Enhancing Customer Satisfaction Analysis with a Machine Learning Approach: From a Perspective of Matching Customer Comment and Agent Note

Qiang Wei
School of Economics and Management,
Tsinghua University,
Beijing 100084, China
qiangwei@tsinghua.edu.cn

Xiaowei Shi
School of Economics and Management,
Tsinghua University,
Beijing 100084, China
sxw17@mail.tsinghua.edu.cn

Quan Li
BT China Research Centre, Beijing, China
quan.li@bt.com

Guoqing Chen
School of Economics and Management,
Tsinghua University,
Beijing 100084, China.
chengq@sem.tsinghua.edu.cn.

Abstract

With the booming of UGCs, customer comments are widely utilized in analyzing customer satisfaction. However, due to the characteristics of emotional expression, ambiguous semantics and short text, sentiment analysis with customer comments is easily biased and risky. This paper introduces another important UGC, i.e., agent notes, which not only effectively complements customer comment, but delivers professional details, which may enhance customer satisfaction analysis. Moreover, detecting the mismatch on aspects between these two UGCs may further help gain in-depth customer insights. This paper proposes a machine learning based matching analysis approach, namely CAMP, by which not only the semantics and sentiment in customer comments and agent notes can be sufficiently and comprehensively investigated, but the granular and fine-grained aspects could be detected. The CAMP approach can provide practical guidance for following-up service, and the automation can help speed-up service response, which essentially improves customer satisfaction and retains customer loyalty.

1. Introduction

Customer satisfaction is deemed to be an important determinant of customer repurchase, word-of-mouth and customer loyalty, and is related to company's long-term profits [1]. Studies on customers' satisfaction contribute to the continuous improvement and innovation of products and services, thereby enhancing company impression and brand.

With the booming of social media and Web 2.0 platforms integrated into customer service systems, User-Generated Content (UGC), e.g., online reviews or comments, from customers become a new and important information source to probe or detect customers' moods, feedbacks, sentiments, key concerns, etc. Customer comment (CC) usually includes customer's preference for the service or attention to some particular aspect of the product, which may imply customer's satisfaction [2]. Though not well-structured like designed questionnaire, CC may contain more diversified and novel features from customer's viewpoint, i.e., showing potential in service improvement [3][4]. The timeliness of CC may further benefit agile service response. Therefore, incorporating CC into analyzing customer satisfaction attracts many research efforts, especially in online shopping platforms [5][6][7].

Nevertheless, analyzing customer satisfaction only with CC may be easily biased and risky. First, customers are usually emotional, especially in a negative mood, leading to biased attitude exposure [8]. Second, the non-professionalism of customers makes CC weak on content quality 9], especially in specialized or technical service context. Third, CC is usually colloquial and linguistically uncertain, i.e., easily leading to semantic ambiguity [10]. Thereafter, how to alleviate the above mentioned distortion while preserving the benefits in CC is a research challenge.

In fact, the biggest deficiency in customer comments is that, little constructive or potential solution could be distilled or inferred, due to its "strong emotion, weak professional" characteristics. For instance, in Telecom service context, a typical CC, like "*I am satisfied*" or "*Whilst talking to me he kept disappearing and putting me on hold and wasn't really listening to me*", is lack of granular information, and can hardly enlighten further technical treatment or solution.

Incorporating other valuable information source is a promising direction. But seldom does a real-world customer service system, e.g., a call center, can achieve or provide further information beyond survey and customer comments. Though it can be observed that, recently in many online platforms, the merchant's

HICSS

specific feedback to a customer's (negative) comment could be presented, which, however, only shows positive feedback in a routine format, and does not deliver informational details.

In this paper, we obtained the valuable data from a well-known UK telecom company X. In the customer service department of X, to effectively evaluate customer's satisfaction, for each incoming call, the company collected not only the traditional survey information (e.g., customer profile, fault/complaint details, service type, etc.) and the customer's Net Promoter Score (NPS) (from 0 totally dissatisfied to 10 fully satisfied) on the call afterwards, but also two types of text UGCs on the call, i.e., customer comment (CC) and agent note (AN). Agent note is provided by the well-trained employee of the call center who handles the call, which briefly summarizes the call and records key technical comments on the faults/complaints in it, and sometimes proposes alternative/potential solutions. Since AN is for internal business tracking and not open to customer, it is more informational and knowledgeable. For the example of CC "*I am satisfied*" (NPS = 7), the agent notes "*customer called to update bb (broadband) contract and buy new phone. Transfer it to sales and retention department*", showing the business is not closed yet and needs to be followed by other department. More importantly, this AN further elucidates why NPC is just 7 (i.e., since the problem is not solved) while the CC is totally positive (i.e., on the agent attitude aspect). For the second example of CC "*whilst talking to ...*" (NPS = 2) with only strong negative mood on bad service attitude, the corresponding AN "*cx (customer) called in to query latest bill. cx informed that the extra 8 pounds on her bill came from calls made outside the plan. cx wanted cheaper deal but already on 18.99, also asked if there was a cheaper for employees and was transferred to sale*" not only points out the price aspect is the main concern of the customer, but also provides some alternative suggestion for following-up.

