1. INTRODUCTION

1.1. Problem Description

A typical customer service call center employs a few hundred to a few thousand agents organized into specialized agent pools. Agents in a particular pool receive intensive training in resolving a specific set of customer service issues. Commonly, call centers have different pools of agents to handle different issues such as billing inquiries, technical support requests, sales inquiries, etc. Before the advent of automated routing solutions, an incoming call would be routed to the next available agent who would then either handle the call himself or transfer it to the appropriate agent pool. For a transferred call, the agent who first picks up the call has to engage in a dialogue with the caller in order to identify the reason for the call. Given that a majority of calls end up being transferred, the call center incurs a substantial cost in terms of lost agent time.

With the advent of touch-tone based Interactive Voice Response (IVR) systems, call centers were able to deflect some of the costs by putting the user through an, oftentimes long, series of menus, which ask the user to make a numerical choice using the telephone keypad. While touch-tone IVR systems clearly represented an advance in terms of allowing the user to choose for themselves the service they need, these systems suffer from a crucial drawback: the user has to understand each menu and make the “correct” numerical choice in order to obtain the appropriate service. Many users either subvert the system by pressing random numbers in the hope of being able to talk to a human or abandon the call and hang up. Empirical evidence shows that 30-50% of the callers don’t get to the right place using touch-tone menus. Calls that end up being routed to the wrong agent are called misrouted calls. The misroute rate is a key measure of the usefulness of any automated call routing solution. BBN Call Director aims at reducing misroutes, providing easy access to automated self-fulfillment, and increasing customer satisfaction.

1.2. Speech in the call center

The first introduction of speech recognition into the call center workflow was in the form of speech menus: applications that allow the user to say one of a specified list of phrases. Examples of this approach are applications that ask the user to “Press or say 1”, or “For billing inquiries say billing, for repair inquiries say repair”. Speech menus are friendlier than touchtone interfaces but do not address the core problem in automated call routing - systems that ask the user to choose from a set of specified choices put a huge cognitive load on the user. The user has to quickly learn to match their request with one of the available phrases, many of which have cryptic wording or do not offer a precise match to the users’ problem.

Speech-enabled natural language call routing offers the potential to sharply reduce the misroutes and abandoned calls by allowing the user to specify in his or her own words the reason for the call. Natural language (NL) systems shift the cognitive load from the user to the system: now the system has to figure out what the user wants. This paradigm shift represents a major improvement in the user experience and in the operational efficiency of the call center. From a technological perspective, NL call routing is made possible by the combined use of statistical n-gram language models for speech recognition and a statistical topic identification system.

While the core NL routing technology itself has been available for a while [1,2], its commercialization has taken longer than expected. We believe that the major obstacle in the path of extensive deployment has been the lack of a solid methodology that proves the benefits of the technology, optimizes the benefits, and minimizes the operational and financial risks of deployment. BBN’s Call Center Assessment practice is an attempt at filling this crucial gap. The assessment methodology, described in more detail later, is a rigorous
procedure for evaluating the performance of the existing (typically touch-tone) IVR to establish a baseline, suggesting data-driven improvements and changes to the call-flow, projecting potential savings from deploying speech, customizing and optimizing the deployed speech-enabled routing solution (BBN Call Director), and finally assessing the real benefit of the speech-enabled solution vis-à-vis the baseline in terms of agent labor savings.

Section 2 contains a description of the technologies that constitute the BBN Call Director product. In Section 3 we describe the BBN call center assessment methodology and how it fits into the overall process of deploying natural language call routing in call centers with minimal risk. Section 4 deals with practical issues related to real deployments and Section 5 discusses typical performance numbers for the system. Section 6 is a brief review of the salient points in this paper.

2. BBN CALL DIRECTOR

BBN Call Director is a NL call routing system that uses a statistical language model for speech recognition and a statistical topic identification (TID) system to identify the topic for the call. The use of a statistical TID system makes the topic classification robust against speech recognition errors. Experimental evidence shows that the topic classification accuracy can be high even in the presence of a high speech recognition error rate. This robustness is achieved by exploiting the redundancy of topic-related information in callers’ response and partly by selecting an appropriate keyword set (see Section 2.2.1).

