



Data Mining in Call Centers: The Overlooked Interaction between Employees

Mr. James Lohse

Jim Lohse is a Computer Science student at UVU interested in Data Science and Machine Learning for the technical support industry.

Dr. Reza Sanati-Mehrizy, Utah Valley University

Reza Sanati-Mehrizy is a professor of Computer Science Department at Utah Valley University, Orem, Utah. He received his M.S. and Ph.D. in Computer Science from the University of Oklahoma, Norman, Oklahoma. His research focuses on diverse areas such as: Database Design, Data Structures, Artificial Intelligence, Robotics, Computer Aided Manufacturing, Data Mining, Data Warehousing, and Machine Learning.

Dr. Afsaneh Minaie, Utah Valley University

Afsaneh Minaie is a professor of Computer Engineering at Utah Valley University. She received her B.S., M.S., and Ph.D. all in Electrical Engineering from University of Oklahoma. Her research interests include gender issues in the academic sciences and engineering fields, Embedded Systems Design, Mobile Computing, Wireless Sensor Networks, Nanotechnology, Data Mining and Databases.

Data Mining in Call Centers: The Overlooked Interaction between Employees

Abstract

Many data mining techniques have been applied to technical support call center interactions between customers and employees. There remains a vast, largely untapped source of data related to intra-employee interactions. In typical call centers, there are chat rooms for more experienced first-level support technicians to help newer, less experienced technicians. In addition, when a first level technician needs more assistance, they will contact a second-level employee by phone or chat.

Mining these records of inter-employee interaction could lead to improved interaction and eventually to a machine learning system that can substitute for the second-level employee in the initial stage of assisting the less trained first-level employees. One of the authors has worked in several technical support call centers and therefore has detailed knowledge of their operations and areas that can be improved. This paper was inspired by this author noticing much of the communication between various levels of employees is not retained or analyzed, leading to a phenomenon of repeating and recreating the same efforts day after day, training class after training class.

This paper proposes a system of studying the interactions between first- and second-level technical support employees using data mining and machine learning techniques. The proposed system would be inserted into the normal flow of information between first-level tech support employees and the second-level co-workers they consult with when they need help. The overall goal of this proposal is to intercept solutions that are provided on an ad-hoc basis and capture them into a corporate database that will present these solutions in the future, as they are learned. This type of system would be flexible enough to adapt to new products and new questions as new questions are posed to second-level experts.

Introduction

In a typical technical support call center, there exists a two-tier structure. There are first-level technical support workers of varying skill, ranging from those barely out of a training class to those who are ready for promotion to the second tier. The most skilled and productive of these experienced employees are promoted to second-tier. Most first-level technical support employees are hired primarily for personality, with the assumption they can be trained for specific technical requirements.

We will look at previous data mining work done on call center data sets to see if that work can be extended to studying the interaction among-first tier employees and their interactions with second-tier employees. Interactions among first-level employees in chatrooms will also be studied.

Some of the benefits of mining the interaction between first level employees and matching it with metrics like call time is to bring calls to second level employees faster, to improve the customer experience. If a call is going to be escalated, it would be better for the customer experience to do this quickly, rather than letting a first-level employee perform unproductive troubleshooting, only to repeat that troubleshooting when the second-level technician takes over a call.

Another area that can save time for customers is when a second level technician has to take time to

decide to take over a call, even when the first level employee does not want to let go of the call. At times, a second-tier technical support person will decide it's best for the customer for them to take over the call. While it's good for first-level employees to show initiative, there are times where a call should clearly be escalated. Making this decision can take several minutes, while the customer is on hold. Identifying ways to shorten this process would lead to happier customers. Instead of waiting on hold up to several minutes and let the two levels of employees negotiate and then escalates this call to a second level employee, a sequence mining algorithm can predict the result of this conversation and escalate the call to a second-level employee without going through the hold process. These types of problem are nearly always classification, not clusters, "since the prediction is made based on some historical data"¹ based on many records of previous calls and their outcomes. In some cases when completely new product is released, it may become a problem related to clustering, while the data set needed for classification is built up.