Clearly, the add-on of AN can effectively complements CC with more granular and technical details, as well the CC complements the AN with customer's sentimental feedback, both of which enhance the customer satisfaction analysis. More importantly, since AN and CC formulate the same event from different angles and highlight different aspects of the event, comparing the consistency and inconsistency on aspects between AN and CC, i.e., conducting matching analysis, may further help gain in-depth customer insights. For the second example, the customer only highlights the poor service attitude aspect in CC, and the agent emphasizes the price aspect but ignores the service attitude intentionally or unintentionally (i.e., mismatch on aspects "price" and "attitude"), revealing the agent did not well address the attitude aspect and

problem on price aspect has not been solved in the call. The mismatch on aspects between customer and agent may significantly lead to dissatisfaction of customer (e.g., NPS = 2 of example 2). Therefore, investigating AN together with CC through matching analysis may strength customer satisfaction analysis to a large extent. Concretely, from matching analysis, the difference between the understanding of AN and CC on aspects of service, e.g., problem definition, service attitude, price, quality, solution proposal, etc., can be detected and targeted. Then further solution guidance or customized service response could be designed to focus on the mismatched aspects. For instance, of example 2, price issue and service attitude should be the key aspects in the following-up call-back.

Due to the short free-text nature of AN and CC, it is challenging to conduct matching analysis. Manual processing is neither effective nor efficient, especially in the large-scaled UGC environment and quick-response call center context. Furthermore, extracting aspects, which is normally divergent, ambiguous and sparse in UGCs, by manual annotation is not only infeasible but quite biased. Therefore, a machine learning based method will be helpful.

This paper will propose a machine learning based approach of conducting CC and AN Matching analysis and further Predicting the NPS, namely CAMP approach. In this approach, first, a Convolutional Latent Semantic Model (CLSM) is deployed to extract the latent aspects. Therefore, a pair of CC and AN can be mapped into two semantic latent vectors on the learnt dimensional space. Based on distance calculation, the (mis)matching scores can be derived. Next, the sentiment of CC can also be calculated with Bayesian classifier after feature extraction. Finally, together with the sentiment of CC and the (mis)matching scores between CC and AN, a prediction model is built to estimate the NPS value. Therefore, the higher (mis)match score on an aspect, the more dissatisfaction on the aspect, and the more improvement is required on this aspect for following-up customer service.

This paper is organized as follows. Section 2 briefly reviews the relate work. The technical details of CAMP approach are introduced in Section 3. In Section 4, an empirical study with the approach on a real-world data is extensively discussed. Section 5 concludes the paper.

2. Literature review

2.1. Customer satisfaction analysis

Customer call centers have emerged as an important platform for providing higher customer satisfaction [11]. Companies directly communicate with their customers

through call centers, which playing a crucial role in CRM [12]. In call centers, human agents or automatic transponders resolve problems/complaints from customers [13]. A customer may be satisfied after telephone if his or her problem is solved [14].

According to the expectation confirmation theory, consumer satisfaction or dissatisfaction is defined as an assessment of the difference between the pre-experience expected and perceived performance [15][16]. And related research shows that the quality of interactive services has a more important impact and is a crucial factor in determining the overall satisfaction of customers [17].

As for customer satisfaction in call centers, a study from Feinberg research [18] showed that the percentage of calls, the average speed of answer, service level, even the amount of after call work and other 13 call center operational aspects did not affect caller satisfaction. But this research was in banking/financial domain. In another Feinberg research in 16 industries [14], only two aspects affecting caller satisfaction were found significant were found significant, i.e., the percentage of calls closed on first contact and average abandonment rate. Miciak and Desmarais found that the first call resolution does not always achieve high customer satisfaction [19]. Furthermore, researchers found that most of the aspects used to effectively manage call centers are not positively correlated to customer satisfaction [14, 18-19]. The above researches reveal that key aspects impacting customer satisfaction are quite divergent, dynamically changing and context-sensitive, and need to be further explored.

Some research paid more attention to refine and construct the customer satisfaction aspect framework in customer service. Bitner and Hubbert divided satisfaction into two categories [20], i.e., attribute satisfaction (e.g., perceived satisfaction with a certain product attribute) and overall satisfaction (e.g., on environment, sales people, quality and transaction process, etc.). Westbrook argued that satisfaction with retail institutions is the accumulation of customer satisfaction for salespeople, store environments, products, and other factors [21]. Though reasonable and well-structured, the designed aspects are quite static and cannot cover the divergent and changing customers' intents. It is more noteworthy that, existing research depend heavily on the questionnaire survey from customers, which cannot evade the self-reported bias and bounded-rationality of customers. Therefore, this paper introduces another piece of valuable UGC, i.e., AN, together with CC to conduct a matching analysis, which may achieve a deep understanding of customer satisfaction from a more comprehensive and de-biased viewpoint.

2.2. Sentiment analysis on customer satisfaction

Sentiment analysis (SA) is a method for identifying the ways in which sentiment is expressed in texts and whether such expressions include positive or negative opinions on a certain product or service [22]. SA has been widely applied in various areas in real-world business, such as market and FOREX rate prediction, business analysis, attribution prediction, marketing intelligence, etc. Customer satisfaction analysis is one of the hot areas using SA [23]. Kang and Park assessed and visualized the customer satisfaction in mobile service using a combination of SA and VIKOR approaches [24]. Garcia-Cumbreras et al. projected a novel application of SA in recommender system by categorizing users according to the average polarity of their comments [25]. Miranda et al. proposed a tool for aiding the evaluation of customer satisfaction by sentiment analysis to assess the client satisfaction regarding services of online job search company [26]. Wang et al. [27] presented a maximum likelihood estimation method to determine the significant components of customer satisfaction in free-form text. Farhadloo et al. presented a method to model the aspects of overall customer satisfaction from unstructured-form written opinions using a Bayesian method [28].

