



Text Analytics

Natural Language Processing
over Sales Call Transcript

Jane Li
Analytics Center of Excellence
Cox Communications
7-29-2020

About me

WORK

Principle Analyst/Lead Analytics Engineer, Cox Communications
Data Architect, Cox Media Group
Data Engineer Manager, The Home Depot
Application DBA, Coco-Cola Enterprise
Product Manager, Wachovia
Sr. Software Engineer, Enterprise Rent-A-Car
Marketing Manager, Lucent Technologies (China)

SKILLSET

Python, R, Spark, big data technology stack
GraphDB, NoSQL, RDBMS
Data visualization
BI
Cloud (AWS, Azure, GCP)
DevOps, CI/CD



SCHOOL

BSc in Economics, Shandong University, China
MBA, the University of Leeds, England
OMSA, Georgia Tech, US (present)

CONNECT

Email Jane.li@cox.com
LinkedIn <https://www.linkedin.com/in/lijuan98>
GitHub <https://github.com/jli-aic>

Research Problem

Is there something in the initial sales call which may contribute to failed self-install vs. successful self-install?

-- last-ditch attempt

How to drive product strategy and enhance customer experience by gaining insights from transcripts data

“What may contribute to the failed Self-Install for new connects?”

All line of businesses touched
900~ data attributes analyzed
Dozen's ML models fitted

- Product
- Equipment
- Work Order
- Network
- Finance
- Sales
- Marketing
- Weather

- Demographic
- Interaction
- House
- Account
- Psychographic
- Sociographic
- Competitor
- 3rd Party

What

A set of techniques, linguistic, statistical and machine learning.

It includes,

- Rules-based
- ✓ Unsupervised ML-based
- Supervised ML-based
- Hybrid

How

To model and structure text data

Why

Business intelligence
Exploratory data analysis (EDA)
Research
Investigation

Where

VOC/CRM/CFM
Competitive intelligence
Social intelligence
Brand management/ reputation/ awareness
Document analytics
Semantic/ cognitive search

May Text Analytics help?

-- “80 percent of business-relevant information originates in unstructured form, primarily text”

Natural Language Processing (NLP)

--“you don’t know what
you don’t know”

What

A component of text analytics

How

Algorithms for linguistic (language-driven) analysis

Supervised or Unsupervised

Why

To process, analyze and understand large amount of text data

Where

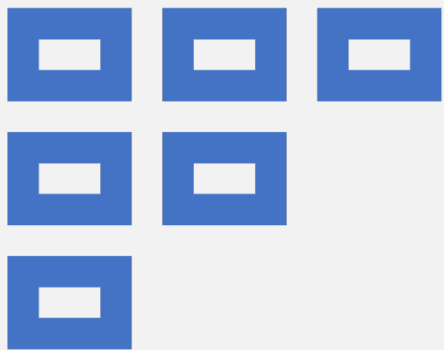
Automatic summarization

Part-of-speech tagging

Disambiguation

Entity extraction and relations extraction

Disambiguation and natural language understanding and recognition



NLP Tools

- ✓ MonkeyLearn | NLP made simple
- Aylien | Leveraging news content with NLP
- IBM Watson | A pioneer AI platform for businesses
- Google Cloud NLP API | Google technology applied to NLP
- Amazon Comprehend | An AWS service to get insights from text
- ✓ **NLTK | The most popular Python library**
- Stanford Core NLP | Stanford's fast and robust toolkit
- ✓ TextBlob | An intuitive interface for NLTK
- SpaCy | Super-fast library for advanced NLP tasks
- ✓ GenSim | State-of-the-art topic modeling



Call



Raw Data



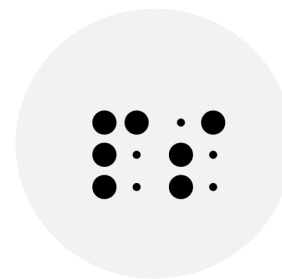
Data Preparation



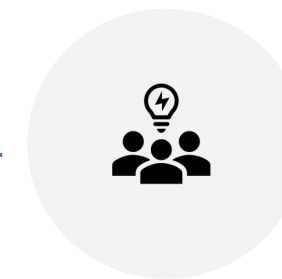
EDA



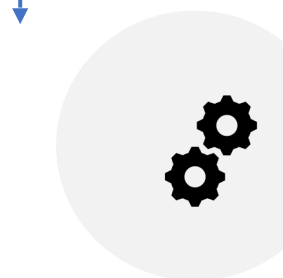
Sentiment Analysis



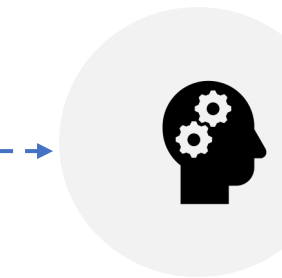
Word Cloud



Top Words



Topic Modeling



Insight Extraction



Validation and Action

Analytics Process Flow

Key Elements

How does it work?

