January 17, 2007

RT0702
Knowledge Management 8 pages

Research Report

A Conversation-Mining System for Gathering Insights to Improve Agent Productivity

Hironori Takeuchi , L Venkata Subramaniam, Tetsuya Nasukawa, Shourya Roy and Sreeram Balakrishnan

IBM Research, Tokyo Research Laboratory
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

A Conversation-Mining System for Gathering Insights to Improve Agent Productivity

Hironori Takeuchi*, L Venkata Subramaniam†, Tetsuya Nasukawa*
Shourya Roy† and Sreeram Balakrishnan†

*IBM Research, Tokyo Research Laboratory
Shimotsuruma 1623-14, Yamato-shi
Kanagawa 2428502 Japan
{hironori, nasukawa}@jp.ibm.com

†IBM Research, India Research Laboratory
IIT Delhi, Block-1
New Delhi 110016 India
{lvsubram, rshourya, srbalakr}@in.ibm.com

Abstract

In this paper we look at the methods for analyzing transcriptions of recorded calls of customer-agent interactions in a contact center to see how agent productivity can be improved. The aim is to obtain actionable insights to improve agent performance by automatically analyzing such transcripts taken from a car rental help desk. These were analyzed to discover steps that agents are taking to convince customers to make bookings and pick up later. Customers book cars from a particular vendor if they are satisfied on various parameters such as rates, car models, pickup locations, etc. In particular we aim to discover specific traits of agents that result in car bookings and pickups. Based on the analysis it is shown that specific actions by the agents result in better pickup. After implementing such actionables, over a period of one month the booked car pick-up rate improved by 1.72% after adjustment for seasonal effect. We also propose an automated technique to identify key segments of customer agent interactions and using these segments we demonstrated an effective way of identifying cases lacking compliance to prescribed guidelines.

1 Introduction

"Contact center" is a general term for customer service centers, help desks and information phone lines. Many companies today operate contact centers to sell their products, handle customer issues, and address product-related and services-related issues. Such support includes dialog-based (voice or online chat) and email support that a user receives from a professional agent. Gigabytes of data are produced every day in the form of speech audio recordings, transcripts, call records, emails, etc. This data is valuable for analysis from many aspects. For example, this data can be analyzed to identify how customer behavior is changing by tracking changing patterns in agent-customer interactions or to looked for shortcomings of individual agents by correlating the output of interactions with the number of customer complaints handled. In this paper we are primarily talking about *inbound* contact centers where customers call for service rather than *outbound* contact centers where agents from contact centers call to offer services.

Contact centers employ a handful of people, generally known as *Quality Analysts* (QAs), to analyze the agent-customer interactions. These analysts randomly sample calls and listen to the conversations between agents and customers. As a result, only a very small (1-2%) fraction of the total call volume is sampled. The analysts rate agents based on these calls on various metrics such as *communication skills*, *problem identification*, *problem resolution*, etc. The analysts also categorize calls into different categories based on the nature of the problems or queries. Examples of domain-independent categories could be *resolution issue*, *delayed response*, *repeated calls* whereas categories in a specific domain such as technical support could be *installation issue*, *network outage*, etc. This kind of manual analysis suffers from some serious problems. Most important among them being *lack of consistency* or *subjectivity* between different QAs. Quite often one analyst may mark a conversation as showing an agent lacking proper communication skills while another analyst sees the same data as showing a resolution problem. Secondly only a small fraction of the calls can be analyzed manually, which introduces significant sampling bias in the analysis. Also, the manual analysis process is obviously slow - they need to listen to multiple calls as long as 30 minutes, depending on the nature of the business. Last but definitely not the least, we have observed the manual analysis process can be quite shallow in nature and more often than not the analysis results in reports showing the *whats* but not the *whys*. In this paper we propose techniques using which we could quickly identify business insights that domain experts could not find in spite of they spending days analyzing the same data.

Text Analytics can play a vital role in improving the efficiency of this analysis process as well as doing deeper and more insightful analysis. However application of text analytics is not very common in contact centers because contact centers do not have in-house text analytics capabilities. Also owing to the confidential nature of data, the business information contained in emails, call logs, and call transcriptions can not be made public. Hence such data never reaches the text mining community. We have been interacting services organizations including contact centers and in this paper we will share our experiences of applying text analytics techniques on really "real-life" data.