However, due to the scarcity of agent/employee's UGC like AN, existing researches only focus on analyzing the customer comments and rating, which essentially limit further understanding customer satisfaction.

3. The CAMP approach

3.1. The Framework

In this section, we first introduce preliminary concepts in customer comments and agent notes then define the problem of CC and AN matching analysis. Finally, we present the framework of CAMP.

Customer comment is the customer 's evaluation of the whole process about each order, which includes product, service, personnel and other information in the form of free text. Agent note is provided by the well-trained employee of the customer service department who handles the customers' complaints, which briefly summarizes the complaints and records key technical comments on the faults/complaints. AN is also in the form of free text but is much more standardized and professional.

The research problem of matching analysis for CC and AN could be formally defined. Given a pair of CC and AN about the same order, the problem is to learn low-dimensional semantic vector representations $y_C \in$

R^{1*d} , $y_A \in R^{1*d}$ for CC and AN, respectively. And matching score vector M $\in R^{1*d}$ can be constructed, representing the deviation between the semantics of CC and AN, where i represents the number of semantic dimensions.

Note that, the output of the problem consists of two parts: (1) matching score vector $M = \{m_1, m_2 \dots m_d\}$, where m_i represents semantic gap between CC and AN on the aspect i. The larger the value, the more mismatched it is. (2). A predicted score NPS for user satisfaction. Based on the above two results, the key aspects (i.e., mismatched) leading to customer dissatisfaction can be further detected.

Basically, matching analysis is to compare the similarity of CC and AN in each aspect. However, aspects should be extracted and determined first. Due to the latent aspects and free-text nature of CC and AN, only lexical matching can be inaccurate. So, CC and AN in form of free-text need to be processed and compared on the semantic level. Typically, some probabilistic topic models such as probabilistic LSA (PLSA) and Latent Dirichlet Allocation (LDA) are often used for semantic similarity matching. Nevertheless, these models are often trained in an unsupervised manner using an objective function that is only loosely coupled with the evaluation metric for the retrieval task. Thus, their performance on matching task on two differently expressing texts, e.g., like CC and AN in this context, is not as good as expectation [29]. Therefore, this paper

refers to the technical treatment in [30], i.e., a deep learning method called CLSM, to map text into potential semantic space, and calculates similarity for each learnt latent aspect.

Centered on the CLSM method, the CAMP approach includes three components: sentiment analysis with CC, CLSM based matching analysis of CC and AN, and a prediction model for NPS. Figure 1 shows the overall framework of the approach. The input of the model is CC and AN, and the output contains not only the predicted NPS of customer satisfaction, but the matching scores on granular aspects of CC and AN vector pair. Concretely, the sentiment analysis part first divides CC into positive/negative classes and extracts feature based on the information score. After training the feature vector by Bayesian classifier, the sentiment score of CC can be calculated. The CLSM part first uses word hashing to map each pair of sentences from CC and AN into a pair of vector spaces. Next, for each vector of sentence, a convolutional neural network is constructed to find global context features and reduce the features to get a low dimension semantic vector. When the low dimension semantic vector pair is derived, the matching score for each extracted aspect can be calculated by distance. Finally, in the prediction model part, both matching scores and sentiment score are merged into a model to predict the fine-grained customer satisfaction NPS.
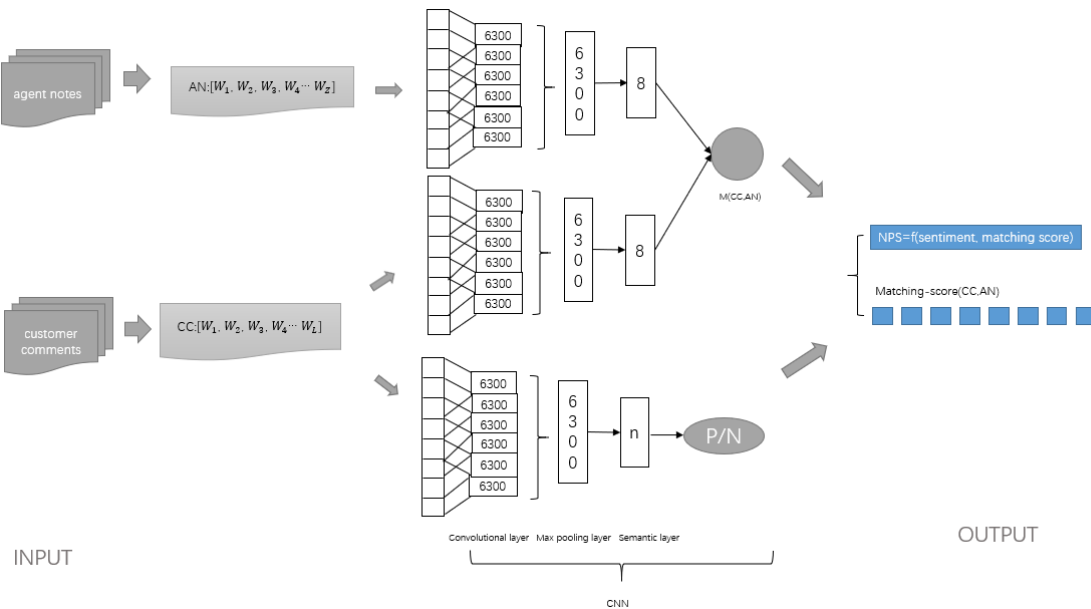