Voice



In- and out-bound calls stored as audio files



Text



Audio data converted to text



Data Preparation



Cleaning
Tokenizing
Matrixing

Python Libraries



Pandas
Scikit-learn
Re
Nltk
Textblob
gensim

NLP Techniques



EDA
Word counts
Word Cloud
Sentiment analysis
Topic modeling

Communication



Interpret topic
Visualization
Extract insights



Python Libraries

Data gathering



Join transcripts per UCID



Extract common words
Extract noun



Standardize format
Remove unnecessary characters



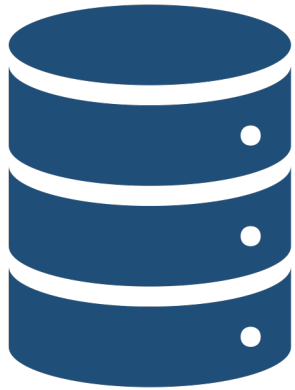
Stemming
Remove stop words



Corpus
Document-Term Matrix

Data Curation

To output clean, organized data in a uniformed format for analysis/analytics



NLP Techniques

EDA

Top words

Word Cloud

Sentiment Analysis

Topic Modeling



Communication

Fact and Observation

Visualization

Insight

Turn into Action or Decision



Notebook

<https://github.com/jli-aic/textanalytics-calltranscript>

Insights



**Diverse
Demographics**



Word Cloud



**Sentiment
Analysis**



**Topics
Modeling**



Opportunity

Less is more

Don't hang yourself on one tree

It may take a whole village

Business knowledge takes more than one third of the pie

Ready for another language?

Fail fast

Requirement deserves a write up and review!

Additional Notes

Sky is unlimited ...



Helpful Resources

Natural Language Processing | Dan Jurafsky, Christopher Manning

https://www.youtube.com/watch?v=3Dt_yh1mf_U&list=PLQiyVNMpDLKnZYBTUOISI9mi9wAErFtFm

www.towardsdatascience.com

www.stackoverflow.com

www.datacamp.com



QA session





Thank You



Text Analytics

Natural Language Processing
over Sales Call Transcript

Jane Li
Analytics Center of Excellence
Cox Communications
7-29-2020

About me

WORK

Principle Analyst/Lead Analytics Engineer, Cox Communications
Data Architect, Cox Media Group
Data Engineer Manager, The Home Depot
Application DBA, Coco-Cola Enterprise
Product Manager, Wachovia
Sr. Software Engineer, Enterprise Rent-A-Car
Marketing Manager, Lucent Technologies (China)

SKILLSET

Python, R, Spark, big data technology stack
GraphDB, NoSQL, RDBMS
Data visualization
BI
Cloud (AWS, Azure, GCP)
DevOps, CI/CD



SCHOOL

BSc in Economics, Shandong University, China
MBA, the University of Leeds, England
OMSA, Georgia Tech, US (present)

CONNECT

Email Jane.li@cox.com
LinkedIn <https://www.linkedin.com/in/lijuan98>
GitHub <https://github.com/jli-aic>

Research Problem

Is there something in the initial sales call which may contribute to failed self-install vs. successful self-install?

-- last-ditch attempt

How to drive product strategy and enhance customer experience by gaining insights from transcripts data

“What may contribute to the failed Self-Install for new connects?”

All line of businesses touched
900~ data attributes analyzed
Dozen's ML models fitted

- Product
- Equipment
- Work Order
- Network
- Finance
- Sales
- Marketing
- Weather

- Demographic
- Interaction
- House
- Account
- Psychographic
- Sociographic
- Competitor
- 3rd Party

What

A set of techniques, linguistic, statistical and machine learning.

It includes,

- Rules-based
- ✓ Unsupervised ML-based
- Supervised ML-based
- Hybrid

How

To model and structure text data

Why

Business intelligence
Exploratory data analysis (EDA)
Research
Investigation

Where

VOC/CRM/CFM
Competitive intelligence
Social intelligence
Brand management/ reputation/ awareness
Document analytics
Semantic/ cognitive search

May Text Analytics help?

