MACHINE LEARNING OPPORTUNITIES IN FREIGHT TRANSPORTATION OPERATIONS

NORTHWESTERN NUTC/CCIT
October 26, 2016

Ted Gifford
SCHNEIDER IS A TRANSPORTATION AND LOGISTICS LEADER WITH A BROAD PORTFOLIO OF SERVICES.
MACHINE LEARNING APPLICATIONS

• Pricing and Cost Estimation – Spot & Contract
• Forecasting Capacity – IC Drivers
• Driver Turnover
• Driver Safety
• Market Strength Forecasting

TEXT ANALYTICS

• Capturing Lost Sales - Demand Forecasting
• Customer Relations – Assessing Account Health
OPERATIONS CENTER
TEXT ANALYTICS

TURNDOWN CAPTURE

- 10,000+ Tendered Orders/day
- Turndowns (Lost Sales) vary 5% to 20%
- Accurate Demand Forecasting requires understanding of total demand
- Difficult to capture via standard data entry

CUSTOMER RELATIONS

- ~ 500 Customer Service Reps
- Gauge Customer Sentiment
- Improve CSR communication skills
- Pro-actively address issues that lead to churn or freight reduction
- Coaching and discover of systemic issues.
SOLUTION APPROACH (COMMON)

- 60K Messages/day – email, chat, phone
- Collect/Store - Hadoop/MapR
- Preprocessing
  - Parsing – Regular Expressions
  - Tokenizing – NLTK (Python)
  - Feature Extraction via TF-IDF
- Classification – ensemble using
  - Logistic Regression
  - Stochastic Gradient Descent
  - Multinomial Naïve Bayes
  - Support Vector Machine
  - Confidence score voting to determine message topic
DISTRIBUTION BY MESSAGE TOPIC
TURN DOWN ANALYTICS PROCESS FLOW

Data Acquisition
- SQL Server Chat Data (Sqoop)
- Microsoft Outlook Email Data (Java)

Architected Layer
- Data Formating and Filter
  - Hive Queries
- Sentiment Java Stanford Sentiment Libraries
- Structured Content Java Stanford NLP Libraries

Machine Learning
- Email / Chat CR Feature Extraction Regex Tagging ML Classifier
- Email / Chat TD Feature Extraction Regex Tagging ML Classifier

Python and Scikit-Learn Libraries
- Hive Architected Layer
  - Architected Tables
    - xoexml
    - xochat
    - email_attachments
  - Sentiment and Structured Content
    - comctn_4ntmntscr_trndwn_str_cntnt
- Machine Learning
  - chat_cap_rqs
  - chat_tdn
  - eml_caprqst
  - eml_tdn

Structured Content Extraction
- Match to COM
  - Oracle PL/SQL Procedure
  - ODS
    - xo_s_order_dcsn_trndwn

Search and Exploration
Enteza Visualization Application
Cover Data Ingestion Process
TURN DOWN CAPTURE

• Select messages associated with capacity request
• Use Content Extraction to create structured records
• Identify specific entities that define an order... customer, geography, dates and times
• Uses Stanford NLP libraries as well as other pattern matching, and data exclusion techniques.
• Data is cleansed, normalized, matched to normalized locations, dates, and customers in Enterprise Master Data
• Checked against other sources to prevent duplication.
• Posted to Order System via Sqoop and ODI.
SENTIMENT ANALYSIS

• Use Sentence-based analysis

• Recursive Neural Tensor Networks (Socher et al.)
• Stanford NLP Library – Deep Learning
• Sentiment Treebank includes sentiment labels for 215,154 phrases in parse trees of 11,855 sentences
SENTIMENT ANALYSIS

- Sentiment packages trained and built with social media and B2C communication do not work for B2B.
- In B2B communication expression of dissatisfaction or issues of concern is almost always ‘emotionally neutral’.
- Using a business-specific set of reference phrases, the Stanford Treebank is about 50% accurate.
- Underlying technology and algorithms (RNTN) is valid, but needs to be trained with business-specific and industry-specific phrases.
- Re-training with a set of 2200 business sentences, we have raised accuracy to 71%.
- Current plan is increased the training set until 80% accuracy is achieved.
SENTIMENT ANALYSIS

**Positive:**
We are good to go on this load!!

**Negative:**
It’s no longer acceptable to miss appointments without communication.

This order is loaded, and we still haven’t discovered why we are creating another service failure.

**Negative situation:**
Due to the weather advisory in the Gulf Region we are asking that you notify us of any potential weather delays and provide updates of managed loads that will be affected.

The driver is updating a late ETD 8/9 1630 for delivery due to equipment related issues.
THANK YOU

giffordt@schnieder.com