The block diagram in Figure 1 graphically illustrates the sequence of operations involved in the BBN Call Director. Rectangular blocks are operations to be performed and oval boxes represent models that need to be trained. As shown in Figure 1, the caller is first greeted with an open-ended prompt, usually, “Please tell me, briefly, the reason for your call today”. The caller’s response to the prompt is then fed to the BBN HARK speech recognition engine. Next, the recognized text from the speech recognition system is used by the topic ID engine for classifying the call to one of the destination topics. If the confidence in the topic classification is not high enough, the system engages the caller in a directed re-prompt, e.g., a prompt that lists the various options available to the caller. Callers whose responses result in a valid topic are routed to a topic-specific agent pool (or to automated fulfillment) and the remaining callers are routed to a default agent pool.

2.1. Speech Recognition Engine

The speech recognition engine used in the BBN Call Director is the BBN HARK recognizer [3] designed specifically for real-time telephony applications. BBN HARK is a Hidden Markov Model (HMM) based speech recognition engine that supports finite-state grammars as well as bi-gram language models.

In order to effectively model the variations in caller responses, we typically train a bi-gram language model with an appropriate back-off strategy [4]. This language model is typically trained on transcribed caller responses collected from the site as well as from a rich set of in-house language model training data. To further boost recognition accuracies, we also recommend training site-specific acoustic models using audio data collected from the deployment site.

The BBN HARK recognizer supports continuous and discrete density models for modeling the acoustic features. Typically, the discrete-density system is faster than the continuous-density system. For the call routing application we have found that the faster discrete-density models provide sufficiently accurate recognition.

2.2. Topic Identification Engine

BBN Call Director uses a statistical topic identification system to identify the topic for each call. The TID system [2] uses a multinomial model for keywords and incorporates two different classifiers: a Bayesian classifier and a Log Odds classifier. In addition, the TID system also supports rejection.

2.2.1. Keyword Selection

Our experience with topic identification has been that some words are more important than others. For example, if the callers’ response to the initial greeting is, “I am calling to dispute a charge on my bill”, then the words bill and charge are important from the perspective of topic classification whereas the words am, to, on, etc. are clearly irrelevant. Also the set of words that are useful in identifying a topic varies from topic to topic; this set of useful words is often referred to as a keyword set. In order to boost the performance of the topic classification algorithms, we automatically identify the keywords for each topic using information theoretic measures such as the Kullback-Leibler (KL) measure. The recognition grammar for the speech engine is also designed to provide better recognition performance on the keywords.

2.2.2. Multinomial Model for Keywords

A caller’s response can be defined as a sequence of words \( r = \{r_i\} \), where each word \( r_i \in W = \{w_1, \ldots, w_M\} \). \( W \) is a keyword set of \( M \) words and includes a non-keyword symbol which substitutes for any word in the response that is not a keyword. We define \( R \) to be the set of all possible caller responses and \( T = \{t_1, \ldots, t_N\} \) as the set of all system topics. The probability density functions of the caller’s response conditioned on topics can be modeled as a multinomial distribution as follows:

\[
P(r | t_j) = \prod_{i=1}^{M} p(r_i | t_j)^{n_i(r)}
\]

\( n_i(r) \) is the number of times words \( w_i \) occurs in \( r \). The parameters \( p(w_i | t_j) \) of the multinomial model are trained using
maximum likelihood (ML) estimation. Given a finite set of labeled training utterances \( X = \{ (t^i, r^i) \in R \times T \}, \) for each \( i = 1, \ldots, M \) and \( j = 1, \ldots, N \), we compute \( n_{ij} \) as the number of occurrences of word \( w_i \) in all responses labeled for topic \( t_j \). The ML estimate for \( p(w_i | t_j) \) is:

\[
p(w_i | t_j) = \frac{n_{ij} + M_j}{\sum_{i=1}^{M} n_{ij} + M_j}
\]

where, \( M_j \) denotes the number of unique words that occur in topic \( t_j \). The estimation equation above uses the well-known Bell-Witten back-off strategy to account for words that were not seen in the training data for a certain topic. In addition, we also use the ML estimate for the prior probabilities of each topic - by using the frequency of the topics’ occurrence among the training samples in \( X \).

