CallAn: A Tool to Analyze Call Center Conversations

Balamurali AR, Frédéric Béchet And Benoit Favre *

Abstract Agent Quality Monitoring (QM) of customer calls is critical for call center companies. We present CallAn, a tool to analyze the call center conversations based on different conversation parameters. CallAn presents the users with timeline analysis of the conversation along with different meta and voice parameters in the form of an analytic dashboard. The user can compare the values with average values of these parameters. Objectivity in evaluation using CallAn can reduce the supervisor bias and also reduce the effort for quality monitoring.

1 Introduction

Quality Monitoring (QM) of call conversation is a critical task in a typical call center (Rafaeli et al., 2008; Witt et al., 2004). It assess whether agents who is attending the call is ensuring the quality standards prescribed by the agency. This standard can be behavioral and procedural in nature. Another reason for such monitoring is to take vibe of the customers toward the standard procedural operations and the manner in which it is performed.

To ensure the quality of the call on various dimension, these centers follow QM questionnaires. These questionnaire evaluate different behavioral and procedural aspects of the conversation. The quality is evaluated by QM supervisors based on the parameters mentioned on the QM questionnaire. Ideally all the conversations are to be monitored. However, due to staffing and other logistical issues, not every calls are monitored. Typically, a call center gets 200-4500 call per day. It is difficult to monitor all the calls. Usually, 2 to 10 calls per agent per month are monitored (Kop-

Balamurali AR, Frédéric Béchet,Benoit Favre
Aix-Marseille Université, CNRS, LIF UMR 7279, 13000, Marseille, France, e-mail: balamurali.ar,frederic.bechet,benoit.favre@lif.univ-mrs.fr

* The research leading to these results has received funding from the European Union - Seventh Framework Program (FP7/2007-2013) under grant agreement n610916 SENSEI

parapu, 2015). Evidently, manual process of monitoring all the calls is cumbersome process. Apart from this, there is a considerable bias among the QM supervisors. This can lead to high disagreement on the assessment.

Studies exists which deals with *queuing*, *staffing* and *prediction* problems in a callcenter (Soyer and Tarimcilar, 2008; Taylor, 2008; Whitt, 1999, 2006). Apart from these dimensions, other research focus on improving technological and social environment of the call centersDeslauriers et al. (2007); Taylor (2008); Weinberg et al. (2007). In this system paper, we present a Call Analyser (CallAn) *tool to analyze call center conversations*. Given there are many calls per day and there exists bias among the QM supervisors, a flagging based system is devised. The conversations which are flagged are then presented to QM supervisors using CallAn. Using CallAn, QM supervisor can analyze the conversation based on different parameters with respect to the average agent response. CallAn also provides features like sentiment analysis on the conversation thread along with the audio description for each turn. Additionally it also act as an annotation bench for new QM parameters. This can aid them in concluding quality of the call with less bias. Through this approach, every call can be sampled, thus improving the overall quality of response.

2 Corpus Description

This study has been done on the RATP-Decoda (French) conversation corpus Bechet et al. (2012). The corpus focuses on conversations between agents and customers of a Paris public transport authority call center. Calls contain various queries pertaining to *lost objects*, *route planning*, *traffic information*, *complaints*, etc. This conversation corpus contains structural and semantic annotations extracted as a part of previous research. For the objective of this work, QM related annotations have been added through the EU project SENSEI.

To build the classification models used in this prototype, the Decoda conversation corpus has been annotated with the QA form mentioned in Table 2 by 5 annotators. The audio and respective transcript was provided to the annotators. The annotators are professional quality monitoring supervisors in one of largest call center company in Europe (*Teleperformance*), partner of the SENSEI project. Based on the specific questions mentioned in table 2, they had to mark the conversation as *PASS*, *FAIL* and *NA*. The category PASS reflect that annotator is satisfied with specific objective mentioned in the QM questionnaire. If they are unsatisfactory, then they are marked as FAIL. If the annotators do not have sufficient information to make decision they are marked as NA. This includes cases in which the service does not provide any actions of up-selling, or if agents do not collect specific information of the customer like name, surname etc, or if the agent's objective (qualitative or quantitative) is ambiguous.

