Do people still miss Steve Jobs as the CEO of Apple Inc.?
A Text Mining Approach: Comparing SAS® and R

Pranav Karnavat and Anurag Srivastava
Decision Sciences, Shanti Business School, Ahmedabad, Gujarat

Introduction

Twitter is a powerful social media website which had over 500 million users worldwide. Tweets posted can be analyzed to get insight about relationships and patterns hidden inside the textual data. In this paper tweets are collected about Steve Jobs – post his sad demise to find out what customers think of Apple Inc., now using text mining technique in SAS and R.

Objectives

•To apply text mining technique and analyze tweets regarding Steve Jobs.

•To identify the top of mind recall terms related to Steve Jobs using Concept Link Diagram.

•To compare the results of text mining technique executed on SAS® and R.

Benefits of Text Mining

Recognizes trends and business opportunities
Text Mining transforms unstructured text into numeric form that surmise the collection. This data then becomes input to full range of predictive and data mining modeling techniques

Increased research efficiency
Ability to extract information automatically cuts down the time spent on ensuring coverage of domain knowledge in the literature review process.

Unlocking hidden information
There is probability that there may be underlying connections between different subtopics which can be understood only with automated analysis

Broader economic and social benefits
There is potential for new radical and incremental innovation with wider economic benefit and including innovative service development.

Output of SAS

•Output of Terms matrix
•Role by frequency matrix
•Frequency by Weight
•Concept Link Diagram

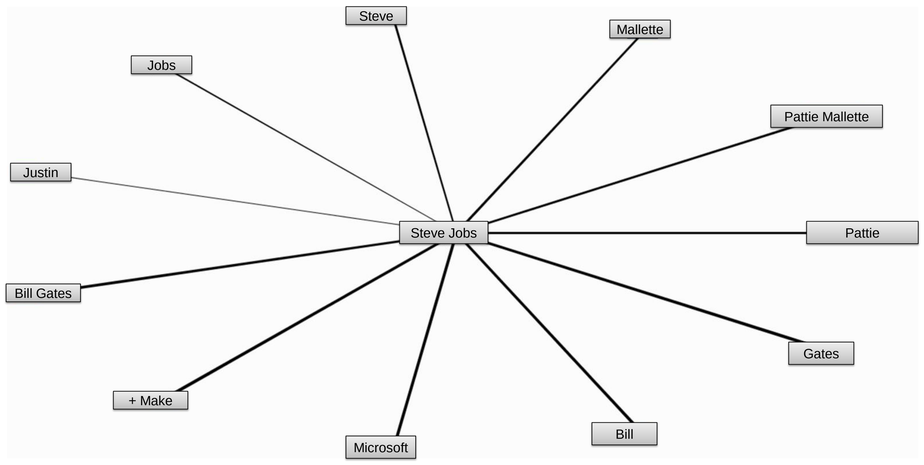
Output of R

•Term Document Matrix
•List of Frequent terms

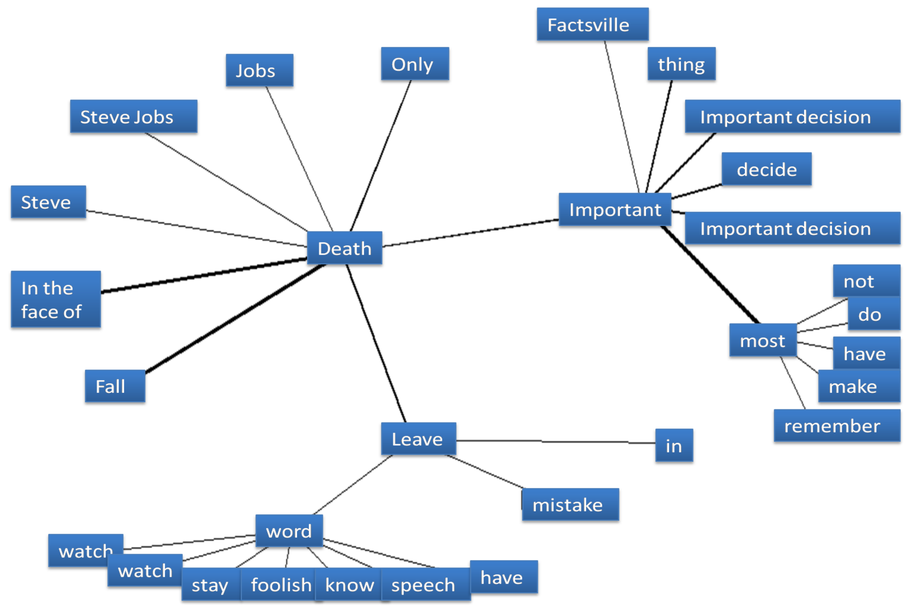
Comparing SAS® and R

Both the software fetches 1500 tweets and fetches same dataset. However, R is limited to displaying only the frequencies of terms whereas SAS® creates Concept Link Diagram showing relationships and strength between terms, which is a step further in the insight obtained. SAS® classifies the terms in different entities whereas there is no scope for classification of terms in R. Apart from classification SAS® also classifies different terms according to their similarity with the document sets.

Results



Concept Link Diagram on term “Steve Jobs” in SAS



Concept Link Diagram on term “Death” in SAS

References

Miner et al, G, 2012,Practical Text Mining and Statistical Analysis for Non Structured Text data application,1st edition,USA, Academic Press.

Weiss et al., S, 2005, Predictive methods for analyzing unstructured information, USA, Springer.

%GetTweet: A New SAS® Macro to Fetch and Summarize Tweets, Satish Garla and Goutam Chakraborty, Oklahoma State University, SAS, 2012.

Kathy Lange and Saratendu Sethi, What are people saying about your company, your products, or your brand?, SAS Institute Inc.

The Value and Benefit of Text Mining, Diane Mcdonald and Ursula Kelly, JISC, 2012.

Package ‘twitteR’, Jeff Gentry, February 20, 2012.

Package ‘tm’, Ingo Feinerer, February 15, 2012.