Figure 1. The framework of CAMP

3.2. Bayesian sentiment classification

Due to the ambiguity and divergence on sentiment in the text of CC, it is hard to define a corpus containing overt emotional words with sentiment tendency. Thus, the dictionary-based sentiment analysis is not well-suited. A machine learning based Bayesian classifier is adopted, i.e., a set of CC expressing positive emotion and a set of CC with negative emotion are selected as training sets, based on which, a sentiment classifier can be built. With the classifier, all CCs can be classified into positive or negative categories.

Technically, for CCs, feature selection should be conducted at the word level. Firstly, all of words and Bigrams are chosen as initiative features. Next, each feature's positive information score and negative information score can be calculated, and information gain of each word/Bigram is used to reduce feature dimension. In order to retain the sufficient information, while ensuring the positive one can be better distinguished from the negative, a metric showed in formula (1) is constructed to evaluate features. After that, the feature vectorization of each customer comment is brought into a Bayesian classifier for training and the sentiment score of each CC can be obtained.

$$(positive_{score} + negative_{score}) * (positive_{score} - negative_{score}) \tag{1}$$

3.3. Letter-trigram based Word-n-gram Representation

Before processing in CLSM based matching, the sentences of both CC and AN need to be pre-processed to map into a vector space. Traditionally, each word can be directly represented by a hot word vector based on bag-of-words method. The dimension of the vector is the natural size of the vocabulary. However, though sparsity of words, the vocabularies sizes are often very large in real CCs, bag-of-words vector representations will be quite exhaustive for model learning. Therefore, we adopt a technique called word hashing proposed by Huang [29] to reduce the dimensionality of the bag-of-words term vectors.

Moreover, word hashing is able to map the morphological variations of the same word to the points that are close to each other in the letter n-gram space. Even a word is unseen in the training set, it can to some extent be processed in word-based representation.

By word-n-gram method, a word is represented by a letter-trigram vector. However, the mapped words so far are still independent, which are lack of semantic information. Then, the CLSM method is deployed, based on a CNN, to learn the linkages between words,

i.e., discovering semantics. In the constructed CNN, sliding windows are used to reserve semantic information.

3.4. CLSM-based matching

In the CLSM-based matching process, a typical 3-layer Convolutional Neural Network (CNN) is built. At the convolutional layer, if words within each CC or AN sentence are semantically similar, they will be projected into vectors that are close to each other. The output of the convolutional layer is a variable length sequence of feature vectors, whose length is proportional to the length of the input word sequence. A complete word window can be formed anywhere in the sentence. For example, if the size of sliding window is three, the first three words of sentence "fttp fiber repair fva broadband" can be extracted based on the window (e.g. fttp fiber repair), then mapping every three words into a m-dimensional vector space by letter-trigram. Next, the three vectors should be concatenated and finally mapped into a 3*m-dimensional vector space. Figure 2 shows a convolutional layer using a 3-word contextual window.
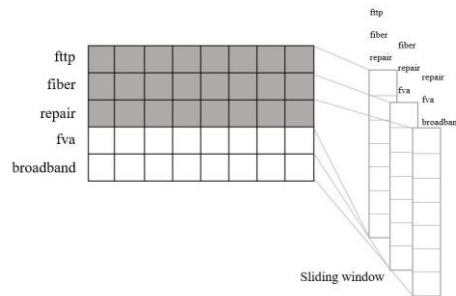

Figure 2. The Word-trigram

A sequence of local contextual feature vectors is extracted at the convolutional layer. These local features need to be aggregated to obtain a sentence-level feature vector. In CCs or ANs, many words do not have significant influence on the semantics of the sentence, so we want to suppress the non-significant local features. To do it, a max operation can be applied, also known as max pooling, to force the network to retain only the most useful local features produced by the convolutional layer. So, the highest neuron activation value will be selected across all local word n-gram feature vectors at each dimension.

With the derived sentence-level feature vector, one more non-linear transformation layer is applied to extract the high-level semantic representation. Semantic layer can be used to reduce the dimension, because, generally, lower dimension vector is of better interpretability.

Based on the obtained vector, cosine distance is chosen to calculate the matching score, as showed in (2), where y_C and y_A represent the vectors of CC and AN, respectively. At the training stage, the goal is to minimize the gap between $D(C, A)$ and true matching score (3). The residual will propagate backward in the CNN, and converge the model by stochastic gradient descent (SGD). Finally, the coefficients of each layer and the low dimension semantic vector can be obtained.

$$D(C,A) = \frac{y_C^T y_A}{\|y_C\|\|y_A\|}, y_C \in R^{1*d}, y_A \in R^{1*d} \quad (2)$$

$$MinL = \sum_{[c,a] \in CA} [(1 - NPS_{ca}) - |\widehat{D}_{ca}|]^2 +$$
$$\lambda \sum_{c \in C, a \in A} (\|y_c\|_2 + \|y_a\|_2) \quad (3)$$

$$m_i = |v_i - f_i|, m_i \subset M \in R^{1*d} \quad (4)$$

In the testing stage, the vector pairs (trained from each CC and AN pairs) are coupled to calculate the matching score on each aspect, as shown in (4). The appropriate dimension size should be chosen based on balance between accuracy and interpretability.