We looked at methods for analyzing transcriptions of telephone interactions between customers and contact center agents in a car rental transaction. In such a transaction, a customer calls the contact center with the intention of booking a car, making changes to an existing bookings, asking rates etc. It is observed that customers book cars from a particular vendor if they are satisfied on various parameters like *rates*, *car models*, *pickup locations*, etc. After making the booking the customer picks up the car on the mentioned date and from the mentioned location. It is important to note that not all successfully converted calls get eventually picked up and a call adds to revenue only when a car is picked up. Typically in the car rental business about half of the bookings never get picked up. We are trying to automatically answer some of the key questions faced by QAs with the objective of improving agent productivity.

- Is there any association between agent behavior and call conversion/pick-up?
- Can some typical agent actions boost call conversion/pick-up?
- Are agents following preset guidelines during interaction?

- Does following guidelines during the conversations have any impact on call conversion/pick-up?

Organization of the Paper: We start by describing related work in relevant areas. Section 2 also talks about the contact center car rental dataset and the details of the manual transcripts. The following section describes the system for analyzing the calls. Section 5 provides the results both in terms of the actionable insights found by the system and the results of implementing these insights in a contact center. Section 5 concludes the paper.

2 Problem Definition

The aim of this work is to look at transcriptions of telephonic conversations from car rental transactions to determine patterns in the agent dialogs that typically lead to booking and pickup of a car. We also identify the call flow by identifying segmenting calls into various segments and use the segmentation to evaluate compliance adherence of agents.

2.1 Description of the Dataset

We collected 914 calls from the car rental help desk and manually transcribed them. Figure 1 shows part of a call that has been transcribed. There are three types of calls:

```
AGENT: Welcome to CarCompanyA. My name is Albert.  How may I help you?
.........
AGENT: Allright may i know the location you want to pick the car from.
CUSTOMER: Aah ok I need it from SFO.
AGENT: For what date and time.
.........
AGENT: Wonderful so let me see ok mam so we have a 12 or 15 passenger van
available on this location on those dates and for that your estimated total
for those three dates just 300.58$ this is with Taxes with surcharges and
with free unlimited free mileage.
.........
AGENT : alright mam let me recap the dates you want to pick it up from SFO
on 3rd August and drop it off on august 6th in LA alright
CUSTOMER : and one more questions Is it just in states or could you travel
out of states
.........
AGENT : The confirmation number for your booking is 221 384 699.
CUSTOMER : ok ok Thank you
Agent :  Thank you for calling CarCompanyA and you have a great day good
bye
```

Figure 1. Transcript of a Car Rental Dialog (part of a call)

1. **Booked Calls:** Calls which got *converted* i.e. customer made reservation for a car. Booked calls can get *picked-up* or *not picked-up*.
2. **Unbooked Calls:** Calls which did not get converted.
3. **Service Calls:** Customers changing or enquiring about a previous booking.

The distribution of the calls is given in Table 1. These calls were randomly selected and were spread over one hundred agents. About 15 agents account for 1/3 of the total number of calls.

Table 1. Distribution of Calls

<table>
<tr><td colspan="2">Unbooked Calls</td><td colspan="2">461</td></tr>
<tr><td rowspan="2">Booked Calls</td><td>Picked-Up</td><td>72</td><td rowspan="2">137</td></tr>
<tr><td>Not Picked-Up</td><td>65</td></tr>
<tr><td colspan="2">Service Calls</td><td colspan="2">326</td></tr>
<tr><td colspan="2">Total Calls</td><td colspan="2">914</td></tr>
</table>

In a car rental conversation, the following conversation flow is predefined for the agent. These segments are commonly found in reservation calls and for unbooked or service calls, some of the segments might be missing.

- **Opening** - contains welcome, brand name, name of agent
- **Pick-up and return details** - agent asks location, date and time of pick up and return, etc.
- **Offering car and rate** - agent offers a car with its rate and other special offers
- **Personal details** - agent asks customer's information such as first name, last name, address, etc.
- **Confirm specifications** - agent recaps the reservation information such as the name, location, date, etc.
- **Mandatory enquiries** - agent enquires of the customer about clean driving record, valid driving license, major credit card, etc.
- **Closing** - agent gives confirmation number to the customer and thanks the customer for calling along with a mention of the brand.

Because the call process is defined clearly, all reservation calls look similar in spite of having different results (in terms of pickup). Finding the differences in the conversations that affect the outcomes is very important.

