Proceedings of the 2016 International Conference on Industrial Engineering and Operations Management
Kuala Lumpur, Malaysia, March 8-10, 2016

In-house Crowdsourcing-Based Entity Resolution Using Argumentation

Morteza Saberi[1], Naeem Khalid Janjua[1], Elizabeth Chang[1],Omar Khadeer Hussain[1], Peiman Pazhoheshfar[2]
[1]School of Business, UNSW Canberra, Australia
[2] Young Researchers and Elite Club, Arak Branch, Islamic Azad University, Arak, Iran

***Abstract*— A conceptual framework is proposed in this study to improve Entity Resolution in contact centers. It is stated in the paper that how RFID produce dirty data in CC's databases and how using customer service representatives (CSRs) via argumentation framework deal with issue. Leveraging the power of CSRs put this work as a crowdsourcing technique that combine human and machine together to rich to the high quality of data in CC's databases.**

Keywords—Crowdsourcing; argumentation; contact centers

I. SMART CITY AND DIRTY ISSUE

The Internet of Things (IoT) is a network of physical objects, in which each object is provided with unique IP and is capable of collecting and exchanging data over the network. The applications of IoT has been extensive in various domains and in one of them have led to the creation of Smart Cities where people are able to store, share and transfer data among various smart city applications [1]. The consumer smart card is one of IoT applications which enable smart city users to access and consume various smart city resources managed by different city organizations. A city's habitants are the prospective, current or churned customers of the city's organizations. Thus provision of seamless access to the information sources for their customers is of pivotal importance for the success of smart cities. However, due to various data collection challenges (which results in duplicate, inconsistent or incorrect data storage), the organizations are currently facing a challenge of handling dirty data in an organization's database that creates barriers to the provision of seamless access to and between the information resources [2].

In this paper, we focus on smart contact centers (CCs) as the front line of organizations which provide essential services to the city's habitants. The objective of our research is to pave the way for development of smart cities where an organization's databases administrator can handle and rectify data inconsistency to ensure seamless access to information resources by smart city users. In the next section, the complex operations of CCs are discussed which leads to high turnover of their staff (CSRs).

II. CONTACT CENTERS' COMPLEX OPERATIONS

Contact centers' operations are complex and they require a combination of technology, human talent and task procedures in order to deliver the appropriate and efficient performance [15]. Answering the high number of calls is one of the complexities of these centers, specifically for large organizations. For example, Amazon receives millions of e-mail messages and voice calls annually [16]. 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 difficulty of CCs in doing their tasks and its various challenges leads to the high rate of employee attrition [17], referred to as CSR turnover [2, 18]. This high turnover is reported to be between 20 to 50 percent [19]. According to a recent report, lack of employee continuity and organizational stability, the high costs involved in the induction and training of new staff, and organizational productivity are some of the challenges that arise as a consequence of turnover [20]. It should be noted that the difficulty of CCs tasks is not the only reason for the current high rate of CSRs turnover. Some CSRs leave CCs since they do not see this job as a permanent job with various possibilities of promotion [21]. While that cannot be controlled, but if CSRs do not have a problem with the nature of their job then making their complex jobs easier and more straightforward would decrease the rate of CSRs turnover. This can be achieved by providing them with supporting decision support systems (DSS) to assist them in performing their routine tasks, supporting system that enable them to tap on customer unstructured data to have a better voice of customer etc.

The problem, which is mentioned earlier, presence of dirty data in CCs' databases, is one of the problems that have a negative effect on CSRs job. As mentioned earlier, inefficiency in dealing with CC's dirty data while making customers unhappy and angry makes CSRs stressful and dissatisfied as well [6-8]. In fact, CSRs as the front staff receive anger and dissatisfaction of customers due to having dirty data in CCs database; but they are not in charge and an fault of these kinds of CCs problem. Providing useful decision support systems (DSS) to ease the job of CSRs will in turn easier and decrease their high associated turnover in turn. It is easy to imagine that working environment for *CSR*s of a CC with high duplicate profiles is hard and stressful. By being proactive toward providing supporting systems for *CSR*s, CCs have the power to turn this stressful environment into a better one. Therefore, it is essential for *CC*s to assist their *CSR*s in performing their everyday activities. Cleansing CC's DB is a great help to *CSR* that reduce their stress at work.