Differentiating Between Structured and Unstructured Data-- Honeywell Study

In one study² of contact records at Honeywell, the researchers looked at ways to classify both structured and unstructured data. The structured data included the various times the customer spoke with a technical support engineer at Honeywell, the customer name and other fixed data like call time, while the free text entered by the Honeywell engineer was unstructured. Various tokenization techniques were used, such as replacing instances of IP addresses with the term IPAddress and replacing words like "shut down" with "shutdown." The researchers came up with five categories ranging from keywords to low-frequency terms. It was only the keywords and alphabetical strings that were used with a naïve-Bayes classifier from the Bayesian Knowledge Discovery project.³ The C4.5 algorithm was also used⁴ and 20,816 cases were analyzed.

A similar approach using updated tools could be effective when looking at the interaction between employees in a support environment. There are two sources of text to be analyzed. First, there are chat sessions between employees. Second, when a second-level employee takes a call from a first-level, or simply provides an assist, there are sets of notes that can be analyzed. The first-level employee should be taking complete notes about the nature of the case and what troubleshooting did not resolve the problem. Then the second-level employee will treat the contact as a consultation and let the first-level employee keep the call, entering notes about the advice given. Or, in the case where the second-level employee takes over the call, more complete notes will be available of the actual solution for the first-level employee to view later. It would be helpful if the call tracking / problem resolution system allows the employees to use separate fields to describe the problem, instead of entering a solution, and the employees were encouraged to use this consistently.

Interestingly, this study showed that clustering can be an effective method for analyzing service calls at Honeywell, while classification may be preferred when a large body of existing cases and experience provides a clear delineation of call types.

Prediction of Call Outcomes Based on Employee Skill Level

It's a challenge to predict the best and fastest solution for a situation where a first-level employee seeks help from either a chat room or by contacting a second-level. If the employee skill level could be part of the equation, better estimates might be possible. The question is, how do we measure the quality of the first-level employee seeking assistance?

Another study from 2012 looked at sales and marketing call centers to predict the job performance

using a naïve Bayes classifier.⁵ In a studying of 1037 sales agents in 2009 in an insurance environment, “it has been shown that, socio-demographic attributes are not suitable for predicting performance.” While this study focuses on performance and employee turnover, it is instructive in the way it approaches classification of good employees. The study is based on a call center in Santiago, Chile. This call center has a high standard for keeping employees based solely on sales performance related to insurance products.

Each employee is hired for a one-month time frame. At the end of the month, if they have not met certain specified sales goals, they are not retained as an employee. This process is repeated at the end of the second and third months. If the employee meets goals three months in a row, they are kept as a regular employee. This approach does introduce a bias because employees who might perform poorly in the short run, but do well in the long run, are not part of the data set.

Various measurements were looked at that would translate well to a technical support environment, primarily a measure of the number of hours active on the phones and the number of hours actually spent talking to customers. These indicate motivated employees who do not take a lot of sick or personal time, providing the most opportunity for the best performance. Some data cleaning was necessary, for example to eliminate employees who were hired before the study period began.

One result of the study is “from the operational point of view, the inference over the model allows us a form of control and monitoring of sale agents performance, to the point that it is possible to detect early (for example, in the first month of work) the likelihood that person will achieve above minimum.” This could be used in reverse, to indicate for those with a likelihood of achieving below minimum, the second-level employee taking a consult could know to take the call over more quickly. The first-level employee can still learn from these cases by studying the notes made later by the second-level who took over the call.

The conclusions of the study mostly relate to retaining high-performing employees in a position that does not require much education or professionalism – this description translates well to first-level technical support organizations where employees are often hired on personality, not experience. The important part of the study is the part of the model that classifies employees as high-performing, so management can focus on the lower-performing ones. In context of this paper, we are looking at the method used to do the classification.

Methods to “Listen In” on Call Center Phone Calls – DECODA

The DECODA project⁶ is summarized in the paper, “DECODA: A call center human-human spoken conversation corpus.”⁷ This paper does work in the area of analyzing, among other things, how polite the call agent is to the customer. In the area we are studying, if a system can detect that a first-level employee has not been polite to the customer, we can program a system to recommend a second-level employee take over the call – before the first-level contacts them!

