Next Generation of Interactive contact centre for efficient customer recognition : Conceptual Framework

Morteza Saberi, Elizabeth Chang, Omar Khadeer Hussain
School of Business,
UNSW Canberra,
Australia

Zahra Saberi
Department of Industrial Engineering,
University of Tehran,
Iran

Abstract— Contact centers, as the organization’s touch point, have a considerable effect on customer experience and retention. It has been shown that 70% of all business interactions are handled in contact centers. A framework is proposed in this conceptual paper to build cleaned interactive customer recognition framework (CICRF) in CCs. CICRF consists of two integrated modules: cleansing and ICRF. The first module focuses on the detection and resolution of duplicate records to improve the effectiveness and efficiency of customer recognition. The second module focuses on interactive customer recognition in a customer database when there are multiple records with the same name. Cleansing module uses Semi-Automatic deduplication process by incorporating three main functions in its design, namely: DedupCrowd, DedupNN and DedupCSR. DedupCrowd is a function that provides training pairs of records for DeduppNN which is a deduplication based neural network. Researchers suggest leveraging human computing power in managing duplicate data which is scalable to the large size of contact centers data. However completion of crowdsourcing tasks is an error-prone process that affects the overall performance of the crowd. Thus, controlling the quality of workers is an essential step for crowdsourcing systems and for that I propose OSQC, an online statistical quality control framework, to monitor the performance of workers. DeduppNN is a neural network based deduplication method that uses output of DedupCrowd for the training purposes. DeduppNN has two features: first is that it is an online deduplication method which is essential for the purposes of customer recognition. Second is that in terms of costs it is much lower in comparison with DedupCrowd. The last function is designed for providing label to pairs when DedupNN is not sure about their label. The intuition behind this function is similar with active learning area which selects appropriate data for labeling. ICRF consists of three integrated sub-modules. The first sub-module (DedupNNSelect) focuses on the detection and resolution of duplicate records to improve the effectiveness and efficiency of customer recognition. The second sub-module determines the level of ambiguity in the recognition of an individual customer in a customer database when there are multiple records with the same name. The third sub-module, depending on the level of determined ambiguity from the second module, recommends to the CSR the series of feature-related questions that need to be asked of the customer for his/her recognition.

Keywords— Customer Recognition; Duplicate detection; Common Personal Names; Predictive Aiding System;

I. INTRODUCTION

In the past two decades, providing efficient service has become essential for organizations, especially service delivered through customer contact centers (CCs). Contact centers, as the organization’s touch point, have a considerable effect on customer experience and retention. It has been shown that 70% of all business interactions are handled in contact centers. Having a customer-focused contact center is thus crucial for each organization. A study by Bain & Company found that, for many companies, an increase of 5% in customer retention can increase profit by 25% to 95%[1]. “Every single step needs to be done well — it is not sufficient just to have a wonderful set of people in the call center, you need to sustain that experience at every point of contact and address the total customer experience (TCE) [2]”.

Contact centers’ operations are too complex. These centers require a combination of technology, human talent and task procedures in order to deliver the appropriate and efficient performance[3]. Answering the high number of calls is one of the complexities of these centers, specifically for large organizations. As an example, Amazon receives millions of e-mail messages and voice calls annually[4]. As discussed in Chapter 2, the current literature focuses more to empower CCs by providing technologies for routing calls, storing data, interactive voice response (IVR) while other aspects of CCs are not considered that assist them in making their operations easier.

The overall objective in this paper is to develop techniques that will assist CSRs in CCs to do their job efficiently. Techniques developed will assist CSRs in the dual process of resolution of duplicate profiles that may exist and in the efficient recognition of customers with a common name. Saberi et al. mentioned four main problems that affect the CSRs in CC from doing their job efficiently[5]. These problems are: a) presence of customers duplicate profiles, b) lack of customers
ID recognition system, c) high rate of CSRs turnover and d) lack of supporting systems for CSRs. The effect of these four problems on each other and on the efficient recognition of the customer is depicted in Figure 1. As this figure demonstrates, CSR turnover increases by having duplicate profiles and lack of supporting systems that will assist them to address it. Also it is evident from the figure; presence of customers duplicate profiles has a negative effect on fast and efficient customer ID recognition process. Lack of having supportive system impacts on the customer ID recognition process, which increases the CSRs turnover.