2.2 Background and Related Work

Today within a contact center there are a variety of textual information sources. There is data, of the interactions between customers and agents, in the form of online chat transcripts, phone conversation transcripts, emails, and SMS. Within a business there are also other information sources such as business documents, e-mail, news articles, press releases, technical journals, patents, conference proceedings, business contracts, government reports, regulatory filings, discussion groups, problem report databases, sales and support notes, and, of course, the Web. Text mining technologies are used to search, organize and extract value from all these information sources [2][11].

There has been a lot of work on specific tools for contact centers. These include call type classification for the purpose of categorizing calls [12], call routing [6] [5], obtaining call log summaries [3], agent assisting and monitoring [7], building of domain models [10], analyzing records of customer contacts [8].

2.3 Contributions of this paper

In this paper we build on the TAKMI (Text Analysis and Knowledge MIning) system [8] to analyze contact center calls of the interactions between agents and customers. We try to analyze aggregated information from a large set of documents rather than focusing on the specific information in each document. We describe ways to extract the appropriate information from the text and then apply statistical analysis to the extracted information. Finally provide visualizations of the results and allow interactive analysis to meet the requirements of analysts working from multiple points of view. We show in this paper that from the collection of calls we can extract useful information about agents' behavior which can be used for improving their performance. We also show an effective way of segmenting calls to identify the call flow and identify lack of compliance to guidelines by agents.

3 System for Analyzing Transcribed Calls

In this section, we discuss our approach to develop a practical text mining system. The three main components of our system are:

1. Sentence boundary detection and call segmentation
2. Concept extraction based on robust natural language processing
3. Text mining for discovering rules and patterns

3.1 Sentence boundary detection and call segmentation

Call transcription when done manually is a tedious and error prone process. Transcribers make many errors while transcribing such as not inserting punctuation marks, skipping words, spelling mistake, putting incorrect case. Automatic speech transcription system generally does not make spelling mistakes (as words come from language model) but even automatically transcribed text does not contain sentence boundaries. However many applications such as information extraction and natural language processing benefit from (or even require) a sentence structure. We use the partially labeled set from the manual transcriptions to learn and improve the sentence boundaries by identifying phrases which typically appear at the beginning or end of a sentence. The sentence boundaries, though not always perfect, but help in the concept extraction phase that follows.

Calls to a contact center typically follow a predetermined call flow where a sequence of events or actions happen. Example of such events or actions could be agent gathering pickup and return details from customer, customer expressing displeasure about high rates being offered. Hence a call can be broken up into *segments* based on the particular action being performed in that part of the call. For example all calls contain a *greeting* segment and a *conclusion* segment. Depending on the domain they may also contain segments for asking *customer details*, *verification of details*, etc. Typical segments found in a car rental process are shown in section **??**. Like sentence boundary detection, call segmentation helps in concept extraction phase by making identification of concept easier. In many cases correct extraction depends on the context of the concept being extracted. The phrase *credit card* may get mentioned in many parts of a call - a customer asking the agent if payment by a particular credit card is acceptable, the agent asking for credit card details from the customer, the agent verifying credit card details from the customer, etc. An easy way to verify if the agent has verified credit card details is to look for the phrase *credit card* in *confirm specifications* segment.

Call segmentation is done using a supervised technique where a subset of the calls is manually segmented which forms the training set. From this set of manually segmented documents we extract two sets of keywords for each segment:

1. *Frequent keywords* obtained by taking the trigrams and bigrams with the highest frequency in each segment. Unigrams are avoided because most of the high frequency words are stopwords (like, the, is etc).
2. *Discriminating keywords* obtained by taking the ratio of the frequent phrases (unigrams, bigrams and trigrams) in a particular segment to their frequency in the whole corpus with preference being given to trigrams.

Top 10 keywords are chosen as representative keywords for each segment. For new calls, each speaker turn is tagged as belonging to a particular segment based on the keywords present in it and the information about the segment to which the preceding turn belongs. Details of the segmentation technique and results can be found in [1].

3.2 Information extraction

In this section, we will describe how important concepts are extracted from call transcriptions. we use the term "concept" as a representation of the textual content in order to distinguish it from a simple keyword with the surface expression. For the information extraction, transcribed text is sent through *annotators* or *Text Analysis Engines (TAEs)* implemented within the IBM UIMA framework [4]. UIMA (Unstructured Information Management Architecture) is a software architecture in which a user can develop as well as use text analytics components implemented by others and has recently been publicly released[1]. We have developed several document level TAEs within the UIMA framework to extract important concepts from transcribed calls.