One of the main problems identified in this paper is that of anonymizing data before it can be released to a research project. As noted in several articles recently, including the Wall Street Journal⁸ and New York Times⁹, supposedly anonymous data isn’t necessarily anonymous. According to a paper in the journal Science¹⁰, using just four data points, 90% of anonymized credit card transactions could be tied back to the individual making the purchase. The chances of identifying an anonymous person increased when the price was among the data, and women were more susceptible than men to re-identification in credit-card transaction based studies. From the abstract, “We show that knowing the price of a

transaction increases the risk of reidentification by 22%, on average. Finally, we show that even data sets that provide coarse information at any or all of the dimensions provide little anonymity and that women are more reidentifiable than men in credit card meta data.”

For a company working internally, this is not an issue, so it’s worth looking at DECODA to see what success they had. They did remove all references to names, addresses and other personal data, before transcribing the audio. This was a manual labor intensive process where they covered over the personal data with a beeping sound. Next they broke the call up into time-based parts, calling this a “segmentation” process. They chose to divide the call process by, “opening, problem presentation, problem resolution, closing. Then a speaker segmentation process is performed in order to separate the turns of each speaker.” Because they recorded all the audio on a single channel, they did not have the easy ability to separate the caller from the first-level employee – this was also a fully manual process that could be automated based on using multiple channels or on identifying the speaker by voice pattern.¹¹

Finally was the transcription where, again by a manual process, each sentence was broken out by listening for pauses. Much advancement has been made in automating this three-step process that would need to be applied in a call center environment where the results are expected to be real-time.¹²

Architectures That Rely on Too Much Technology

In reviewing a Chinese paper¹³ that mostly focused on the technical architecture of a call center for agricultural products, we can see one approach to call centers that is perhaps relying on too much technology and not enough human touch. They focus on ACD and IVR systems to manage the customer before they get to a human being, such that “queuing strategy ... is one of the core technologies of Web call center, queuing strategy can improve the service efficiency of the call center, customer call is assigned to the appropriate position, provide better service for customers.” There is a place for assigning more valuable customers faster service, but keeping the human touch is always a challenge.

In healthcare and other industries that work with many older people, acceptance of IVR systems is low¹⁴, because “users tend to perceive them as obstacles installed by companies to keep callers from reaching expensive human agents.”¹⁵ There are systems that have humans monitor the IVR interactions and actively help a person navigate their way through. There are even systems that allow humans to inject pre-recorded segments into a conversation, such that the user doesn’t realize they are listening to a computer.¹⁶

All of these systems would be inferior to a system that would let a human being take a call with very little prompting from an automated system. Often callers greet first-level employees with frustration at the difficulty they had with an automated system that, for example, could not understand them reading their serial number. At this point the employee must ask for the serial number again, frustrating the caller even more. This problem would be remedied by the approach suggested in this paper.

Case Study of a Large Consumer Electronics Company

In the experience of the authors, one of whom has worked in call centers of various sizes, doing support for a very large and well-known company, too much emphasis is placed on managing the customer interaction, but not enough on capturing and retaining the mass of information that flows by every day. We have seen situations as described through this paper where a first-level tech support agent has a few

key resources to get help. When that agent first gets out of training, he may know only 10-15% of what he eventually needs to know, and may understand even less of the companies policies and procedures relating to repairs, updates, returns, upgrades, warranties and support policy.

That newly trained first-level employee can turn to three key resources: a knowledge base, a chatroom with fellow first-level employees and a call to a second-level employee. Because a common metric is to see how often a first-level employee can resolve an issue without going to the second-level, there is an incentive to spend an inordinate amount of time trying to own the problem. Using a chatroom can be inefficient for reasons outlined later.

Let's take a look at how big companies commonly manage their employee chatrooms. First, there is not one big chatroom for all employees. Instead there are a number of chatrooms with a manageable number of employees so the questions don't go by too quickly to be answered. Second, there's still a problem of multiple people asking questions at the same time, so a question scrolls off the screen before anyone can notice it. This forces the person asking the question to repeat themselves, which makes other people asking questions frustrated that their question now has low priority.

