

Intro to Text Mining

Ken Brooks

May 31, 2015

Ted Kwartler / Director of Customer Success at DataRobot

You will learn how modern customer service organizations use data to understand important customer attributes and how R is used for workforce optimization. Topics include real world examples of how R is used in large scale operations to text mine customer notes, broadly understand satisfaction surveys and forecast call arrival patterns. Concepts will be presented in a manner suitable for business professionals but code examples will be shared for those interested in more than high level explanations.

Workflow

- unorganized state > organized and distilled
- corpora > cleaned corpus > analysis
- set working dir
- load libraries
- make custom functions and specify options
- read in data and preprocess
- perform analysis and save

Two approaches

- semantic using syntactic parsing
- bag of words
- Term Document matrix (terms x documents)
- you then want to make it a matrix to handle sparsity
- can do Term freq inverse doc freq
- synonyms: `gsub()` can specify specific one
 - Otherwise you'll need a natural language dictionary
 - can also stem them down to a base, but it makes it difficult to reconstruct intent

VCorpus holds in memory