2.2.3. Bayesian Classifier

Given the probabilities \( p(r|t_j) \) and a priori probability distribution \( p(t_j) \) for the set of topics \( T \), a Bayesian classifier is constructed for maximizing the posterior probability \( p(t_j|r) \forall j \), where,

\[
p(t_j|r) = \frac{p(r|t_j)p(t_j)}{p(r)}
\]

The TID system returns a list of topics which have the probability \( p(t_j|r) \) above the rejection threshold specified to the system. The topic with the highest probability is generally used to route the call to a particular agent pool. In the case where no topic has a probability \( p(t_j|r) \) above the rejection threshold, the TID system returns a NULL topic to the BBN Call Director application.

2.2.4. Log Odds Classifier

Given the probabilities \( p(r|t_j) \) and a priori probability distribution \( p(t_j) \) for the set of topics \( T \), a log odds classifier is constructed to maximize the posterior topic log odds for the caller response:

\[
\log \left[ \frac{p(t_j | r)}{1 - p(t_j | r)} \right] = \log p(t_j | r) - \log \left[ 1 - p(t_j | r) \right] + \log p(r | t_j) - \log \left[ 1 - p(r | t_j) \right].
\]

Again as in the Bayesian classification framework, the TID system returns a list of topics with posterior log odds greater than the rejection threshold.

3. CALL CENTER ASSESSMENTS

BBN’s Call Center Assessment [5] is a patent-pending methodology for measuring, analyzing, and improving the performance of an IVR system. At the core of our assessment methodology is the idea quantifying IVR benefit in terms of agent seconds saved by analyzing thousands of live end-to-end calls. The agent-seconds-saved metric is a comprehensive, objective measure of the usability and cost-effectiveness of IVRs. The agent-seconds-saved metric helps to accurately estimate the cost-savings potential for call flow redesign as well as objectively compare alternative IVR designs.

In an IVR assessment, we record thousands of live calls in an unobtrusive fashion, and apply automated tools to determine the complete IVR event sequence for each call. Often, the IVR event sequence can be obtained from IVR logs. When such IVR event logging is not available or practical, we use algorithms, that we have developed, to automatically infer the complete IVR event sequence from the call recordings by detecting specified prompts and associated spoken or keyed-in caller responses. For calls that are transferred to live agents, the dialog between the agent and the caller is transcribed and significant dialog events are identified. Thus, we create a database of event traces for complete calls and compile accurate statistics on how successful callers really are in the automated system. We can also estimate the amount of missed automation and number of misrouted calls. The assessment is a crucial component in the development and deployment of NL call routing in four ways: building the case, guiding the design, proving the benefit, and minimizing risk.

Before moving towards NL call routing in any call center, we typically conduct a baseline assessment in order to quantify the performance of the existing IVR and measure the potential for improvement. Only if we identify a significant opportunity for improving routing, which may include routing to specialized agent pools or automated self-fulfillment, do we proceed towards a NL call routing solution. Besides helping to build a case for considering NL call routing, the initial assessment helps minimize financial risk for the call center because it typically provides easy-to-implement improvements for the existing system. These improvements can yield substantial and immediate cost savings that help offset the cost of configuring the NL call routing system.

If the data from the initial assessment justifies the business case for NL call routing, we perform a trial of BBN Call Director in the production environment. Such a trial serves two purposes: configuring a system that’s ready for deployment, and proving the benefit of NL call routing. Thus, the call center can defer the decision on capital investments that are necessary for a deployment until the benefits are proven and the business case is backed by solid evidence.

The first phase of a trial is the design of the speech-enabled solution and here the initial assessment provides important guidance. From the annotation of caller-agent dialogs, we infer the distribution of true reasons for calls. The call reason distribution is useful in identifying the categories for routing destinations. It also helps to focus the design efforts on the frequently requested services or services that are otherwise important for the customer, thus preventing us from wasting development effort.