The interesting thing about these chatrooms is they are usually not monitored by a machine learning and predictive analytics system, even when they represent a huge data set that lends itself for a classification approach. Often the nature of a chat is something like this:

Questioner: "I have a customer who is not getting email on their phone, how I do handle that?"

The questioner was probably trained on this but in the pressure of the moment does not take the time to research the knowledge base, instead they want someone to give them an easy answer. In some cases a more experienced employee will actually walk them through the steps to resolve the problem. More commonly, someone will post the document number of the knowledge base solution that walks the questioner through troubleshooting.

Explanatory Answer: "Try first to check the internet connection, then see if they can log into a web-based version of their mail.

Reference Answer: "Take a look at TS3899, it will walk you through all the steps."

If the first-level employee has more questions, it's possible that other employees will also jump in to help. A system would need to be devised to track what help the employee has been offered, to factor in what we will call the "flail factor." The flail factor is a measurement of how much external help an employee needs in helping a customer. There are many ways this could be measured, based on length of call, number of chats during the call looking for help, and perhaps even a "raise your hand I am flailing" button that the employee could press to indicate their distress before they decide to escalate to the second-level employee tier.

Room For Improvement in Chat As Well

Similarly, for first-level employees managing multiple chat sessions, one step is already eliminated – the voice does not need to be transcribed in real-time, as the customer and employee are already communicating by text. More research needs to be done in this area to identify existing papers that would shed light on this subject.

Proposal

This paper is proposing to locate willing call center operations in the Salt Lake City and Utah Valley areas of Utah. This area, known as the “Silicon Slopes.”¹⁷ has a core group of data scientists who hope to make Utah a leading area of data mining, predictive analytics and machine learning. This area is a prime location for the research proposed in this paper because it hosts many call centers from companies as diverse as Visa, NuSkin, Adobe, EMC, Bluehost, Domo and several contract call centers.

Between the University of Utah, Brigham Young University and Utah Valley University, there is a critical mass of all levels of university students studying machine learning and data mining. Not only is there a data mining focus at Utah Valley University, the University of Utah also has a program and infrastructure specializing in Data Management.¹⁸

The plan being proposed here is to approach these companies to determine who is willing to sanitize and share their data sets for analysis. The audio recordings and chat sessions that take place between various internal employees would be analyzed with two main emphases. The first emphasis would be to capture questions that are repeatedly posed and reduce the workload on second-level employees. The second emphasis is to notice when the questions being asked are not known questions, but instead signify a new trend in terms of what the first-level employees are struggling with.

Currently trends are spotted by the way that first-level employees classify cases. This depends on two things that are not always present: new employees that are trained in how to properly classify cases, and first-level employees taking the time (and having the presence of mind in the middle of a call) to properly classify cases. Human-based classification is only good at spotting call trends when the employees use the system properly.

This proposal has a beneficial side effect that classification would be done automatically, therefore increasing the ability to spot new trends and redeploy call-taking and training resources accordingly. To implement this proposal, the following the proposed system (figure 1) would be inserted into the normal flow of information between first-level tech support employees and the second-level co-workers they consult with when they need help.

