“Social Media Intelligence” with KNIME

Social Media is a hot topic. But the actual task of monitoring and understanding what customers are saying about an organization via social media touch points has not been easy until now. Dymatrix, a KNIME partner specializing in all forms of Customer Intelligence and with a very solid reference pedigree, has used KNIME to develop the first application capable of capturing not only a wide variety of social media touch point data, but also of combining that data with an organization’s own customer intelligence.

“When we talk about analyzing social media data, we are primarily talking about capturing and mining text data from a number of different sources and transforming it into something intelligible. The challenge is three-fold and involves: retrieving the social media data, mining it for textual insights, then surface it in a form suitable for marketing and management,” says Stefan Weingärtner, management consultant with Dymatrix. Dymatrix utilized its own expertise to design social media engines that can pull all relevant information from social media networks such as Twitter. “KNIME is the perfect data mining platform for text mining, easily reading all the data we have collected. The KNIME text mining nodes are extremely powerful and – unlike other tools specifically for social media monitoring – the nodes can be customized to cover languages other than English.”

Understanding social media text mining means answering such questions as: which touch points are being used? By whom? Who is listening to whom? Who is talking positively (or negatively) about our products or services? What is the tone of the conversations? How do we compare in these conversations to our competition or others we wish to benchmark against?

“Most importantly, using the KNIME platform means that we can combine social media data with other existing customer intelligence within an organization to develop truly new insights, which can be surfaced in multiple ways, depending on the target audience,” says Weingärtner.

The DynaSociacian application from Dymatrix packages the social media engines together with a series of KNIME workflows. In addition, the application has a full front end giving non-data mining experts the ability to not only set up, but to run and – most importantly – visualize information in a special front end. “A good example of DynaSociacian is demonstrated here by QVC, the extremely successful television shopping organization. By using the software, QVC was able to get a good idea of the tone of, but also of the items being promoted. In this case, we did the example in German.”

Stefan Weingärtner
Executive Board
Dymatrix Consulting

“...actually monitoring and understanding what customers are saying about an organization via social media touch points has not been an easy task until now.”
The Chi-Square Test of Independence

The Chi-square test of independence assumes as a null hypothesis that two categorical variables (col1 and col2) are independent. Such a hypothesis of total independence expects random, equally distributed frequencies, $E(ij)$, for each i and j value pair. The Chi-Square variable, which calculates how far the expected frequencies are from the observed frequencies, has an approximate chi-square distribution with $DF = (r-1) \times (c-1)$ degrees of freedom.

The “Crosstab” node selects two of the input data columns and produces two output tables. The cross table contains the contingency table on two selected input data columns, col1 and col2. A contingency table is an r x c matrix of observations for each i and j value pair.

More details on the crosstab node are available on my blog: http://dataminingreporting.weebly.com/blog.html.