-- “80 percent of business-relevant information originates in unstructured form, primarily text”

Natural Language Processing (NLP)

--“you don’t know what
you don’t know”

What

A component of text analytics

How

Algorithms for linguistic (language-driven) analysis

Supervised or Unsupervised

Why

To process, analyze and understand large amount of text data

Where

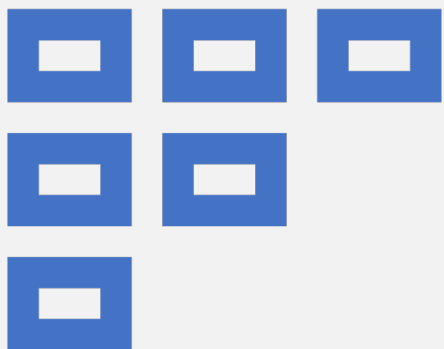
Automatic summarization

Part-of-speech tagging

Disambiguation

Entity extraction and relations extraction

Disambiguation and natural language understanding and recognition



NLP Tools

- ✓ MonkeyLearn | NLP made simple
- Aylien | Leveraging news content with NLP
- IBM Watson | A pioneer AI platform for businesses
- Google Cloud NLP API | Google technology applied to NLP
- Amazon Comprehend | An AWS service to get insights from text
- ✓ **NLTK | The most popular Python library**
- Stanford Core NLP | Stanford's fast and robust toolkit
- ✓ TextBlob | An intuitive interface for NLTK
- SpaCy | Super-fast library for advanced NLP tasks
- ✓ GenSim | State-of-the-art topic modeling



Call



Raw Data



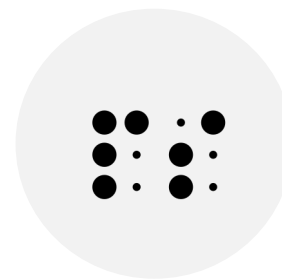
Data Preparation



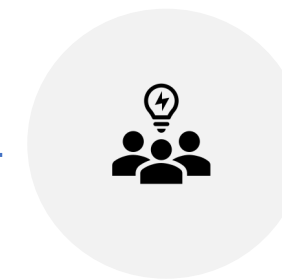
EDA



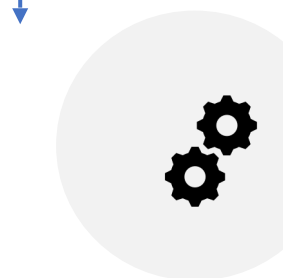
Sentiment Analysis



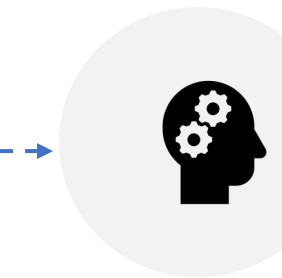
Word Cloud



Top Words



Topic Modeling



Insight Extraction



Validation and Action

Analytics Process Flow

Key Elements

How does it work?

Voice



In- and out-bound calls stored as audio files



Text



Audio data converted to text



Data Preparation



Cleaning
Tokenizing
Matrixing

Python Libraries



Pandas
Scikit-learn
Re
Nltk
Textblob
gensim

NLP Techniques



EDA
Word counts
Word Cloud
Sentiment analysis
Topic modeling

Communication



Interpret topic
Visualization
Extract insights



Python Libraries

Data gathering



Join transcripts per UCID



Extract common words
Extract noun



Standardize format
Remove unnecessary characters



Stemming
Remove stop words



Corpus
Document-Term Matrix

Data Curation

To output clean, organized data in a uniformed format for analysis/analytics



NLP Techniques

EDA

Top words

Word Cloud

Sentiment Analysis

Topic Modeling



Communication

Fact and Observation

Visualization

Insight

Turn into Action or Decision



Notebook

<https://github.com/jli-aic/textanalytics-calltranscript>

Insights



**Diverse
Demographics**



Word Cloud



**Sentiment
Analysis**



**Topics
Modeling**



Opportunity

Less is more

Don't hang yourself on one tree

It may take a whole village

Business knowledge takes more than one third of the pie

Ready for another language?

Fail fast

Requirement deserves a write up and review!

Additional Notes

Sky is unlimited ...



Helpful Resources

Natural Language Processing | Dan Jurafsky, Christopher Manning

https://www.youtube.com/watch?v=3Dt_yh1mf_U&list=PLQiyVNMpDLKnZYBTUOISI9mi9wAErFtFm

www.towardsdatascience.com

www.stackoverflow.com

www.datacamp.com



QA session





Thank You