ID	Quality Monitoring Parameters
1	Agent respects opening procedure
2	Agent listens actively and asks relevant questions
3	Agent shows the information in a clear, comprehensive and essential way
4	Agent manages the objections reassuring the customer and always focusing on client satisfaction
5	Agent manages the call with safety
6	Agent uses positive words
7	Agent follows the closing script
8	Agent is polite and proactive with the customer
9	Agent is able to adapt to the style of client's communication always maintaining professionalism
10	Agent Management: he negotiates the wait always giving reasons
11	Ability to listen

Table 1 Quality Monitoring Parameters Evaluated

3 Conversation Analysis

Conversation style followed in call centers are generative dialogues (Gergen et al., 2004). They focus on solving a problem in a constructive manner. Most often the customer is the lead and pose questions and queries for the agent. The conversation follows a threaded style with each party taking turns to respond to specific queries or reinforce the solution.

3.1 Process Flow

CallAn interface provide QM supervisor to closely observe the agent conversation which is flagged by the automatic system for any quality related relapses. Flagging based system uses text and audio related features to flag conversation as fail based on any of the parameters mentioned in Table 2 (Balamurali AR and Riccardi, 2016). This system has an accuracy of 91%. Individual classification model for each QM parameter is not possible due to low annotator agreement. The process flow diagram is shown in the Figure 1. The pass conversations are directly stored into the call center data center whereas the fail conversations, which are flagged by the system, are presented to QM supervisors through CallAn.

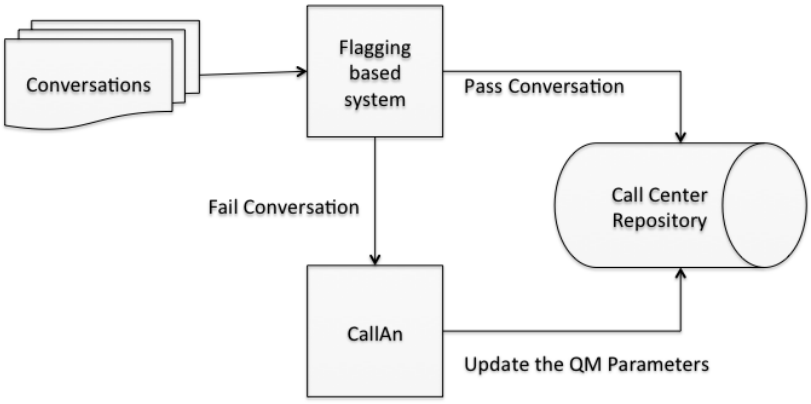

Fig. 1 Process Flow diagram and CallAn interface

3.2 CallAn parameters for user analysis

CallAn provides an extendable analytic dashboard presenting all the parameters that have been analyzed by the flagging system.

- Emotion Parameter: We developed a sentiment lexicon to extract sentiment-related features. An emotion score is calculated based on the normalized sentiment bearing words. The user is also shown the average emotion score over all the conversations that were used to train the flagging based system.
- Conversation Time Parameter: The time spoken by each participant during the call can be a measure of the agent behavior. The conversation time of the agent, customer and the automatic machine is captured using the audio file. The normalized conversation time for each participant is used as features. The average conversation time is also shown.
- Conversation Length Features: The conversation length in terms of words spoken by the agent, customer and the automatic machine is captured using audio file. The normalized conversation length for each participant is presented along with the average values.
- Wait Time Parameter: The waiting time refers to time taken by agent to respond to a query. It can also refer to the agent as well as the customer. Average normalized wait period is extracted for the agent as well as customer. This along with

