2016-07-04-DataVarietyInCapturingCustomerSentiment

UC Berkeley – MIDS 201 Research Design and Applications for Data Analysis

Week 8 Assignment: Case Study

### Domain, Organization and Stakeholders

In March 2016, at FutureCars Inc., an effort called “Recognize and Resolve (RnR)” was initiated for integrating text, voice, transcription, language and emotion and other data into the Customer Issue Tracking System (CITS) to improve customer satisfaction through pre-emptive recognition of customers who are expressing in some form a negative product satisfaction sentiment. The goal is to capture customer sentiment during the early days of the interaction and take corrective/personalized action well before opinions are expressed on public forums like Twitter and Yelp, at which point customer sentiment recovery is likely too late. The effort for implementation of RnR-CITS is projected at $6.8M. Prior to approval, a pilot program is to be run.

This case study document is a summary of the research study conducted as part of the pilot program to understand the effectiveness of the new RnR enhancements. The immediate business impact is to FutureCars Inc. and the decision as to whether or not to budget the requested monies, with an indirect positive impact projected on FutureCars customers.

### Introduction

Customers interact with their product providers in a multitude of ways and at FutureCars Inc. detailed records of every exchange are kept. The initial contact is usually through the phone or an Internet chat session. For persistent issues, two-way emails, data exchanges, phone calls and chats are normal follow up mechanisms. In each such exchange, the proceedings are catalogued and stored, usually on a diverse set of systems. Furthermore, the data takes multiple unstructured formats such as with messages, text, voice and emails, often times in the language that is local to FutureCars Inc.’s global presence. Because of this diverse segmentation, the data pertaining to an issue cannot be quickly accessed and analyzed in its entirety and automated cognitive methods are critically important [1] [2] [3] and FutureCars Inc. recognizes this.

In recent years, natural language processing and machine learning have made enormous strides [4]. These technologies have the potential to reveal cognitive insights from large amounts of unstructured data. These capabilities are now individually available as cloud based managed services and are capable of extracting intelligence from information databases that hold phone calls, catalogs, training manuals, product disclosures, terms and conditions, emails, customer forums, documentation, internet communities and call center logs. The capability to transcribe voice to text for multiple languages, through the use of linguistic analysis and the capability to recognize tone in voice have also taken shape [5] [6]. Capability to recognize three types of tones that can spot emotion, social tendencies, and language style [5] [7] and personality traits [8] have also been developed.

FutureCars Inc. is contracted with a major managed service provider and is using a cognitive ecosystem implement the RnR-CITS architecture for completion of the pilot program.

### RnR-CITS System Architecture and Deployment Overview

The architecture and the deployment of the proposed RnR-CITS system enhancements is shown in Figure 1. The architecture integrates customer and issue specific information from a vast variety of non-traditional data entry points and assimilates information of different formats.

Customer Phone Records

Customer case data Records

Case E-mails, Messenger notes

Voice

transcription

Emotion

Extraction

Text and Tone Records

Alerts

Foreign Language

Voice transcription

Foreign Language

transcription

Internet

Yelp, Twitter, Communities

FUTURECARS Inc. domain

FUTURECARS Inc.

domain

Managed Service Provider domain

Figure 1. RnR-CITS System and Deployment Overview[[1]](#footnote-1)

### Variety of Data in RnR-CITS Pilot

The FutureCars Inc. pilot is not volume or velocity intense but relies upon a variety of structured and unstructured data such as with Text, Voice, Messages, Internet, Languages, Transcriptions, and Translations.

### Data Processing Complexity and Pre-hypothesis data

The RnR-CITS deployment in Figure 1 shows the sequencing of various cognitive modules and the various points of data transformation and translation. Transcription and translation processing of language, voice and video bring very unusual issues to data science in RnR-CITS.

The transformation and translation algorithms for these varied data types tend to initially be quite lossy. That is, while much information is gained via processing, some of the original information is also migrated to a new form partly because the data source is not perfectly comprehensible and also, some information is lost in the transformations. Lossy algorithms are inherently refined over time, through adaptive learning and dictionary refinement. For a well-known language like English, which has the best support, confidence levels in the 30 – 50% through such multi stage lossy processing is not unusual as shown in Table 1.

The implementation is more lossy (< 30% confidence), as compared to English, with other languages and accents.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Language**  **(a)** | **Current Dictionary capability, % of language**  **(b)** | **Current Voice transcription capability,**  **Intelligible voice**  **(d)** | **Current Emotion recognition capability,**  **Qualitative metric**  **(e)** | **Theoretical cumulative worst case metric:**  **(b) \* (d) \* (e) =**  **(f)** | **Theoretical cumulative best case metric:**  **(b) \* (e) =**  **(g)** | **Using a more realistic 65% expectation, the true expected best case metric: is:**  **(f) \* 0.65 = (h)** |
| English | 85% | 80% | Decent - 60% | 0.408 | 0.510 | **0.332** |

Table 1. Sequential points of confidence loss in RnR Cognitive Data Processing

### Implementation baseline via Training Data Set

The pilot implementation per the structure in Figure 1was used to study historical English-centric cases with customer issues that included a comprehensive set of voice, data and message records. Using 40 well-known cases as a training set, the pilot implementation was tested and a validation level of between the Practical expectation of 33% (Table 1, column h) and the Theoretical best of 51% (Table 1, column g) was confirmed in the 35 – 38% range.