After the CLSM processing, not only the latent key aspects can be extracted, but the matching score on each aspect is derived, which may provide granular and customized suggestions for following-up service. This is, higher matching score of an aspect represents more dissatisfaction on the aspect. Further analysis like decoding treatment can be used to interpret the meaning of the latent aspect to help service improvement.

3.5. Predicting NPS

When the matching scores in every aspect have been obtained, together with the sentiment score, a prediction model, e.g., a regression model, can be constructed. The IVs (Independent Variables) are matching scores in every aspect and sentiment score. The DV (Dependent Variable) is NPS value, representing the degree of customer satisfaction. Compared with traditional sentiment analysis, the CAMP approach can not only detect whether customer is satisfied or not, but provide a fine-grained NPS value (e.g., in the range from 0 to 10), showing better predictive power.

4. Empirical analysis on real data

4.1. Data Description

Company X's business mainly includes local, national/international telecommunications, high-quality broadband, Internet products/services, and relative IT solutions. The data comes from X's customer service department. Due to the privacy and legal issues, 512 examples are given in total. Each entry of data consists of three parts: structured data (e.g., customer profile, service description, fault type, time, etc.), unstructured UGCs (i.e., CC and AN), as well as NPS ratings.

4.2. Pre-test with structured data

In order to set up a benchmark for following analysis, a classical regression analysis on structured data w.r.t. NPS is conducted to show whether the structured data imply valuable pattern.

The structured data includes totally 60 items, e.g., the basic profile of customers, the type of service ordered by the customer and related parameters, the duration of the current problem, and NPS, etc. Before statistical regression, the preprocessing is conducted, e.g., filling missing values, transforming qualitative data into categorical, as well as normalization, etc.

Due to the sequential nature of the Variables, sequenced regression is used to fit the correlation between the variables and the NPS. Due to the space limitation, only the statistically significant variables (confidence level > 90%) are listed in Table 1.

Table 1. Significant variables of regression (confidence > 90%)

variable	Meaning	Sig.
X_SPORT	Install sport channel or not	*
BB_LENGTH	How long is the bb service	*
BB_FIBER	Have fiber or not	*
BB_ACCL	Use an accelerator on the copper line or not	*
HUB_TYPE	Hub new or old (the bigger the newer)	**
MARKET_3	Market competition	*
TOTAL_CONTACTS	how many times did communicate with the customer	*
FAULT_LONGEST	how long the last fault was active	*
FAULTS	how many of this type of fault recorded	***

After consulting the experts in X, the above significant variables do be key aspects impacting customer satisfaction, but they are too macro and general. More importantly, they are quite "common sense", i.e., lack of novelty and practical guidance. Furthermore, among the 60 variables, 51 are not significant and R-square is only 0.25, revealing there is a large portion of customer satisfaction or dissatisfaction cannot be explained by the structural indicators. This calls in-depth analysis by incorporating other valuable information source like CC and AN.

4.3. Sentiment classification with CC

Together with CCs, the NPS is deemed as a natural emotional tag. Generally, the NPS no less than 5 is

regarded as positive, otherwise negative. With the constructed Bayesian classifier, the CCs can be trained and classified. We implemented the classifier in PYTHON. Concretely, a chi-squared test is used to help select key features. And total informative scores can be calculated based on the positive/negative information score metric (i.e., (1)). If all words are selected as features without screening, the derived accuracy is 0.48, and F1 value on negative is 0.33, F1 on positive is 0.58, where the overall F1 is 0.55. According to the customer satisfaction context, the identification of negative comments needs to emphasize. Therefore, the F1 on negative comments is selected as key metric, i.e., the higher the better. After tuning with different feature dimensions, the optimal dimension size is selected as 200, i.e., with the highest negative F1 value (Figure 3). Moreover, when dimension size > 200, the negative F1 keeps stable. Then 200 is used for the following analysis.
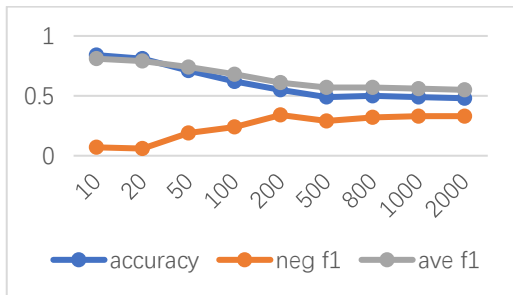

Figure 3. Different dimension of features

It is worthy noted that, CC-based sentiment analysis is easily distorted by customer's ambiguous semantic and biased emotional exposure. In addition, the customer's unprofessionalism bounds the information content in CC. Agents, on the other hand, follow the instructions to reply and are trained to take notes accordingly. Therefore, AN is quite significant to complement CC. However, conducting sentiment analysis on AN is meaningless in the call center context, since agents are required to be always calm and professional, i.e., emotionless or consistently weak-positive. Thus, conducting matching analysis on CC and AN from a semantic learning perspective will be more valuable.