3.2.1 Semantic Information Assignment using Domain Dictionary

Assuming that each domain has important terms for analysis, we make a list of words extracted from call transcriptions sorted by their frequency and ask domain experts to assign semantic categories to words that they consider important. The domain experts are also asked to assign appropriate canonical forms to take care of synonymous expressions or variations in the expressions. This dictionary consists of entries with surface representations, parts of speech (PoS), canonical representations, and semantic categories such as the following example for the car rental domain.

- child seat [*noun*] $\longrightarrow$ child seat [*vehicle_feature*]
- NY [*proper noun*] $\longrightarrow$ New York [s*place*]
- master card [*noun*] $\longrightarrow$ credit card [*payment_methods*]

We observed that the number of frequently appearing words is relatively limited in a textual database, especially when the content belongs to a narrow domain. This is due to not very large number of calls and the fact that the word distribution is not balanced The workload for this dictionary creation has been relatively small in our experience. Terms not appearing in the dictionary are automatically mapped to the corresponding canonical form with semantic categories. This dictionary lookup process is applied to the results of POS tagging system [9].

3.2.2 Expression Extraction using Pattern Matching

In natural language, there are many ways to express the same concept. In this section, we will show how concepts can be extracted using user defined patterns of grammatical forms or certain expressions and domain dictionary. User is expected to define some patterns in which he is interested. For example to identify how car rental agents are putting their requests or mentioning value selling phrases, user defined phrases could be as follows.

- please + VERB $\rightarrow$ VERB[*request*]
- just + NUMERIC + dollars $\rightarrow$ mention of good rate[*value selling*]
- wonderful + rate $\rightarrow$ mention of good rate[*value selling*]

This allows us to associate communicative intentions with predicates by analyzing grammatical features and lexical information. For example, for "rude", we can define followings expression patterns.

- X was rude. $\rightarrow$ rude[*complaint*]
- X was not rude. $\rightarrow$ not rude[*commendation*]
- Was X rude? $\rightarrow$ rude[*question*]

The previous dictionary looking up process assigns semantic categories to each word without considering any features around the target word. In this process, we can assign opposite semantic categories such as intention, though it takes more costs to prepare patterns for the expression extraction compared to the domain dictionary construction.

[1] http://www.alphaworks.ibm.com/tech/uima

3.3 Text Mining

Once appropriate concepts have been extracted from the documents, we can apply various statistical analysis methods in data mining to the set of concepts as well as to the structured data. As a result, even a simple function that examines the increase and decrease of occurrences of each concept in a certain period may allow us to analyze trends in the topics. Also, the semantic classification of concepts enables us to analyze the content of the texts from the viewpoints of various semantic categories. In the following subsections, we introduce some analysis functions in TAKMI [8] that are useful for gathering insights.

3.3.1 Relevancy Analysis with Relative Frequency

The basic idea of relative frequency is very simple. It compares the distributions of concepts within a specific data set featured with one or more concepts with the distribution of the concepts in the entire data set. For example, given conversational data from the car rental reservation center, the distribution of the phrases in the ”method of payment” category within a set of data associated with a specific call type may be different from the distribution of that ”method of payment” over the entire data set. Within 42 unbooked calls whose reason for being unbooked is ”not meeting requirements”, a ”debit card” is found in 13 calls (31%), whereas it is found in only 31 calls (3.4%) of the entire 914 calls. In this case, the relative frequency of ”debit card” for the not-meeting-requirements-unbooked records is $9.3(=\frac{13/42}{31/936})$. This indicates that the density of ”debit card” within the not-meeting-requirements-unbooked records is about 9 times higher than for the entire data set. By sorting phrases in the category based on the relative frequencies, relevant concepts for a specific data set are revealed.

3.3.2 Two Dimensional Association Analysis

In order to obtain valuable insights on the stronger relationships among concepts, simple applications of the well-known association rule extraction techniques used in data mining do not work well in text mining. Unlike basket analysis, many of the items, such as the words and phrases in a text, tend to have some dependencies upon each other. For example, they may form a compound word, or they may have other grammatical relationships with each other, such as between a verb and its object. Thus, the application of association rule extraction by considering the text data as a basket full of concepts expressed by words and phrases usually results in a list of item sets that correspond to typical compound words and predicate-argument pairs.