© IEOM Society International

III. MOTIVATION AND RESEARCH DIRECTION

The ways by which customers communicate with their companies have changed because of constant development and adoption of information and communication technologies (ICT). While on the one hand, this eases the way by which a customer can contact an organization (for example email, online form etc.) on the other hand it creates many challenges for the companies to maintain the quality of information in CRM systems at a desired level of accuracy. This is because when customers use several communication tools it may produce dirty and noisy data which will have a negative effect on the working and performance of the CRM systems [3]. For instance, according to TDWI report, the dirty data issue costs US businesses more than $600 billion a year. One result of such noisy and dirty data is the creation of duplicate profiles which is introduced in the CRM systems. This is a critical issue as it prevents CSR's from addressing the customers' queries efficiently using CRM systems.

Emergence of smart cities is another challenge for CCs as it brings another type of threat to their database quality in the viewpoint of dirty data. The literature states that use of smart city technologies such as RFID produce massive amount of dirty data in organization database [4]. Most of the RFID devices are sensitive to environmental factors, and result in inaccurate reading of tags which is stored in the organization database [4]. RFID technology has a pivotal role to assist in the track and trace of supply chains [5]. Therefore, the storage of dirty data leads to different snow balling issues such as items intractability etc.

It has been established in the literature that the difficulty of CCs tasks and data related challenges lead to a high rate of employee attrition in an organization [6]. Making CCs to comply with the smart cities structure equipped with technologies like RFID make CCs job even harder and increase the rate of customer service representatives (CSRs) turnover. For example, a customer contacts a CSR to follow the status of his complaint in relation to the damage to his product and wants to know when he would receive it after being repairing. Having a lack of accurate information prevents the CSR to give accurate response to that customer which results in customer dissatisfaction.

In the literature, Entity resolution (ER) is used as a general term to represent methods which deal with duplicate profiles in one or two different databases [7]. In this work, the term of data matching problem (DMP) is used which is more general than ER and covers any consistency in RFID based produced data. Let us assume a customer contacts a CSR to follow his queries regarding a product's warranty. As it is mentioned earlier, if the tag reader reads incomplete electronic product code (EPC) or even reads incorrect data (which is named as Ghost data) it is possible that the CSR find this as an inconsistency. Using his knowledge through the in In-house crowdsourcing which has been studied recently and also applied in the database community [8-10] will help CSRs to address this. By considering the massive amount of data being produced with different schemas by RFID based technologies, using argumentation based system can be useful. Using collective intelligence for automatic cleansing in future via an argumentation based system is useful if the CSRs operation load in their normal duty is considered.

In this study, we adopt a hybrid approach for data matching problem (DMP) and used judgments of CSRs against the machine results and make a decision on final output of argumentation based system. The CSRs indicatess the crowd where each of them has familiarity with the targeted database and also has a high level of experience thus distinguishing them from the normal crowd. we propose an In-house Crowd DMP approach using argumentation in present study using such a crowd. We have designed and developed an algorithm that target cases identified by CSRs during their interactions with customer. The interactions can be in any forms of communication which are available to the customers. Improving the efficiency of DMP process is the objective of identifying and sharing only one DMP with CSR crowd. A CSRs needs to be more efficient in achieving DMP while carrying out other duties. To ensure this, it is needed to ask only one verification question from them to express their viewpoint about inconsistency. This question will be identified and forwarded by an active CSR who is responsible for DMP.

Example. Let us assume a company has various stores in Australia. This company makes use of RFID technology and has its own RFID configuration settings. Figure 1 shows a typical configuration of EPC.

Fig. 1. EPC Example

Michael Smith (with original profile A) contacts a CSR via e-mail to claim the warranty of his product that has an issue with the screen.

© IEOM Society International

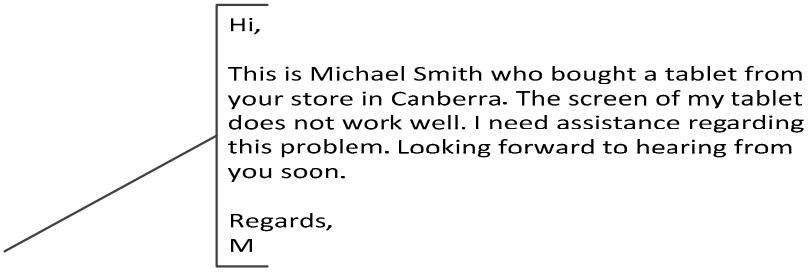

Fig. 2. Michael Smith Email Contact

Table 1 shows results to a CSRs query which has two customer profiles named Michael Smith. Table 2 shows the products bought by them and information was captured using RFID technology.