In this paper a Cleaned Interactive customer recognition framework (CICRF) is proposed to address mentioned problems. CICRF with all its subcomponents are explained as follows.

II. SOLUTION OVERVIEW

We propose Cleaned Interactive customer recognition framework (CICRF) to address mentioned problems in the first section. The proposed Cleaned Interactive customer recognition framework (CICRF) has two broad aims. (a) First is cleaning of CC’s database (DB) and (b) second is of an interactive framework for customer recognition solution for CCs. Figure 2 depicts these two aims of CICRF conceptually.

In this section, we broadly explain how these two aims are successfully achieved by CICRF thereby ultimately addressing the four identified problems.
Cleansing:

Presence of duplicate profiles in the CC’s databases is one of the key problems in the efficient operation of CCs. CICRF addresses this problem in two phases: Periodic deduplication and Online deduplication. The process of Periodic deduplication targets the whole CC’s data base while the process of deduplication on a real time basis only focuses on cleansing those tuples that are associated with the contacted customer. These associated tuples are the ones that have either a similar or the same name with the contacted customer.

Semi-Automatic deduplication module (DedupSAUT) is proposed for the purposes of the periodic deduplication and is responsible for the cleansing of CC’s DB while neural network based classifier module is proposed for the online deduplication phase. The process of Periodic deduplication targets the whole CC’s database while the process of deduplication on a real timer basis only focuses on cleansing those tuples that are associated with the contacted customer.

Fig. 3. Tow aims of CICRF Static Cleansing and Online cleansing

Figure 3 depicts, conceptually, how these two deduplication phases perform over the CC’s DB and what their differences are. This figure shows that periodic deduplication starts at $t_0$ and continues in the time periods of $t_1, t_2, \ldots, t_n, \ldots$. Also it is evident from the figure that when a customer contacts with the CC, online deduplication phase is initiated by running $DedupNN$ over the associated data similar to that of the contacted customer name. At time $t_1$ when it is a time for periodic deduplication process to run, $DedupNN$ is run but on the whole CC’s DB. It is also possible that in periodic deduplication phase, we clean CC’s DB by running either $DedupCrowd$ or combination of $DedupNN$ and $DedupCSR$ when $DedupNN$’s performance degrades.

Interactive customer recognition: An interactive customer recognition framework (ICRF) is developed to this end. ICRF relies on three integrated modules: $DedupNNSelect$, predictive aiding (PA) and decisive feature selection (DFS). As it is discussed before, $DedupNNSelect$ detects duplicate profiles and merges them into original ones before the recognition phase. Predictive aiding informs $CSR$ the level of difficulty of customer recognition and $DFS$ determines the sequence of questions that should be asked from customer for her recognition by $CSR$. The main aim of $ICRF$ is to provide efficient and
fast recognition framework which has a positive impact on both sides: CSR and customer. Recognising the customer in an efficient and accurate way increases the satisfaction which leads to a positive impact. Also this supportive environment make CSR feel better in the complex and challenging environment of CCs.

**Four identified challenges:** Figure 4 demonstrated the proposed solution, CICRF. This figure highlights the difference between two deduplication phases: periodic deduplication and online deduplication along ICRF’s targeted problems.

![Modules to address identified problems](image)

**III. CICRF**

CICRF is the framework which is proposed to address identified problems in this thesis by focusing on duplicate profiles and customer recognition issues. Figure 5 demonstrates CICRF’s two main components, namely, cleansing and ICRF, with its subcomponents. It is vital to briefly introduce all these subcomponents in this paper to assist readers in the better understanding of the proposed solution, CICRF. In Section A, I introduce the first aim of CICRF which is the cleansing phase with its subcomponents. In Section IV, I introduce the ICRF with its three subcomponents.

![CICRF and its subcomponents](image)
A. Cleansing

As I mentioned earlier, CICRF has the ability to perform deduplication in two ways: periodic deduplication and online deduplication. These two ways are designed to address the identified problems namely the presence of Dirty Data and duplicate customer profiles. This deduplication assists CCs to clean their database and is very useful in customer recognition process by making it more efficient and fast.