In order to extract the valuable relationships among concepts, it is important to pre-identify what kinds of relationships can be valuable. For example, given a set of car rental conversations, it may be valuable to know what kinds cars get booked from a given location. Then we would need to target expressions that indicate car types such as ”full-size,” and ”SUV” as well as places such as ”New York,” ”Los Angeles,” ”Seattle,” and ”Boston.”

Once we set up such car categories and location names, we need to develop a dictionary and information extraction rules for identifying mentions of each item as described in Section 4.2. For example, ”SUV” may be indicated by ”a seven seater,” and ”full-size” may be indicated by ”Chevy Impala.” As a result, we can fill in each cell in a two-dimensional table as in Table 2 by counting the number of texts that contain both the column and row labels.

However, because of the differences in recall and precision for information extraction for each concept, the absolute numbers may not be reliable. Still, if we can assume that the recall and precision for extracting each concept are coherent over the whole data set, we can calculate indices showing the strengths of the associations for each cell compared to the other associations in the table.

One simple measure of the correlation between a vertical item and a horizontal item would be

$$\frac{N_{cell} \times N}{N_{ver} \times N_{hot}} = \frac{N_{cell}/N}{(N_{ver}/N)(N_{hor}/N)} \quad (1)$$

Table 2. Two Dimensional Association Analysis

		Vehicle type category			
		SUV	mid-size	full-size	luxury car
Location category	New York				
	Los Angeles				
	Seattle				
	Boston				

which is the point estimation of the exponential of the mutual information, given the number N as number of records in all of the data, N_{cell} as the number of records with both the horizontal and vertical items, N_{ver} as the number of records with the vertical item, and N_{hor} as the number of records with horizontal item. However, it can be inaccurate when the value of N_{cell}, N_{ver}, or N is not sufficiently large. To avoid this problem, we use the left terminal value (smallest value) of the interval estimation instead of the point estimation. Then the cells indicate smaller values than those obtained by the point estimation considering the uncertainty of the three density values in the right-hand member. By using this type of measurement, we can identify pairs of concepts that exhibit stronger relationships than other pairs.

4 Analysis and Results

4.1 Analysis Approach

In this paper, we are using telephonic conversation data from a car rental help desk. Such business-oriented conversations have the following properties.

- The conversational flow is concretely defined in advance.
- There are a fixed number of outcomes and each conversation has one of these outcomes.

We first identify the key concepts from the call transcriptions and group them under appropriate semantic categories. We hypothesize that there are two types of customer intentions at the start of call, *strong start* and *weak start* depending on willingness to make a booking, and such customer intention can be changed by the agent's actions. Under this classification, we prepared the following semantic categories:

- Customer intention at start of call: From the customer's first or second utterance, we extract the following intentions based on the patterns.
  - Strong start: *would like to make a booking, need to pick up a car, want to make a car reservation, . . .*
  - Weak start: *can I know the rates for booking a car, I would like to know the rates for a full size car, . . .*

  Under our assumption, the customer with a strong start are referred to as a **booking_customer** and the customer with a weak start just wants to know the rates and is referred to as a **rates_customer**.
- Discount-relating phrases: *discount, corporate program motor club, buying club* . . . are registered into the domain dictionary as discount-related phrases.
- Value selling phrases: we extract phrases mentioning good rate and good vehicle by matching patterns related to such utterances.
  - mention of good rate: *good rate, wonderful price, save money, just need to pay this low amount, . . .*
  - mention of good vehicle: *good car, fantastic car, latest model, . . .*

Using these categories, we tried to find insights to improve agent productivity.

4.2 Analysis Results and Actionable Insights

Table 3 shows the result of two dimensional association analysis between customer types and pick up results for 137 reservation calls. From these results, we see that 67% (47 out of 70) of the booking_customers picked up the reserved car and only 35% (13 out of 37) of the rates_customers picked it up. This supports our assumption and shows that the pick up result can be predicted from the customer's first or second utterance.

Table 3. Association between Customer Types and Pick Up Results

Customer types extracted from texts based on customer intent at start of call	Pick up result	
	pick up	not picked up
booking_customer(w/ strong start)(70)	47	23
rates_customer(w/ weak start) (37)	13	24

The results show that cars booked by a rates_customer tend to be in the "not picked up" category. Therefore, if we can find any actions by agents that convert such customers into "pick up," then the conversion rate as well as revenue will increase. In the booking_customer case, to keep the "pick up" ratio high, we need to determine if there are any specific agent actions that concretize the customer's intention.