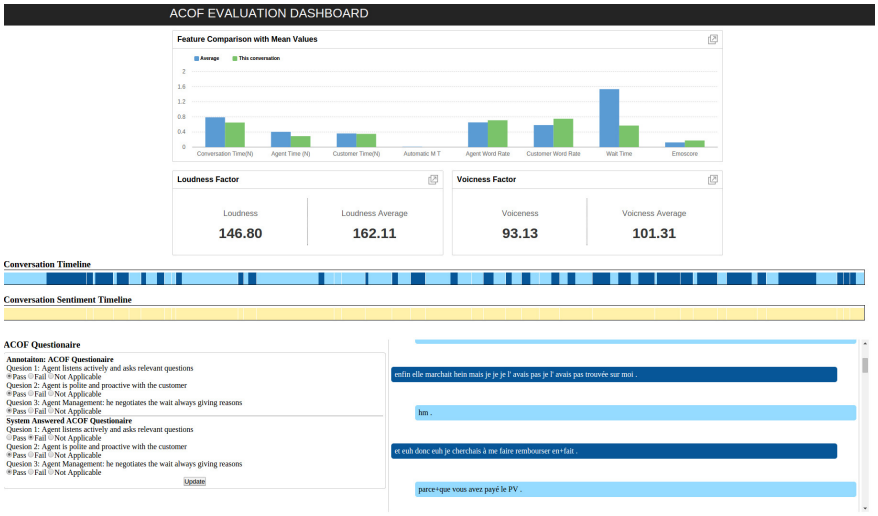

Fig. 2 CallAn User Interface

the average waiting time over the complete set of conversations in the training set is also presented.

- Speech Parameters: The fundamental frequency, voicing probability and the loudness contour of the conversation along with the average values is presented for the user to analyze the specific conversation.
- Conversation Time-line: The entire conversation is broken as per the turns. The user is provided with the facility to click each turn. This would in turn activate the audio part for that turn along with highlighting of the part in the conversation thread.
- Sentiment Time-line: Apart from the conversation Time line, a sentiment time line of the conversation is provided. Using this time line, the users can see the problematic areas of the conversation as detected by a sentiment analysis engine.

The demo video of CallAn can be found at `https://youtu.be/cauXQoyKWMQ`

4 Conclusion

In this system paper, we presented CallAn: a call analyzing tool for quality monitoring of call center conversations. CallAn presents the users with an extendable dashboard populated with information about different parameters of the conversation. Using this tool, we hope to mitigate the effort required for quality monitoring by the supervisors and also help in having a better understanding about the agent behavior towards different kind of conversations.

References

M. Danieli B. Favre F. Béchet Balamurali AR, E. A. Stepanov and Giuseppe Riccardi. A flagging based approach for automatic quality monitoring of call center conversations (under review). In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference*. IEEE, 2016.

Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, and Eric Arbillot. Decoda: a call-centre human-human spoken conversation corpus. In *LREC*, pages 1343–1347, 2012.

Alexandre Deslauriers, Pierre LEcuyer, Juta Pichitlamken, Armann Ingolfsson, and Athanassios N Avramidis. Markov chain models of a telephone call center with call blending. *Computers & Operations Research*, 34(6):1616–1645, 2007.

Mary M Gergen, Kenneth J Gergen, and Frank Barrett. Appreciative inquiry as dialogue: Generative and transformative. *Advances in appreciative inquiry*, 1(1): 3–27, 2004.

Sunil Kumar Kopparapu. *Non-Linguistic Analysis of Call Center Conversations*. Springer, 2015.

Anat Rafaeli, Lital Ziklik, and Lorna Doucet. The impact of call center employees' customer orientation behaviors on service quality. *Journal of Service Research*, 10(3):239–255, 2008.

Refik Soyer and M Murat Tarimcilar. Modeling and analysis of call center arrival data: A bayesian approach. *Management Science*, 54(2):266–278, 2008.

James W Taylor. A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science*, 54(2):253–265, 2008.

Jonathan Weinberg, Lawrence D Brown, and Jonathan R Stroud. Bayesian forecasting of an inhomogeneous poisson process with applications to call center data. *Journal of the American Statistical Association*, 102(480):1185–1198, 2007.

Ward Whitt. Dynamic staffing in a telephone call center aiming to immediately answer all calls. *Operations Research Letters*, 24(5):205–212, 1999.

Ward Whitt. Staffing a call center with uncertain arrival rate and absenteeism. *Production and Operations Management*, 15(1):88, 2006.

L Alan Witt, Martha C Andrews, and Dawn S Carlson. When conscientiousness isnt enough: Emotional exhaustion and performance among call center customer service representatives. *Journal of Management*, 30(1):149–160, 2004.