Semi-Automatic deduplication (DedupSAUT) and DedupNNSelect are proposed to perform the periodic and online deduplication respectively. It should be also noted that DedupNN is a part of DedupSAUT but its scope is different to DedupNNSelect which is used in online phase. When Neural based classifier is used in the online phase (DedupNNSelect) it just runs on part of CC’s DB while in periodic phase (DedupNN) it targets the whole CC’s DB.

Before explaining and discussing various components of DedupSAUT, it is suitable to clarify that why DedupSAUT is a semi-automatic approach and not a fully automated one. As the first note it should be noted that developing automatic deduplication method is an ideal solution for any cleansing project. However, having such an automatic deduplication approach, in this thesis for the CC’s DB, is hard to achieve along with maintaining a high level of accuracy due to the dynamic and complex nature of CC’s DB. One way to provide an automatic deduplication method is to build a rule based expert system or machine learning based classifier that will automatically is able to classify pairs either as duplicate pairs of non-duplicate records while these kind of classifiers usually are not perfect and suffer from classification errors (Type I & II) [6]. Hence, feeding them with human feedback is one way which increases their accuracy. Thus I increase the accuracy of deduplication task by making it semi-automatic with the higher cost in terms of time, complexity to ensure its correct answer. I use neural based classifier, DedupNN, as the core deduplication method by hybridizing it with two other main components: DedupCrowd and DedupCSR which form a semi-automated deduplication system as shown in Figure 6.

**Fig. 6. DedupSAUT: Semi Automated Deduplication**

DedupCrowd is a crowdsourcing based deduplication module [7, 8]. It is developed to provide training data for DedupNN to assist it in its classification purposes. Once DedupNN is convergence to the good results at the end of its training, then the use of DedupCrowd is not required thereby decreasing it overall cost of cleansing. Even after DedupNN training, its performance can be degraded due to the dynamic and complex nature of CC’s DB. To address this DedupCSR is developed for retraining DedupNN. In such scenarios, either DedupCSR or DedupCrowd can be used depend on the number of pair that need to be clarified. If (N) is greater that the threshold (M) that is defined to justify the use of cost required by DedupCrowd. Alternatively if N is less than M (the costs required by DedupCrowd) then DedupCSR is used. In other words, DedupCSR is used when the number of records that DedupNN is not able to classify accurately is small since it is not practical to trigger DedupCrowd for such a small number of HITs. In fact, it is not beneficial for workers to just answer small number of HITs with little compensate. It should be noted that by trigging DedupNN, DedupSAUT is able to classify records
as either duplicate or not-duplicate. Also it is important to note that DedupNN is updated by feedbacks which are provided by either DedupCrowd or DedupCSR. Figure 6 demonstrates the mentioned process of DedupSAUT through graphical chains.

1) DedupCrowd: Crowd based algorithm for deduplication

DedupCrowd is a crowd based algorithm for deduplication that leverages the power of human in the deduplication process for cleansing the CC’s database. DedupCrowd is beneficial in three ways: firstly it provides a reliable deduplication process from an accuracy perspective; secondly it provides training data for DedupNN which is used by DedupNNSelect one of ICRF’s module and finally it assists us in identifying those CSRs that can be used in DedupCSR module. As Figure 7 depicts, DedupCrowd relies on four integrated modules: Pairs identification, Workers’ error estimation, Monitoring rules and Majority of voting.

Pairs’ identification module is the first module which is initiated on the start job of DedupCrowd. This module identifies pairs (HITs) that should be posted to workers for their labelling. After posting identified pairs to the workers, Workers’ error estimation module is initiated to estimate their associated errors during the completion of HITs. The outputs of this module are fed into Monitoring rules module to determine the reliability of workers job. This module either allows workers to continue their job due to their good performance or evicts them due to the low quality job. Finally the labelling results from workers, who remained in the system, are used to determine the final label for each pair. Majority of voting scheme is used to do this end by labelling each pair to one which is used by filtered workers in their labelling.