Table 4 shows how mentioning discount-related phrases affects the pick up results for rates_customer and booking_customer. Table 5 shows how mentioning a good rate or a good vehicle affects the pick up results. From these tables, it can be seen that mentioning discount-related phrases and value selling phrases affects the customer's final status in both the booking_customer and rates_customer cases. In particular, for a rate_customer, the probability that the booked car will be pickedup, $P(pickup)$ is improved to 0.476 by mentioning discount-related phrases. This means that the customer is attracted by offering discounts and this changes the intention from "just checking rate" to "making a reservation here".

Table 4. Association between Mention of Discount-Related Phrases and Pick Up Results

$Rates\_customer$	Pick up result	
Mention of discount-related phrases by agents	pick up	not picked up
yes(21)	10	11
no(16)	3	13
$Booking\_customer$	Pick up result	
Mention of discount-related phrases by agents	pick up	not picked up
yes(40)	30	10
no(30)	17	13

As results, we derived the following actionable insights.

- There are two types of customers in reservation calls.
  - **Booking_customer** (with a strong start) tends to pick up the reserved car.
  - **Rates_customer** (with a weak start) tends not to pick up a reserved car.

Table 5. Association between Mention of Value Selling Phrases and Pick Up Results

$Rates\_customer$	Pick up information	
Mention of value selling phrases by agents	pick up	not picked up
yes(21)	10	11
no(16)	3	13
$Booking\_customer$	Pick up information	
Mention of value selling phrases by agents	pick up	not picked up
yes(40)	30	10
no(30)	17	13

- In the **rates_customer** case, the "pick up" ratio is improved by mentioning discount-related phrases.
- In both the **rates_customer** and **booking_customer** case, the "pick up" ratio is improved by mentioning value selling phrases.

4.3 Improving Agent Productivity

By acting on the actionable insights derived from this analysis in the actual car rental process, we verified improvements in the pick up ratios. Before doing the evaluation, we discussed these results with the agents having good pick up ratio. It was found that she was subconsciously dividing the calls based on a customer's strong or weak start in her and was dealing with the customers accordingly. However, before we spoke to her she had not consciously recognized the strong associations that we had found. For the evaluation, we divided the 125 agents in the car rental help desk into two groups. One group of 22 agents was trained based on the insights from the text mining analysis. The remaining 103 agents were not told about these findings. We compared these two groups over a period of two months to see how much the actionable insights contributed to improving the agent's performance. As the evaluation metric, we used the pick up ratio - that is the ratio of the number of "pick ups" over the number of reservations.

In the month following the training, the pick up ratio of the agents trained based on the insights increased by 3.75%. The average pick up ratio for the remaining 103 agents increased by 2.03% for this month. Before training, the ratios of both the groups were comparable. The seasonal trends in this industry cause bookings and pick ups to go up or down. However this group maintained this improvement the following month also. Considering this trend, it can be estimated that by acting on the actionable insights the pick up ratio was improved by about 1.72%. As a result, the contact center trained all of its agents based on the insights from the system.

4.4 Evaluation of Compliance to Guidelines

Finally we present some findings based on call segmentation and that can help in identifying lack of compliance to guidelines. The agents are supposed to perform some key tasks in each segment of the call. Example of such key tasks could be *verifying if the customer has a valid credit card, ask for future reservation*. Some of these tasks are mandatory for agents as missing them may have legal implications as well. Segmentation of calls into segmentation helps to identify cases where such key tasks are not performed. For the same we take a set of 60 calls and divided them manually into two sets depending on whether the call contains the key task or not, *positive* and *negative* set respectively. Now for each key task we define patterns based on concepts extracted from previous steps. For example, to check if the agent has confirmed that *the credit card is not a check/debit card* we can look for the keywords *check,cheque,debit,which* *is not,that is not*. We claim that searching such patterns is meaningful only when they are searched in the right segment and searching in the entire call body will lead to a lot of spurious matches. For example, confirming if the customer has a clean driving record should be present in *Mandatory checks*) and compare the result with the pattern matches in the entire call. The comparison results are shown below for the following key tasks:

1. Ask for sale
2. Confirm if the credit card is not a check/debit card
3. Ask the customer for future reservations
4. Confirm if the customer is above 25yrs of age
5. Confirm if the customer has a major credit card in his own name

Key Task	#1	#2	#3	#4	#5
No. of Calls	41	32	38	20	48
With Seg.	38	32	38	20	48
Without Seg.	35	19	1	12	1

Table 6. Statistics for Negative Instances

From the above distribution of negative instances i.e. calls missing some key tasks we can see that a large number of instances are being missed out without segmentation. The reason being the keywords indicative of such key tasks are likely to occur in different parts of the call and not only in the right segment. For example, consider the key task 3 i.e. the agent asking the customer for future reservations we look for keywords like *anything else,any other,help,assist*. These keywords are likely to occur in not only in *closing* segment but in other segments like greeting as well. So by looking at the entire call it is not possible to capture the information if the agent has performed a particular key task or not. Hence segmentation can really improve the task of identifying lack of compliance.

5 Conclusions and Future Work

In this paper, we have proposed methods for analyzing data from conversations to see how agent productivity can be improved. It was found that in business-oriented conversations as found in car rental help desks, the customers' intention at start of a call has an effect on the final result and that it can be changed by the actions of the agent. By analyzing the data from actual conversations, it appears we can get valuable insights considering the real voices of the customers. With the text mining technology used in this paper, it is very important to define concepts (e.g. categories) for the analysis to get useful insights. Constructing proper concept and domain dictionary needs an expert's skill. As a part of the future work we would like to devise automatic methods to extract the proper concepts and dictionary entry candidates from conversation data.

Acknowledgment

The authors would like to thank Raghuram Krishnapuram, Hideo Watanabe, and Koichi Takeda at IBM research for their supports. The authors also appreciate Jatin Joy Giri at IBM Daksh for providing domain knowledge about the car rental process and help in constructing dictionaries.

References

[1] S. Challa, S. Roy, and L. V. Subramaniam. Analysis of agents from call transcriptions of a car rental process. In *Proceedings of the Language, Artificial Intelligence and Computer Science for Natural Language Processing applications (LAICS-NLP)*, 2006.

[2] W. F. Cody, J. T. Kreulen, V. Krishna, and W. S. Spangler. The integration of business intelligence and knowledge management. *IBM Systems Journal*, pages 697–713, 2002.

[3] S. Douglas, D. Agarwal, T. Alonso, R. M. Bell, M. Gilbert, D. F. Swayne, and C. Volinsky. Mining customer care dialogs for "daily news". *IEEE Trans. on Speech and Audio Processing*, 13(5):652–660, 2005.

[4] D. Ferrucci and A. Lally. Accelerating corporate research in the development, application and deployment of human language technologies. In *Proceedings of the HLT-NAACL workshop on software engineering and architecture of language technology system*, pages 67–74, 2003.

[5] P. Haffner, G. Tur, and J. H. Wright. Optimizing svms for complex call classification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Hong Kong, April 6-10 2003.

[6] H.-K. J. Kuo and C.-H. Lee. Discriminative training of natural language call routers. *IEEE Trans. on Speech and Audio Processing*, 11(1):24–35, 2003.

[7] G. Mishne, D. Carmel, R. Hoory, A. Roytman, and A. Soffer. Automatic analysis of call-center conversations. In *Conference on Information and Knowledge Management*, Bremen, Germany, October 31-November 5 2005.

[8] T. Nasukawa and T. Nagano. Text analysis and knowledge mining system. *IBM Systems Journal*, pages 967–984, 2001.

[9] M. S. Neff, R. J. Byrd, and B. K. Boguraev. The talent system: Textract architecture and data model. In *Proceedings of the HLT-NAACL workshop on software engineering and architecture of language technology system*, pages 1–8, 2003.

[10] S. Roy and L. V. Subramaniam. Automatic generation of domain models for call centers from noisy transcriptions. In *Proc. of COLING/ACL 06*, pages 737–744, Sydney, Australia, July 2006.

[11] D. Sullivan. *Document Warehousing and Text Mining*. John Wiley and Sons, Inc., New York, 2001.

[12] M. Tang, B. Pellom, and K. Hacioglu. Call-type classification and unsupervised training for the call center domain. In *Automatic Speech Recognition and Understanding Workshop*, St. Thomas, U S Virgin Islands, November 30-December 4 2003.