Finally, once the trial system is configured we measure its performance using the same process that was used to evaluate the IVR baseline. We measure improvements in the main components of IVR automation: capture of customer account numbers, correct routing of callers to specialized agent pools or automated services, automated delivery of information to callers, and transaction completion. By applying our assessment methodology [5], improvements in partial call automation are translated into estimated average savings in agent seconds per call.

4. DEPLOYMENT STRATEGY

Developing high performance speech-recognition systems typically requires collecting and manually transcribing an adequate amount of audio (caller) data from the deployment
In BBN Call Director, training the topic models requires that the training data be labeled with the appropriate topics. However, the performance improvement from site-specific training data comes with two negative effects, (a) increased deployment expenses, and (b) slower pace of deployment. To reap the benefits of rapid deployment along with the additional benefits derived from the use of site-specific training data, BBN has evolved the following 2-phase deployment strategy.

4.1. Phase 1: Bootstrapped Models

In the first phase of the deployment that typically lasts for about two weeks, we deploy the BBN Call Director application using stock acoustic models along with language/topic models trained on a selected subset of available in-house data. Typically, the in-house data contains training data for most of the topics for a given site. In the case of topics for which there is no in-house training data, we artificially generate topic-training data by applying simple transformation rules to convert from existing topics to new ones. During this process, we also use the data collected during the initial assessment.

4.2. Phase 2: Training on site-specific data

With the bootstrapped models in place, we collect new training data from the incoming calls using the BBN Call Director application itself. By studying the application logs, we analyze the topic-wise performance of the detection engine as well as the topic classification engine. Based on this topic-wise analysis, we select (more) training data for topics that exhibit the most need for improvement. The selected training data is transcribed and the acoustic, language, and topic models are appropriately updated.

To further reduce the cost of training the models with site-specific data, we have developed unsupervised training procedures that use a small amount of transcribed training data along with large amounts of un-transcribed acoustic training data [6]. Researchers at BBN [6,7] and elsewhere [8] have earlier demonstrated that the recognition performance of models trained using un-transcribed training data can approach the performance obtained by training on manually transcribed data. The key to the success of the approaches outlined in [6-8] is the use of a reasonably good bootstrap acoustic model and a powerful task-specific language model – both of which are available in the context of the BBN Call Director.

Using the unsupervised procedure in [6] to train the BBN HARK recognizer results in an increase in the speech recognition word error rate (about 5% absolute) as compared to the performance obtained using manually transcribed data for training but the topic classification accuracy was not significantly affected by the increase in recognition error rate. Further improvements in performance can be obtained by a better selection of bootstrap models and by using the transcribed training data from other deployment sites. Since the overlap in topics is significant from one site to another, a substantial amount of data from one site can be used directly to train language and topic models for building a reasonable bootstrap system for another site.

5. SYSTEM PERFORMANCE

Topic classification accuracy in the BBN Call Director reaches 80-90% or higher, with some variation among different sites and varying levels of rejection. But the real metric of the performance of the BBN Call Director is the reduction in misroute rate, the savings in terms of agent-seconds, and the impact on customer satisfaction. On all three measures, BBN Call Director significantly outperforms touch-tone IVR systems. In a representative customer survey conducted by an independent third party, 84% of the customers surveyed preferred speech input while only 14% preferred touch-tone. In the most recent trial, the use of the BBN Call Director reduced the number misrouted calls by about 28% resulting in a savings of 2-4 minutes for each of those calls.

6. CONCLUSIONS

In this paper, we have presented an overview of the natural language call routing solution that BBN currently offers. As discussed, the call routing solution along with the deployment methodology brings a unique approach to the call center marketplace by fusing together expertise from a variety of different, but related, spheres: speech recognition, language processing, human factors, and call flow design. We believe that speech-enabled technologies will be an integral part of the customer response technologies of the future. While the business case for putting “voice into the network” is rapidly evolving into a compelling story, speech recognition is only one part of the infrastructure – interpreting and associating known meanings to the recognized text is the other part.

Language processing technologies such as topic identification and information retrieval allow relevant information to be extracted from the recognized text. The extracted information is then fed into various existing back-end processes in a manner that delivers tremendous savings to the enterprise.

7. References