Figure 7. DedupCrowd

Four integrated modules of DedupCrowd are explained briefly in the following sections:

a) Pairs identification module

Usually the percent of duplicate profiles in any DB is not a high number and majority of plausible pairs are very dissimilar which are not duplicates for sure[9]. Thus, it is not necessary to post all the possible pairs into crowds which are \( \frac{n(n-1)}{2} \) pairs by considering \( n \) as the number of tuples in CC’s database. It is an easy task to filter dissimilar pairs and justs post pairs into crowd that are similar. Distance metric is used to calculate this similarity and a heuristic threshold is set to identify candidate pairs, HITs, for crowd labelling.

b) Workers’ Error Estimation module

To monitor the performance of workers, estimating their error during completion of HITs is essential. This estimation is done via Workers’ Error Estimation module and its output is used by the Monitoring rules module for monitoring the performance of the workers. The hybrid gold- plurality algorithm is proposed as the worker error estimation algorithm [7].

© IEOM Society International
c) Monitoring Rules module

The Monitoring Rules module monitors the performance of workers and evicts workers with too poor performance. A multi-rule quality control (QC) system is proposed for this purpose and is designed in a way that current and past errors of workers are taken into account [10]. This system consists of four rules that are used to determine the status of workers at the end of each batch either as ‘is within control’ or ‘get suspended’.

d) Majority of voting module

Majority of voting (MV) is a handy and popular aggregation approach in the crowdsourcing which simply selects the most frequent label for the final label of each pairs [11]. I use weighted MoV (WMV) to determine the final label more accurately by considering the output of Workers’ Error Estimation module as the weights. WMV assists DedupCrowd to determine the final label of each pair by aggregating active workers labelling. As it is mentioned, DedupCrowd provides with DeduppNN training data. DeduppNN is explained briefly in the next section.

2) DedupNN

DedupNN is a classifier based on artificial neural networks (ANNs) [12, 13]. This module has the role of finding duplicate profiles and merging them into original profiles. This is an online deduplication method which is essential for customer recognition process to make it more accurate and faster. It is possible that a duplicate detection algorithm returns the wrong output. This error can be one of the following error types: false negative or the false positive error. If it is false negative, the algorithm fails to match two records, classifying a true matched pair as a non-match. If it is false positive, the algorithm classifies two distinct (different) records as a matched pair which is incorrect. As it was mentioned earlier, labelled paired records by crowd via DedupCrowd are used as the training data for ANN training purposes. If this trained neural network fails to accurately categorize some pairs into either duplicate or not-duplicate, individual feedbacks are used in this regards. These individuals can be either crowd or CSRs based on the number of these pairs. If the number of uncertain pairs is high DedupCrowd is used again while DedupCSR is utilized for the case of low number of uncertain pairs. DedupNN output is forwarded to Predictive Aiding module as the cleansed data.

3) DedupCSR: CSR based algorithm for deduplication

DedupCSR is an algorithm that use ANN as its classier is not accurate always like other classifiers [14] and retraining it with more data will improve its accuracy [15]. Thus more training data is needed which can be obtained in this case via crowd or CSRs. The selection between Crowd and CSR is done based on the number of uncertain pairs. If this number is small, CSRs feedback are used otherwise DedupCrowd is utilized again.

DedupCSR relies on two integrated modules and a decision variable. Question selection module and Compare module are two modules that construct DedupCSR. The decision variable is the number of uncertain pairs that its impact on DedupCSR is explained above. Sudden drift in pattern of duplicate profile is possible in CCs database and this is reason for observing some uncertain pairs by DedupNN. Figure 8 depicts DedupCSR flowchart by highlighting its two main modules.
When a customer contact CCs after end of his conversation, Question selection module is used to post a pair for CSR’s labelling. This labelling assists DedupCSR in calculating the ability of CSRs in deduplication process. To complete this calculation the output of CSR’s label should be compared with other workers’ labels which be done via Compare module.