TABLE I. EXAMPLE CANDIDATE RECORD PAIRS FOR THE FIRST ARGUMENT

RecID	Name	StrName	Suburb	Date of Birth (dot)	Phone Number
A	Michael Smith	Luxton	Coculter	01041981	0471785412
B	Michael Smith	Depot	Davidson	02041981	0470125658
C	Michael Smith	Coulter	page	04051977	0482156321

TABLE II. CUSTOMER BOUGHT PRODUCTS

RecID	Name	Manufacturer	Product Type	Unique Item
A	Michael Smith	41 (Sony)	12 (LCD)	0125364
B	Michael Smith	43 (Samsung)	13 (Tablet)	9852101
C	Michael Smith	43 (Samsung)	null	9852103

A *CSR* who looks after this e-mail decide to forward his argument to another *CSR* for his/her viewpoint. Equations 2 & 3 show their argumentations that are support and attack in argumentation literature terminology. In fact, two different CSRs identify Michael Smith with two different profiles which need to be clarified.

$$\boldsymbol{Support}:$$
$$\text{SimilarCity}(Q, \text{A}), \text{Samename}(\text{A}, \text{B}) \rightarrow RecID(A)$$
$$\textbf{Attack}:$$
$$\text{SimilarProduct}(Q, \text{B}), \text{Samename}(\text{Q}, \text{B}) \rightarrow \sim RecID(A)$$
$$\sim RecID(A) \rightarrow RecID(B)$$
$$\boldsymbol{Attack\ to\ other\ two\ arguments}:$$
$$\text{ProblemReader}\left(Prod_{Type}\right) \rightarrow \sim RecID(A, B)$$
$$\sim RecID(A, B) \rightarrow RecID(C)$$

IV. RESEARCH METHODOLOGY

In present research, in-house crowd entity resolution approach using argumentation is proposed. In order to ensure that the stated objectives will be achieved, four phases are proposed as follows:

A. Phase 1 – Majority of voting Scheme

To incorporate CSR feedback to the DMP result, their aggregate opinion is essential. Since it is not practical to use many CSRs for DMP, this phase just acts as a preprocessing phase. In case there being a conflict between CSRs opinions, then argumentation-based conflict resolution algorithm is used for conflict resolution.

B. Phase 2 - Develop argumentation conflict resolution algorithm

There are number of areas where argumentation-driven reasoning helps us to recognize and resolve the contrasts among the participants. Various argumentation formalisms have been proposed in the literature: philosophical argumentation and logic-based argumentation. In this research, we will do thoroughly analyze the existing argumentation-based frameworks and will design and develop an argumentation-driven methodology for entity resolution. The selected framework allows the participants to generate their arguments from underlying evidences/observations, resolve conflicts and produce the final result. In my case, it will be either matched or unmatched. CSRs either support one of the existing arguments or attack the existing argument by producing a counter-argument. The selected argumentation-driven methodology determines the final label for the forwarded pair. This label is then forwarded to DMP algorithm for final decision and deduplication.

© IEOM Society International

C. Phase 3 - Validation of the proposed methodology

It is essential to assess the DMP performance before and after using argumentation based system. This assessment shows whether the argumentation based system feedback is useful or not. The available gold standard data for DMP is that is available in public domain and used for its evaluation and assessment.

V. Conclusion

The issue of dirty data is very important to tackle for contact centers. Relying on machine learning techniques it does not provide a reliable data cleansing mechanism. In this study, the conceptual framework was presented that consider leveraging the power of human in data cleansing along argumentation. This study can be a point of departure to incorporate various elements of CCs in data cleansing.

References

[1] R. G. Hollands, "Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?," City, vol. 12, no. 3, pp. 303-320, 2008.

[2] D. R. Moseley, "Consumer oriented smart card system and authentication techniques," Google Patents, 1993.

