text Analytics and Network Mining: Results
Text Analytics and Network Mining: Results
Conclusions
Conclusions

- Is Carl Bialik from WSJ is an influencer and ... neutral.
- Most influencers are actually neutral.
- Worth it keep informed

- dada21 is talking positively about each topic. Worth it to pamper him/her.
- Of the negative talking users, pNutz though obnoxious, is not the main worry. Catbeller is.
Where can I find all this?

White paper, Workflows, and Data is available on the KNIME web site:

http://www.knime.com/white-papers (section Social Media)

Resources

- **KNIME** ([www.knime.org](http://www.knime.org))
- **BLOG** for news, tips and tricks ([www.knime.org/blog](http://www.knime.org/blog))
- **FORUM** for questions and answers ([tech.knime.org/forum](http://tech.knime.org/forum))
- **EXAMPLE SERVER** for example workflows
- **LEARNING HUB** ([www.knime.org/learning-hub](http://www.knime.org/learning-hub))

- **KNIME TV** channel on ![YouTube](https://www.youtube.com/channel/UCQ9z09QWQj4uWQ0RZaQc1Dw)
- **KNIME** on ![Twitter](https://twitter.com/KNIME) @KNIME
- **KNIME** on ![Facebook](https://www.facebook.com/KNIMEanalytics)
Thank You

Free Copy of KNIME Beginner’s Luck Book at KNIME Press

https://www.knime.org/knimepress

Promotion Code: Barcelona2016