### Core Research Problem

The core objective is to qualify the viability of the RnR-CITS enhancements in the pilot program at a minimum usability level of 33% using real (English centric) customer data

### Null Hypothesis

The RnR-CITS enhancements pilot **can** accurately predict alerts of future customer issues at an accuracy level of at least 33%.

### Alternative Hypothesis

The RnR-CITS enhancements pilot **cannot** accurately predict alerts of future customer issues at an accuracy level of at least 33%.

### Organizational Decision

By accepting or rejecting the Null hypothesis, we establish the viability of the RnR-CITS enhancements and hence make the decision whether or not to go to a business justification study prior to funding the full follow-on implementation program.

### How is the decision made?

The test is run over 12 weeks. The base customer case data is available via SQL databases. The voice recordings were kept in the cloud. This data is fed into the translation and transcription engines via well defined REST API and JSON API. The text and voice transcription software were uploaded with the latest libraries that took raw data as input and returned translated data or customer sentiment as output. A FutureCars Inc. defined threshold is then used to set the alert levels. A human operator then intercepted this alert level when the flag indicated that a customer needed attention. By having a personal conversation with the customer, the operator then determined the legitimacy level of the alert. This data is then used to compute the accuracy of the RnR-CITS system.

Multiple test paths are chosen for testing:

1. Scenario #1: Tone recognition in customer phone records without transcription to text
2. Scenario #2: Tone recognition in customer phone records with transcription to text
3. Scenario #3: Tone recognition in customer messenger notes and emails

If a tested path is successful (shows an accuracy level of at least 33%), the hypothesis is deemed successful in that path.

### Likely Outcome and Actual Results Summary

Given that a testing with a training set preceded the hypothesis testing, it was expected that the Null hypothesis would be confirmed. Each module in the deployment sequence has well understood behavior and well understood limitations and so the overall outcome is predictable.

Testing of the first 60 customer test cases confirmed the likely outcome and confirmed the null hypothesis.

### Next Steps

Since the null hypothesis is true, a business hypothesis will next be formulated and tested. That is, given that RnR-CITS can deliver a minimum of 33% accuracy, is the business justified at that level of accuracy? This can be treated as the business hypothesis that must be tested next. If this business null hypothesis is true, funding of $6.8M will be granted. If this business null hypothesis rejected, the program will be stopped.

If the business is justifiable for the current implementation, the next technical step is to perfect the RnR-CITS architecture to push results to a minimum of 65% level of confidence. This will then be followed by a determination of business viability at the new level of effectiveness. The plan is to accomplish this updated objective in adaptive and dynamic fashion, with the use of translation dictionaries that are constantly tuned and accompanied by updates to the transcription, translation and emotion recognition modules.

### Overcoming biases and fallacies

1. Engineering design is often times marred by Perfection fallacy. That is do something if it works all the way of don’t do it. However, given that a business can at times afford to take incremental steps towards perfection, hypotheses that drive towards 33% acceptability, if the business is viable at this confidence level, can be important stepping-stones to larger confidence targets. FutureCars Inc. executives have questioned this approach and have had to be re-educated. In this pilot, this fallacy was overcome via the initial training set based test that set the expectations for the pilot.
2. Given the complex processing paths in RnR-CITS, there is the possibility of Overfitting and Underfitting bias if the model is trained with too many data sets. To overcome this bias, the training of the model will be stopped once the pilot is successfully completed.

### Cautious Road Ahead for FutureCars Inc.

Getting to a very robust cognitive solution with the complexity of data will take iterative progress. The current process of verifying work through a pilot program protects FutureCars investment. FutureCars executives will also be educated about the “Achieve perfection with pilot results” fallacy.

### References

1. <http://www.cnbc.com/2015/10/06/ibms-rometty-businesses-are-in-a-cognitive-era.html>
2. <https://www.ibm.com/blogs/watson/2016/04/innovative-organizations-using-cognitive-technology-disrupt-markets/>
3. <http://www.techrepublic.com/article/how-ibms-cloud-and-cognitive-computing-tools-help-deliver-personalized-customer-experiences/>
4. <http://www.ibm.com/watson/what-is-watson.html>
5. <http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/tone-analyzer.html>
6. <http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/speech-to-text.html>
7. <https://en.wikipedia.org/wiki/Emotion>
8. <https://en.wikipedia.org/wiki/Big_Five_personality_traits>

1. Foreign language voice transcription and text transcription are shown with dotted lines because they are available but are not a part of the current pilot [↑](#footnote-ref-1)