[3] O. K. H. Morteza Saberi, "Intelligent Online Customer Recognition Framework: Dealing with Common Personal Names," in ICIEA 2014, China, 2014.

[4] D. Xie, Y. Qin, Q. Z. Sheng, and Y. Xu, "Managing Uncertainties in RFID Applications-A Survey." pp. 220-225.

[5] R. Angeles, "RFID technologies: supply-chain applications and implementation issues," Information systems management, vol. 22, no. 1, pp. 51-65, 2005.

[6] A. R. Owens, "Exploring the benefits of contact centre offshoring: a study of trends and practices for the Australian business sector," The International Journal of Human Resource Management, vol. 25, no. 4, pp. 571-587, 2014.

[7] P. Christen, Data Matching Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection: Springer Data-Centric Systems and Applications, 2012.

[8] J. Fan, M. Lu, B. C. Ooi, W.-C. Tan, and M. Zhang, "A hybrid machine-crowdsourcing system for matching web tables." pp. 976-987.

[9] S. E. Whang, P. Lofgren, and H. Garcia-Molina, "Question selection for crowd entity resolution," Proceedings of the VLDB Endowment, vol. 6, no. 6, pp. 349-360, 2013.

[10] J. Wang, T. Kraska, M. J. Franklin, and J. Feng, "Crowder: Crowdsourcing entity resolution," Proceedings of the VLDB Endowment, vol. 5, no. 11, pp. 1483-1494, 2012.

Biography

Morteza Saberi is an outstanding PhD student in the University of New South Wales (UNSW) under the supervision of Professor E. Chang and Dr. Omar Hussain. He has published more than 120 papers in famous academic journals and conference proceedings, of which over 15 papers related to his PhD thesis. He has published over 65 scientific papers, including 20 international Journal papers in his areas of active research, which include fuzzy systms, soft computing, systems modelling and data mining in unstructured context. 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.

Naeem Khalid Janjua is a Post-Doctoral Fellow at School of Business, UNSW, Canberra. Prior to joining the school in January 2014, he worked as a Post-Doctoral Fellow at La Trobe University and as a Research Associate at Curtin University. He received MS degree in Information Technology from the National University of Science and Technology (NUST), Pakistan; and the Ph.D. degree in Information Systems from Curtin University, Perth, Australia, in 2013. He has more than 5 years of experience in information system's design and development in various business environments. He is an Associate Editor for International Journal of Intelligent systems (IJEIS) and International Journal of Computer System Science and Engineering (IJCSSE). He has published an authored book, a book chapter, and various articles in international journals and refereed conference proceedings. His areas of active research are knowledge representation and reasoning, modelling and automation of business processes, enterprise knowledge management and Big Data analytics. He works actively in the domain of making informed business intelligence through the use of Web-based intelligent decision support systems.

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; Ambient Security, Cyber-Physical Systems and Internet-of-Things; Trust and Risks, Intelligent Transportation, Situation Awareness in Ad-hoc enviornment. She is also an Associate Editor for IEEE Transactions on Industrial Electronics (since 2007) and Guest Editor on IEEE Transactions on Industrial Informatics (since 2005), Co-editor in chief for International Journal on Engineering Intelligent Systems. She is the Chair elect for IEEE IES Technical Committee on Industrial Informatics (2014-2015) where she will provide leadership in this area in the world by attracting top

© IEOM Society International

talent and researchers from around world to help define the future research directions. She is honorary member of the Australian Logistics and Supply Chain Society , a member of Council of Supply Chain Management Professionals, , Senior Member of IEEE and she was honoured to be Technical Chair or General chair for over 20 International and IEEE Conferences.

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. His current research interests are Distributed and Grid Systems, Decision Support, Logistics and Supply Chain Management, 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. He was awarded with an APDI Fellowship from the ARC on a Linkage project with Prof Elizabeth Chang as the lead CI in 2011.

Peiman Pazhoheshfar is member of Young Researchers and Elite Club of Islamic Azad University, Arak Branch in Iran. He earned his B.S in Industrial Engineering from Tafresh University and his MS in Industrial Engineering from Islamic Azad University, South Tehran Branch in Iran. His current research interests include data mining , systems modelling, fuzzy systems, econometric modelling and forecasting of supply and demand via artificial intelligence tools. He also has published more than 25 academic papers.

© IEOM Society International