4.4. Matching analysis and NPS prediction

Since the data size is 512, which is far less than the typical size for deploying deep learning analysis, it significantly affects the learning performance. Even so, the results obtained from CAMP are still better than those of only sentiment analysis, and have more explanatory power in granularity.

First, the data need to be pre-processed. For text vector learning, the range of matching score is [0,1], so it is necessary to normalize NPS to [0,1]. Given the word hashing method, a word-3-gram is represented by concatenating the letter-trigram vectors of each word. In this analysis, after removing the lower-cased letters and punctuations, about 2,100 unique letter-trigrams are observed in the training set. After the letter-trigram, every sentence can be represented as a sparse matrix.
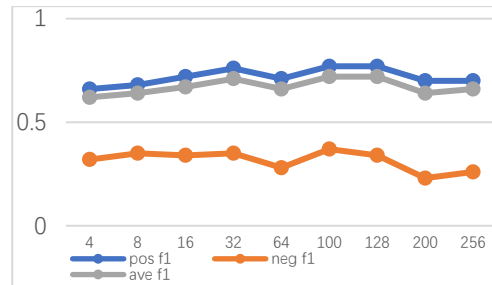

Figure 4. Semantic vector dimensions

Next, the hashed sentence matrix can be taken into CNN and the final low-dimensional latent semantic vector can be generated through learning from the backward propagation of residual. The appropriate dimension of the final semantic vector can be determined by comprehensively considering higher negative F1 value (i.e., better performance) and smaller dimension size (i.e., better interpretability). As shown in Figure 4, the appropriate dimension is selected as 8, indicating 8 key aspects are learned, which is used in the following analysis.

Based on the semantic vector, the matching score on each aspect can be calculated as well. The bigger the (mis)matching score on the aspect, and the more dissatisfaction on the aspect of the customer.

Table 2. Statistical analysis

NPS	Total	0 mismatch	1 mismatch	2 mismatches	3 mismatches	4 mismatches	5 mismatches	6 mismatches	7 mismatches	Average
[7-9]	32	4	14	8	5	1	0	0	0	1.53
[4-6]	98	0	8	25	38	24	4	0	0	2.90
[0-3]	37	0	0	0	5	6	17	4	5	4.95

Table 2 presents the statistical results on the number of mismatching aspects under different NPS levels. For easy understanding, we treat the match score >= 0.6 as significant mismatching. Moreover, since the data is too small-sized and sparse, we categorize the test data into three groups by their NPS values, i.e., [0-3], [4-6], and [7-9]. On a whole, it could be found that, with decline of customer satisfaction (lower NPS), the distribution of

matching scores move to high partition, indicating more mismatching aspects. Besides, the average number of mismatching aspects is larger when NPS is lower. This further verifies that more mismatching aspects between customer and agent on reported fault/complaint may lead to stronger customer dissatisfaction.

Furthermore, based on derived sentiment score and matching scores, a regression model is built to predict the NPS value. First of all, the values of sentiment and the 8 matching aspects, as IVs, should be standardized into range [0, 1]. The learnt regression model is as shown in (5), with R-square = 0.55, which is a significant improvement than that of the pre-test on structured data (i.e., R-square = 0.25). Moreover, the regression results show that NPS is negatively correlated with each matching score (i.e., larger mismatch, smaller NPS), and positively correlated with sentiment score. This is in line with the expectations.

$$NPS = -0.278D_1 - 0.883D_2 - 2.651D_3 - 2.705D_4 - 1.930D_5 - 4.279D_6 - 1.024D_7 - 0.400D_8 + 2.792\text{sentiment} + 11.73 \qquad (5)$$

4.5. Analysis on matching scores

In addition, the CAMP approach can output the specific mismatched aspects along with the (mis)matching scores, which can assist to target detailed problem aspects timely and accurately.

Since the learnt 8 aspects are latent, further decoding treatment should be conducted for interpretation. Unfortunately, decoding treatment needs more data, which cannot be satisfied by on-hand data. Nevertheless, by investigating the original data together with the results manually, some implications could be derived.

Table 3 lists some representative examples. Some derivatives could be obtained. First, for high NPS customer, the matching scores are usually low, i.e., implying less dissatisfaction on aspects, while the NPS is lower with the increase of matching scores, which is consistent to that of Table 2.

Second, on the low customer satisfaction like the 5th and 6th examples, both customer and agent may provide long UGCs for complaints and defenses, respectively. Therefore, these cases provide fruitful information context for matching analysis and show the benefit of CAMP approach. In the 5th example, CC and AN have strong inconsistency on 5 major aspects, i.e., product quality, product performance, agent professionalism, service attitude and overall services efficiency. These are reflected in the 5 large matching scores, which provide an explicit evidence and granular direction for further improvement. Example 6 contains three inconsistency, i.e., product quality, problem-solving

efficiency, and agent professionalism, reflected in large matching scores in the corresponding dimensions.

Third, from the 1st and 2nd examples, the sentiment analysis output opposite sentiments to true NPS values, which shows the ineffectiveness of sentiment analysis only with CC. For the 1st example, the sentiment analysis on such a short sentence may be easily manipulated by negative keyword "problem". For the 2nd example, the fragmentation and mixed semantics in CC make it hard to justify the true sentiment. However, with the complement of AN, the deficiency can to a large extent be conquered with the CAMP approach, e.g., more accurate predicted NPS values. That is to say, the inconsistency between real customer satisfaction and sentiment prediction can be corrected well.