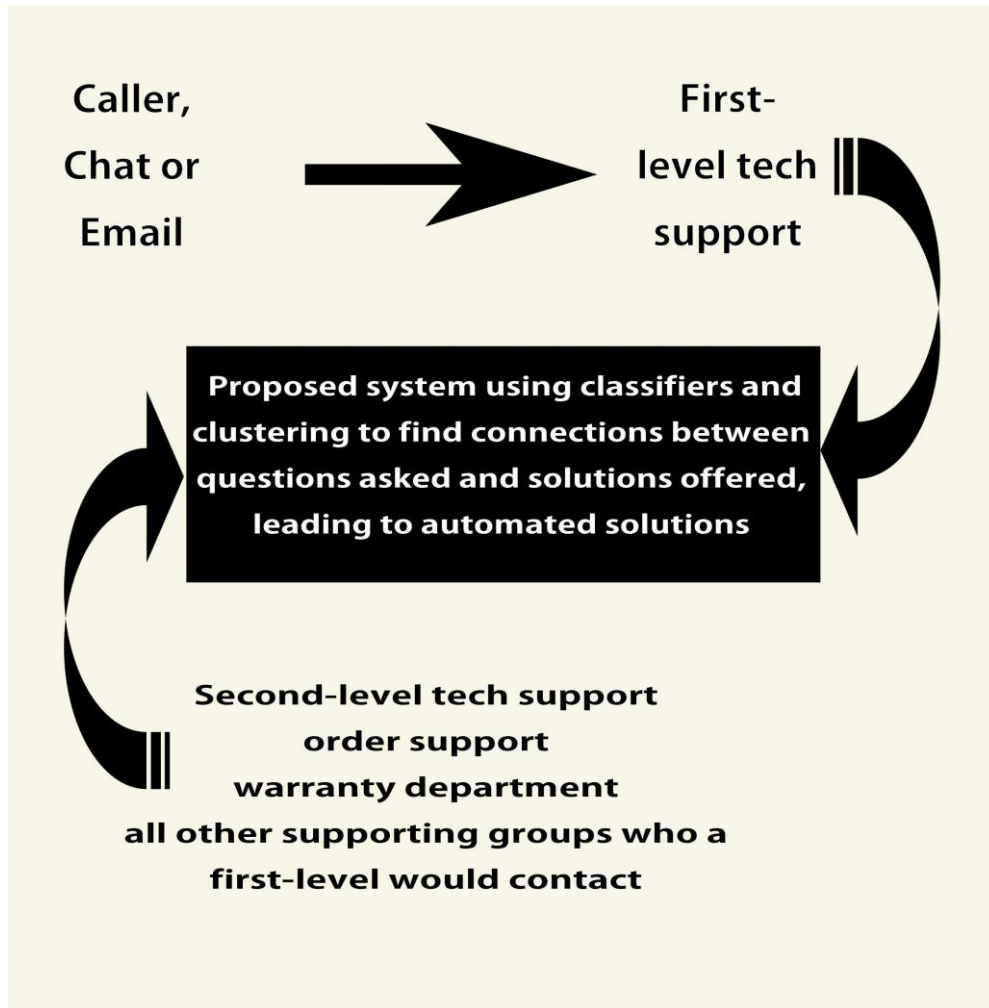


Figure 1: Flow diagram showing where calls and chats are intercepted

Best Approaches for Call Center Machine Learning

To accomplish this proposal, three primary technologies will be used, natural language understanding, probabilistic topic modeling, and random forest trees. Neural networks may be used at a later stage of the process as well.

The first component of the proposed system is natural language processing, and specifically natural language understanding. The chat sessions are already in text form but the conversations we hope to capture and analyze are not. The initial stages of the proposal would focus on chat sessions and move directly to the next section on topic modeling when it comes to processing the phone consultations between first- and second-level employees.

Next up is probabilistic topic modeling¹⁹ outlined in a paper published by Princeton, “Topic modeling algorithms are statistical methods that analyze the words of the original texts to discover the themes that run through them, how those themes are connected to each other, and how they change over time.” The “change over time” part is critical to this approach. As noted in the paper, if a person were to try to read all the chats between employees read transcripts of all the phone consultations, there would be far too much information for that person to pick up all the themes contained therein.

The proper use of topic modeling can identify existing *and emerging* themes. Existing themes are the known problems like email or internet connection issues. Unknown themes may come out during product releases. There may be new features or new bugs, and both cases will be caught by a properly implemented system without being programmed by experts ahead of time.

Finally, once all speech has been converted into text, and that text has been broken down by theme, the next step is to look for correlations between the two parties to the chat or conversation. The approach proposed here is to look for correlations between what the first-level employee is saying and how the second-level employee responds. As noted, another important component of this interaction is to note whether the call or chat is kept by the first-level employee or is escalated to the second-level employee. The correlations can be found by applying the methods of random-forest trees. At this point a domain expert is needed to look at the results and ensure that there is no over fitting of the models produced. Figure 2 represents this three steps process.