Calculating CSRs’ ability in recognition of duplicate profiles as the front line staff is beneficial in some ways: Using their feedbacks for training DedupNN, using capable CSRs in future as the part of workers, putting more weight on their feedback about databases quality issues and using them for training other CSRs through a knowledge management platform. In this thesis we just demonstrate how we can calculate the ability of CSRs regarding duplicate profiles recognition and shows how this calculation is used for improving the performance of DedupNN and do not go further with this point.

a) Question Selection module

The main aim of this module is to identify which pair of records should be asked from the CSR from the given list, as they do not have enough time to answer all the HIT tasks. Integration of Heuristic probabilistic method and active learning based method is used to this end. For having the maximum impact, Whang et al.’s method has been used in which they propose a probabilistic method that estimates the accuracy of deduplication considering the label given by CSR to a pair [16]. The uncertainty sampling is utilized in active learning research [17].

b) Compare Module

In this module, the ability of CSRs in deduplication process is calculated. This calculation is done by comparing their labels with the final label of DedupCrowd in the related pairs. This model consists of two phases: labelling CSR ability and monitoring his ability. The first $m$ labels of CSR are compared with DedupCrowd and if $k$ percent of his labels similar with DedupCrowd labels, this CSR is considered as Good CSR. After this phase, his labelling is monitored during a time by using multi rule QC. It is possible that this CSR do not consider as Good CSR if his performance is considered as out of control.

IV. ICRF

This module is proposed to automatically assist a CSR in the efficient and effective recognition of individual customers in a large customer database with many ambiguous entries[18]. In fact, this module is designed to address two of identified problems: customers with common name recognition issue and complexity of CCs’ operations and lack of CSRs supporting systems. This module has two main features: fast recognition and data quality improvement and relies on three integrated modules: Duplicate detection, Predictive aiding, and Decisive Feature Selection. Figure 9 shows how these three modules working through ICRF. DedupNN is used as their duplicate detection modules since ICRF need an online duplicate detection algorithm to be able perform its job in an interactive manner. Since DedupNN is explained in Section, predictive aiding and decisive feature selection modules are explain in the following.

Fig. 9. ICRF

1) DedupNNSelect

When a customer contacts CC, the CSR should look at customer profiles (Tuples) who share the similar or same name with this customer. Detecting duplicate profiles among these tuples make the process of recognition much efficient and
faster. Therefore, a neural network based deduplication technique which is trained in DedupSAUT (DedupNN) should run over part of CC’s DB which also expedites its implementation. With the similar approach which is used in first module of DedupCrowd, just pairs forward to this classifier that are similar. Therefore just selected pairs are forwarded to the neural network based deduplication technique and this neural network based classifier called DedupNNSelect. Table I depicts the similarity and difference of DedupNN and DedupNNSelect. It shows that they use the same architecture since the trained neural network does not change while they have different domain.

<table>
<thead>
<tr>
<th>TABLE I DEDUPNN AND DEDUPNNSELECT COMPARISON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture</td>
</tr>
<tr>
<td>DedupNN</td>
</tr>
<tr>
<td>DedupSelect</td>
</tr>
</tbody>
</table>

This section reduce the input’s space and make DedupNNSelect implementation even faster which is an essential factor in customer recognition. As it is clear based on the given discussion in this paper, DedupNNSelect is trained and tested before its usage in customer recognition framework. In fact, based on machine learning terminology usage of DedupNNSelect at this stage is its recall phase.

2) Predictive Aiding (PA) Module

“One of the reasons for this high turnover rate is due to CSR’s dissatisfaction with their work environment. There can be many different factors that can lead to such dissatisfaction with the successful identification of the virtual customer being one of them”[5]. PA module is designed and developed in order to assist CSRs in better customer recognition. In fact, PA assists CSRs to predict the level of ambiguity of current tuples. In other words, when customer makes a contact with CSR, PA is able to determine the difficulty of this customer’s recognition process by providing two scalar and linguistic variables.

PA module has two roles in the proposed framework. Firstly, its scalar output is used as a decision variable as to whether to use Customer Recognition or not. If the level of ambiguity of current tuples is low, then we do not use the Customer Recognition module. Figure 2 demonstrates how the Predictive Aiding scalar output is used as the decision variable. Secondly, Predictive Aiding assists the CSR to be proactive rather than reactive in the recognition process. It is clear that CSRs are more effective at being proactive than reactive. The linguistic output of Predictive Aiding assists CSRs in predicting the customer recognition difficulty. In fact, this difficulty is provided to the CSR in the form of a fuzzy linguistic variable. This fuzzy variable has a better communication capability as stated in the natural language form.