Fourth, on high customer satisfaction, usually CC is short. However, less details of customer's latent concerns can be detected only with CC, especially in the case that CC is totally positive while NPS is not very high, e.g., the 3rd example (NPS = 8). This example further asserts that AN does provide more informational explanation. With CAMP, it could be analyzed that two aspects have larger (mis)matching values (> 0.6). After manually decoding with CC and AN, two key aspects, i.e., product quality and agent professionalism, could be targeted as the mismatched aspects. From the combined semantics of CC and AN, it can be concluded that, the quality aspect is key concern of the customer. But the agent solves the problem only by suggesting restart and does not find the underlying quality problem, which is more-or-less about the professionalism aspect. The 4th Example (NPS = 9) reflects similar phenomenon. Although the short CC is absolutely positive, NPS is not perfect 10, which could be reflected by one large matching score, i.e., one not-so-satisfactory aspect. After checking the semantics in CC and AN, the aspect could be the inconvenience caused by replacing box. Therefore, by conducting CAMP match analysis on both CC and AN, more insightful result can be distilled.

Finally, compared with the sentiment score, the NPS values predicted with CAMP are more accurate and more fine-grained than just positive or negative.

By validating the CAMP approach, it can be found that matching CC and AN can credibly enhance and refine customer satisfaction analysis, further providing constructive and customized service suggestions.

5. Conclusion

This paper proposes a machine learning based approach, i.e., CAMP, to refine the customer satisfaction analysis through matching analysis of two type of UGCs, i.e., customer comments and agent notes. With the CAMP approach, not only the semantics and sentiment in CC and AN can be sufficiently and

comprehensively investigated to help further gain customer satisfaction insights, but also the granular and fine-grained aspects can be detected with the matching analysis. This may benefit customer satisfaction analysis on two folds. First, it provides more insights to understand the reasons for customer dissatisfaction, which in turn help employers train the agents to improve service such as speeding up responses, providing more practical and customized solutions for customer complaints. Second, the automation of matching analysis can help speed up service response, which

essentially helps improve customer service and retain customer loyalty.

The limitation of the study could be found in the following aspects: (1) The data size is too small due to the privacy and legal issues, which significantly impacts the performance of the analysis. (2) Since CAMP can only extract latent aspects from machine learning process, decoding treatment is very important for interpretation of the results. Future study will be conducted on these two directions.

Table 3. Examples of result

#	CC	AN	Matching Score	Sentiment Score	True NPS	Predicted NPS
1	*because person extremely helpful in trying to resolve my problem, she kept me informed of the progress and called when scheduled to do so*	*see open complaint*	[0.11, 0.48, 0.12, 0.30, 0.44, 0.28, 0.38, 0.49]	negative	9	7.50; positive
2	*the promised callback did not happen* *I phoned on day as the system was not working* *was told that my request would be listened to this operator* *our phone stopped accepting* *I have had 2 further conversations to rectify this problem*	*order placed* *product not working contact manage*	[0.48, 0.19, **0.81**, **0.60**, 0.56, **0.75**, 0.32, 0.37]	positive	2	4.70; negative
3	*was very efficient in dealing with my problem*	*ex called to report a BB fault; he has been having dropping connection for a few weeks but current has not BB connection at all. We haven't found a network problem, Home checks done and connection refreshed. Customer has restarted his hub and has BB connection. {{MR_M}} will monitor his connection*	[0.12, 0.01, 0.09, 0.59, **0.85**, **0.63**, 0.51, 0.47]	positive	8	7.57; positive
4	*PERSON was friendly, helpful and efficient.*	*Spoken with {{PERSON}}, opened the complaint, she has got screen got stuck on loading, hence box replaced hence closing the complaint with his*	[0.31, 0.13, 0.47, 0.48, 0.24, 0.31, 0.35, **0.64**]	positive	9	9.40; positive
5	*lack of knowledge* *subject on hold far too long without explanation* *40 min call did not listen so had to repeat basic steps no resolution other than wife calling is hit and miss* *maybe uninstall and reinstall my app no escalation after no resolution*	*WIFI 4g calling validated name* *saying on mobile management and resolve mobile not compatible* *its galaxy s9 no fault* *his wife calling is on software update completed switched he says has done everything* *its hit miss sometimes it works and other times* *tried to explain no fault showing incompatible* *however again wife calling does not always work* *impatient and not too helpful rude, wasted 40mins for the above*	[**0.82**, 0.33, 0.42, 0.57, **0.90**, **0.76**, **0.61**, **0.61**]	negative	1	2.68; negative
6	*not happy with being without service from 15:09 until 01:10 having no landline for 17 days*	*customer called to follow up on ll fault as she wanted, have advised will not be compensating her requested to speak to manager got manager as requested*	[**0.62**, 0.02, 0.36, 0.54, **0.92**, **0.96**, 0.27, 0.51]	negative	3	2.78; negative

6. Acknowledgement

7. References

[1] R. N. Bolton, P. K. Kannan, M. D. Bramlett, "Implications of loyalty program membership and service experiences for customer retention and value". Journal of the academy of marketing science, 28(1), 2000, pp. 95-108.

