



The Research and Construction of Complaint Orders Classification Corpus in Mobile Customer Service

Junli Xu^(✉), Jiangjiang Zhao, Ning Zhao, Chao Xue, Linbo Fan,
Zechuan Qi, and Qiang Wei

IT System Department, China Mobile Online Services Company Limited,
Zhengzhou, Henan, China

{xujunli_zb, zhaojiangjiang, zhaoning_zb, xuechao,
fanlinbo, qizechuan, weiqiang}@coms.chinamobile.com

Abstract. Complaint orders in mobile customer service are the records of complaint description, which professional knowledge and information on customer's complaint intention are kept. Complaint orders classification is important and necessary to be established and completed for further mining, analysis and improve the quality of customer service. Constructed corpus is the basis of research. The lack of complaint orders classification corpus (COCC) in mobile customer service has limited the research of complaint orders classification. This paper first employs K-means algorithm and professional knowledge to determine complaint orders classification labels. Then we craft the annotation rules for complaint orders, and then construct complaint orders classification corpus. The corpus consists of 130044 complaint orders annotated. Finally, we statistically analyze the corpus constructed, and the agreement of each question class reaches over 91%. It indicates that the corpus constructed could provide a great support for complaint orders classification and specialized analysis.

Keywords: Mobile customer service · Complaint orders classification corpus
K-means · Annotation rules

1 Introduction

Complaint orders are the description of customer's complaint and recorded by customer service staff, in order to do the mining work and analyze the complaint intention in future. Identifying categories of complaint orders is called complaint orders classification. Research on complaint orders classification in mobile customer service and capture the semantic information is significant for capturing customer's intention and specialized analysis. Therefore, it is necessary to construct complaint orders classification corpus (COCC) in mobile customer service.

Current modes of constructed corpus contain expert tagging, crowdsourcing tagging (such as Amazon Mechanical Turk¹, Crowd Flower²), group tagging. Considering

¹ Available at <https://www.mturk.com/>.

² Available at <https://www.crowdfunder.com>.

accuracy, reliability and economy, this paper focuses on constructing COCC in mobile customer service based on group tagging. An annotated complaint order is taken from COCC.

Complaint order 1:

*****客户反映，本机上开通两个校讯通，收取3个校讯通功能费，客户要求查明原因并回复说明，请处理。 费用问题

(*****customer complains that two school newsletters have been opened, but functional fees for three school newsletters are charged. The customer requires to find out the cause and responds to the explanation, please handle it. cost question).

In complaint order 1, complaint order is separated into two parts by tabulator key, which are the content of complaint order and question category.

Recently, the corpus in public are appearing in dialogue [1], micro-blog [2–4], linguistic [5, 6], medical [7], etc. However, there is no public research on COCC in mobile customer service. The lack of COCC has limited the research of complaint orders classification.

This problem motivates us to construct COCC. Firstly, based on K-means algorithm and business knowledge, we divide complaint orders into 8 categories: marketing activities question, unknowing customization question, information security question, service question, cost question, business processing question, network question, and business use question. Secondly, business experts craft detailed annotation rules. Each complaint order is annotated by two skilled annotators respectively. Then, we repeatedly discuss and revise about bifurcation points with annotators, business experts without participation in annotating the corpus, aiming at determining inconsistent labels of complaint orders, then we construct the COCC. Finally, we statistically analyze the constructed corpus. As we know, COCC is the first publicly available and large dataset in mobile customer service field so far, which could provide a great support for classifying complaint orders, mining semantic information and conduct specialized analysis.

2 Related Work

Researches on corpus constructed have attracted considerable attention [8–12]. In customer service field, Yin et al. [13] identify 8 new metrics, named as tones, to describe emotional information by collecting and labeling online conversations of customer service on Twitter. They solve the problem that conventional metrics do not fit emotion analysis of online customer service conversation. Quan and Ren [14] introduce a blog emotion corpus for Chinese emotional expression analysis, which contains manual annotation of eight emotional categories (expect, joy, love, surprise, anxiety, sorrow, angry and hate), emotion intensity, etc. Chen and Nie [15] describe a parallel text mining system that finds parallel texts automatically on the Web, and generate Chinese-English parallel corpus for training a probabilistic translation model which translates queries for Chinese-English cross-language information retrieval. Feng et al. [16] propose Uyghur emotional words corpus construction based on CRFs.