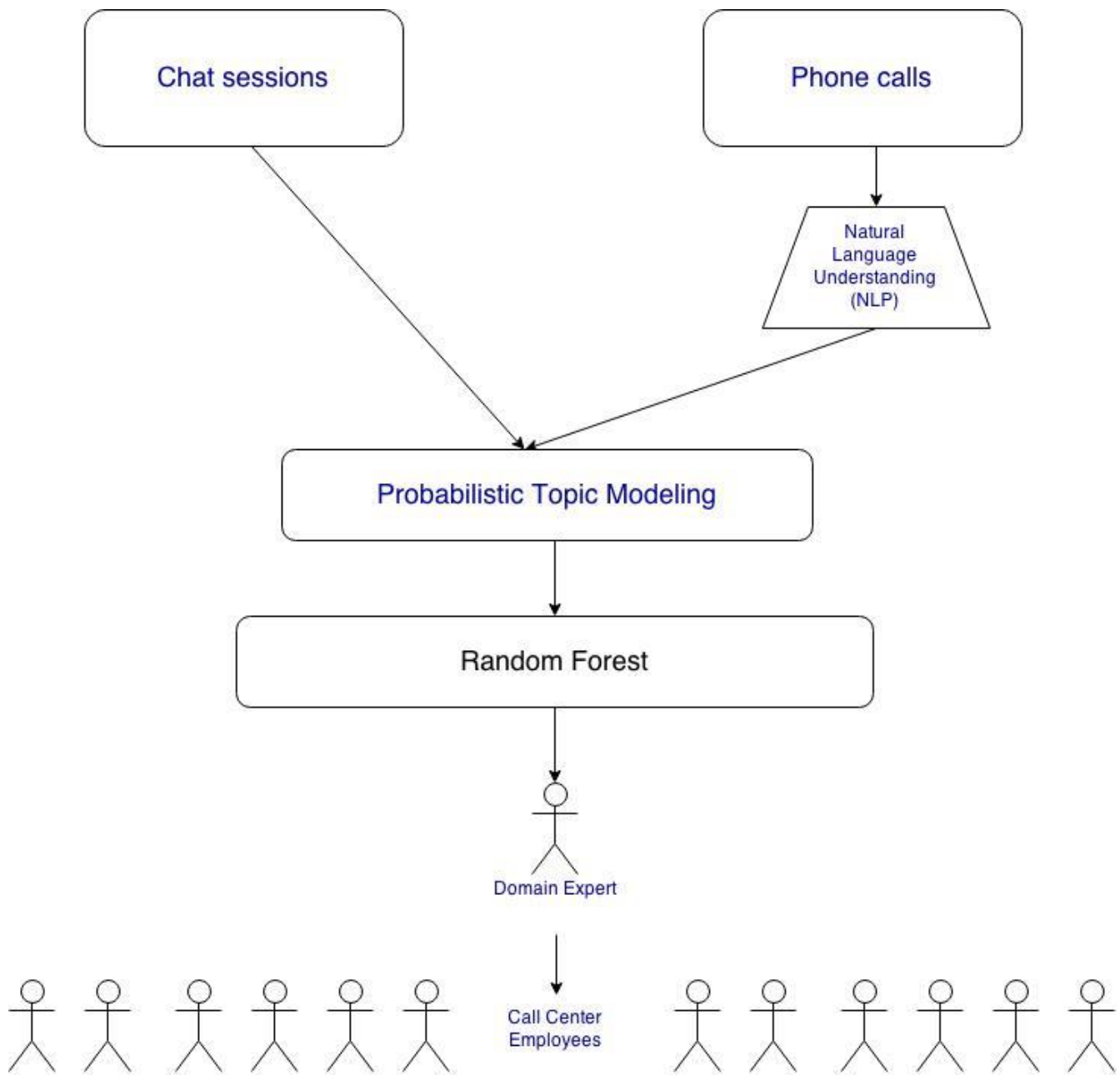


Figure 2: Placement of technologies in call center flow

Conclusion

By combining different proven technical approaches in a new way, call times can be dramatically reduced and customer satisfaction can improve. Even for an organization that already scores well on customer support, there is room for improvement. In the preceding case study, we propose to study a system that will do several things.