[2] K. Zhang, R. Narayanan, A.N. Choudhary, "Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking", WOSN, October 2010, pp. 11-11.

[3] E.W. Anderson, F.D.R. Lehmann, "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden", Journal of Marketing, 58(3), 1994, pp. 53-66.

[4] T. Song, J. Huang, Y. Tan, et al., "Using User-and Marketer-Generated Content for Box Office Revenue Prediction: Differences Between Microblogging and Third-Party Platforms", Information Systems Research, 30(1), 2019, pp. 191–203.

[5] H. Onishi, P. Manchanda, "Marketing activity, blogging and sales", International Journal of Research in Marketing, 29(3), 2012, pp. 221-234.

[6] B. Gu, J. Park, P. Konana, "Research note—the impact of external word-of-mouth sources on retailer sales of high-involvement products", Information Systems Research, 23(1), 2011, pp. 182-196.

[7] D. Kang, Y. Park, "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach", Expert Systems with Applications, 41(4), 2014, pp. 1041-1050.

[8] R. Lukyanenko, J. Parsons, "Information Quality Research Challenge: Adapting Information Quality Principles to User-Generated Content", Journal of Data and Information Quality, 6(1), 2015, PP. 1-3.

[9] R. Lukyanenko, J. Parsons, Y.F. Wiersma, "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content", INFORMS, 2014.

[10] D. Coleman, Y. Georgiadou, Y. Labonte, "Volunteered Geographic Information: The nature and motivation of producers.", Int. J. Spat. Data Infrastruct. Res, 4, 2009, pp. 332–358.

[11] J. Anton, J. Monger, D.S. Perkins, "Callcenter Management: By the Numbers", Purdue University Press, 1997.

[12] A.K. Jaiswal, "Customer satisfaction and service quality measurement in Indian call centres", Managing Service Quality: An International Journal, 18(4), 2008, pp. 405-416.

[13] B.K. Moon, J.K. Lee, K.J. Lee, "A next generation multimedia call center for internet commerce: IMC", Journal of organizational computing and electronic commerce, 10(4), 2000, pp. 227-240.

[14] R.A. Feinberg, I.S. Kim, L. Hokama, et al., "Operational determinants of caller satisfaction in the call center", International Journal of Service Industry Management, 11(2), 2000, pp. 131-141.

[15] R.L. Oliver, "Effect of Expectation and Disconfirmation on Postexposure Product Evaluations - an Alternative Interpretation", Journal of Applied Psychology, 62(4), 1977, pp. 480.

[16] R.L. Day, "Modeling choices among alternative responses to dissatisfaction", ACR North American Advances, 1984.

[17] J. Ganesh, M.J. Arnold, K.E. Reynolds, "Understanding the customer base of service providers: an examination of the differences between switchers and stayers", Journal of marketing, 64(3), 2000, pp. 65-87.

[18] R.A. Feinberg, L. Hokama, R. Kadam, et al., "Operational determinants of caller satisfaction in the banking/financial services call center", International Journal of Bank Marketing, 20(4), 2002, pp. 174-180.

[19] A. Miciak, M. Desmarai, "Benchmarking service quality performance at business-to-business and business-to-consumer call centers", Journal of Business & Industrial Marketing, 16(5), 2001, pp. 340-353.

[20] M.J. Bitner, A.R. Hubbert, "Encounter satisfaction versus overall satisfaction versus quality", Service quality: New directions in theory and practice, 34(2), 1994, pp. 72-94.

[21] R.A. Westbrook, "Sources of consumer satisfaction with retail outlets", Journal of retailing, 57(3), 1981, pp. 68-85.

[22] T. Nasukawa, J. Yi, "Sentiment analysis: Capturing favorability using natural language processing", Proceedings of the 2nd international conference on Knowledge capture. ACM, October 23-25, 2003, pp. 70-77.

[23] K. Ravi, V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications". Elsevier Science Publishers, 89, 2015, pp. 14-46.

[24] D. Kang, Y. Park, "Based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach", Expert Systems with Applications, 41(4), 2014, pp. 1041-1050.

[25] M.Á. García-Cumbreras, "A. Montejo-Ráez, M.C. Díaz-Galiano, Pessimists and optimists: improving collaborative filtering through sentiment analysis", Expert sSyst., 40, 2013, pp. 6758–6765.

[26] M. D. Miranda, Renato José Sassi, "Using Sentiment Analysis to Assess Customer Satisfaction in an Online Job Search Company", Business Information Systems Workshops. Springer International Publishing, 01, 2014, pp. 17-27.

[27] H. Wang, Y. Lu, C. Zhai, "Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach", Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, ACM, July 25-28, 2010, pp. 783-792.

[28] M. Farhadloo, R.A. Patterson, E. Rolland, "Modeling Customer Satisfaction from Unstructured Data Using A Bayesian Approach", Decision Support Systems, 90, 2016, pp. 1-11.

[29] P.S. Huang, X. He, J. Gao, et al, "Learning deep structured semantic models for web search using clickthrough data", Proceedings of the 22nd ACM international conference on Information & Knowledge Management. ACM, October 27 - November 01, 2013, pp. 2333-2338.

[30] Y. Shen, X. He, J. Gao, et al., "A latent semantic model with convolutional-pooling structure for information retrieval", Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, November 03-07, 2014, pp. 101-110.