3) Decisive Feature Selection (DFS) Module

The DFS module acts like a guideline by generating the sequence of the most decisive questions. The decisive questions expedite client recognition which improves customer experience. Three approaches are proposed to find these questions from three different schools of thought, namely statistical analysis, information retrieval and machine learning: (i) Levenshtein edit distance in combination with weights based on the Inverse Document Frequency (IDF) of terms, (ii) statistical tests based on the ANOVA[19], and (iii) decision trees based on the C4.5 algorithm. In order to determine the preferred algorithm, two criteria are proposed. It should be noted that this module receives data from the Duplicate Detection module as depicted in Figure 9 and then provides the sequence of the most decisive questions for CSRs.

V. CONCLUSION

In this paper, I proposed cleaned interactive customer recognition framework (CICRF). CICRF relies on two integrated modules: cleansing and interactive customer recognition framework (ICRF). A semi-automatic cleansing approach by leveraging crowdsourcing and neural networks was proposed. In this paper, I show how CSRs’ feedbacks are used in this semi-automatic approach. Three sub-modules also proposed that build ICRF, namely, DedupNNSelect, Predictive Aiding (PA) Module and Decisive Feature Selection (DFS) Module. Various disciplines armed proposed solution to ensure that the proposed solution is able to deliver the right solution: statistic, machine learning[20], natural language processing, statistical quality control, Fuzzy logic [21] [22] and neural networks[23].
Reference:


BIOGRAPHY

Morteza Saberi is an outstanding PhD student under the supervision of Professor E. Chang and Dr. Omar Hussain. He has published more than 140 papers in reputable academic journals and conference proceedings, of which over 15 papers related to his PhD thesis. In his current research interests, which include systems modelling, fuzzy and soft computing and data mining in unstructured context, he has published over 65 scientific papers, including 20 international Journal papers. He was a Lecturer at the Department of Industrial Engineering at University of Tafresh. He is also the recipient of the 2006-2012 Best Researcher of Young researcher Club, Islamic Azad University (Tafresh Branch). He is also the recipient of National Eminent Researcher Award among Young researcher Club, Islamic Azad University members.

Elizabeth Chang is Professor and Canberra Fellow in the School of Business, the University of New South Wales at the Australian Defence Force Academy (ADFA). Her current research focuses on Defence Logistics; Intelligent Transportation, Ambient Security, Cyber-Physical Systems and Internet-of-Things; Trust and Risks, Situation Awareness in Ad-hoc environment; Virtual Collaborative Logistics and underlying technologies including data intelligence and real time big data analytics. In combining IT and Logistics she empowers the research work and professional training in Logistics and Transporation in a way that would not otherwise be possible. She is excellent in working with public and private sector professionals, academics from many fields as well as PhD Students. She is motivated by the idea of ‘creating value by making connections’ and is looking forward to working with you locally and internationally to help build one of the
best research groups in the world in the area of Logistics and Transportation and especially for Defence, public and private sectors, please contact her at elizabeth.chang@unsw.edu.au or e.chang@adfa.edu.au.

Omar Khadeer Hussain is currently a Lecturer at the School of Business, UNSW Canberra. Prior to joining the School in February 2014, he worked as a Senior Research Fellow at Curtin University. Omar's research areas of interests are Logistics and Supply Chain Management, Distributed and Grid Systems, Decision Support, Group Support Systems and their applications to Logistics areas. Omar’s research has been published in international journals such as The Computer Journal, Journal of Intelligent Manufacturing etc. He has won university and faculty level awards from his research and as the main supervisor has supervised 5 PhD students to completion. In 2011, he was awarded with an APDI Fellowship from the ARC on a Linkage project with Prof Elizabeth Chang as the lead CI.

Zahra Saberi is a graduate student at department of industrial engineering, Tehran University. She obtained her BsC in Industrial and Systems Engineering from the sharif University of technology, Iran. She continued her study at Amirkabir university of technology in the filed of supply chain management and logistic. Her current research interests include Logistic and Supply chain management, simulation-based optimization, Data Mining, Expert System.