First an AI / Machine Learning system will monitor these employee chats. When it spots a pattern it can recognize, such as a first-level employee asking for help with mail troubleshooting, it would be

easy for the AI system to inject its answer into the chat, even in such a way the employee has no idea he is being prompted by a computer. The computer could inject, “Try TS3899, have you tried the first step of checking the internet connection”

If the employee then responds to the AI, it might be an existing pattern, or if it is something unrecognized, the AI system could “say” something like, “I am on a call. You might want to contact a second-tier employee about that if the steps in TS3899 don’t help.”

Then, it can be noted in the call center that this particular first-level employee needs help with a mail call. Based on experience, the company might know that these calls can take a long time for new employees to solve. Based on the number of times the caller has contacted the company, based on survey results from previous interactions, and based on the availability of second-level employees, a more experienced person could know to take over the call immediately.

Alternatively, the AI system could alert an experienced employee to contact the struggling first-level employee in the chatroom to offer specific help. Each company would have to design a threshold to decide when to let the new employee benefit from solving his own problems, versus having a more experienced employee get involved in the interest of a better, more efficient customer experience.

The key points, and what we propose to study, are ways to newly monitor the unmonitored data flows inside technical support call centers. Every day corporate memory is created and lost when the solutions in these chatrooms disappear into the ether. A system that would monitor AND retain the knowledge in these chatrooms would be a boon to reduce technical support costs. Mining the notes of interactions between first- and second-level employees would be another great source of knowledge that needs to be retained in corporate database for future use. This would reduce call times and empower newly trained employees, leading to even high rates of customer satisfaction.

References

1. Dong, G., & Pei, J. (2007). Examples of Sequence Mining Applications. In *Sequence data mining* (p. 5). New York: Springer
2. Tan, Pang-Ning, Blau, Hanna, Goldman, Robert & Harp, Steve (2000). Textual data mining of service center call records, KDD’00 417-423
3. <http://projects.kmi.open.ac.uk/bkd/> -- Built for Windows NT
4. J. Quinlan, C4.5: *Programs for Machine Learning*. Morgan Kaufman, 1993
5. Valle, Varas & Ruiz (2012). Job performance prediction in a call center using a naive Bayes classifier. In *Expert Systems With Applications* (pp. 9939-9945) Vol 39(2012) Elsevier
6. <http://decoda.univ-avignon.fr/projet.php>
7. Bechet, F., Maza, B., Bigouroux, N., Bazillon, T., El-Bèze, M., De Mori, R., & Arbillot, E. DECODA: a call-center human-human spoken conversation corpus.
8. <http://www.wsj.com/articles/metadata-can-expose-persons-identity-even-when-name-isnt-1422558349>
9. http://bits.blogs.nytimes.com/2015/01/29/with-a-few-bits-of-data-researchers-identify-anonymous-people/?_r=0
10. <http://www.sciencemag.org/content/347/6221/536.full?intcmp=collection-privacy>

11. Speaker Identification by Speech Spectrograms: A Scientists' View of its Reliability for Legal Purposes, <http://www.haskins.yale.edu/Reprints/HL0090.pdf>, retrieved 2/1/15
12. Ortega, Lewis, vanBuskirk, Wang, Maes (2001) U.S. Patent US6332122 B1, Washington, DC : <https://www.google.com/patents/US6332122> Transcription System for Multiple Speakers, using and establishing identification
13. Zhu, Zhang, Lin & Gan (2014) Design call center management system of e-commerce based on BP neural network and multifractal.
14. Messier, C., Miller, D. I., Gagnon, M., & Talbot, V. (2010). I hate Emily: why interactive voice response systems are unfriendly with older people and how to fix it. *Gerontechnology*, 9(2), 231.
15. Bouzid, Ahmed, and Weiye Ma. "Back to the future: bleeding-edge IVR." *interactions* 17.3 (2010): 18-20.
16. Now you can love your call center, <http://www.perfectpitchtech.com/>, retrieved 2/1/2015
17. Silicon Slopes, <http://siliconslopes.com/> retrieved 3/10/15
18. Big Data Certificate (and other topics related to MS and PhD programs) <http://www.cs.utah.edu/bigdata/> retrieved 3/10/15
19. Probabilistic topic models, <http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf> retrieved 3/10/15