Yang et al. [17] construct the corpus for named entities and entity relations on Chinese Electronic Medical Records, which comprises 992 medical text documents, and inter-annotator agreement of named entity annotations and entity relation annotations attain 0.922 and 0.895, respectively. However, due to the domain specificity, it is very difficult to be applied to the mobile customer service field. Moreover, there is no COCC in mobile customer service. In accordance with the characteristics of customer's complaint intention, we select data and construct COCC.

We select data from June, 2016 to May, 2017 comes from China Mobile Client Service System of 31 provinces. 60 million data are randomly selected, and the number of complaint orders from each province is roughly the same. To ensure privacy, we replace mobile phone number, address and other information in the complaint order with “*****”. Finally, we use 130044 complaint orders with more standardized format to construct COCC. Compared with other corpus, COCC takes full consideration of the structure, grammar and expression of complaint orders, such as regionalism, diversity and wide coverage, etc.

3 The Research and Construction of Complaint Orders Classification Corpus (COCC)

3.1 Classified Labels for Determining of Complaint Orders Based on K-Means and Professional Knowledge

The process of combining K-means and professional knowledge to determine classification labels for complaint orders is shown in Fig. 1. This process contains 5 phases: preprocessing, features extract, K-means cluster, classified label numbers for determining, classified label names for determining.

- (1) In preprocessing phase, firstly, we process desensitization of mobile phone number, address and other information in the complaint orders. Then, the corpus is preprocessed with LTP segmentation toolkit³. Finally, we use Word2Vec toolkit⁴ to train word embedding (the dimension of word embedding is 100) on the complaint orders corpus.
- (2) In features extract phase, we statistically analyze the data and extract two features for comparison: (a) n-gram features (including unigram, bigram, trigram), we extract n-gram features and use tfidf for sentence representations. (b) sentence representations feature, we adopt the way of each word embedding summation in the order to obtain sentence representations.
- (3) In K-means cluster phase, we use K-means algorithm ($5 \leq K \leq 12$) to cluster complaint orders based on unigram, bigram, trigram and sentence representations features respectively. From the cluster results, we can see that: (a) the effect based on unigram feature is not ideal, which may due to that unigram features do not utilize contextual information. (b) we vary K from 5 to 12 with an interval of 1, the

³ Available at <https://github.com/HIT-SCIR/ltp>.

⁴ Available at <https://code.google.com/p/word2vec/>.

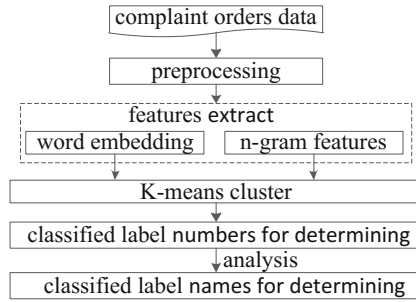


Fig. 1. The process of determining classification labels for complaint orders

performance based on trigram features is still not good. This may be due to that trigram features are easy to generate data sparsity and cause probability distortion. (c) Clustering results based on sentence representations feature are also unsatisfactory, the reason is that: ① complaint orders are too long, which may bring noises to obtain sentence representations by the way of word embedding summation. ② templates of complaint orders are similar, which result in low discriminability of sentence representations. So this paper performs cluster experiments based on bigram features by varying K from 5 to 12 with an interval of 1.

- (4) In classified label number for determining phase, we use error square sum of elbow method to determine the number of classified labels is 8. The relationship between error square sum and K is an elbow shape. The corresponding K value of elbow is the real clustering number of data.
- (5) In classified label names for determining phase, 500 complaint orders from each class are randomly selected and given to the business experts for analysis based on bigram features and $K = 8$, aiming to determine the classification label names.

3.2 Annotation Rules

Making annotation rules of COCC is difficult, because it involves complicated professional knowledge. We craft the following annotation rules of each class based on the analysis of complaint orders and professional knowledge in mobile customer service.

- (1) marketing activities question

Complaint orders that customer complains marketing activities are annotated “marketing activities question”. For example, customers are not satisfied with the rules of marketing activities, and complain that they can’t participate in marketing activities.

Annotation rule: the object of complaint order is about marketing activities, such as participation or withdrawal of marketing activities, dissatisfaction with rules about marketing activities, disagrees of propaganda and practical, without receiving gifts about marketing activities (including telephone fees/flow) on time, quality problems (including error telephone fees), etc.

(2) unknowing customization question

Without being aware of it, customer complains that some businesses are opened, cancel, change should be annotated unknowing customization question, such as unsubscribe.

Annotation rule: complaint orders that the customer clearly state that a product/business has been opened, changed or cancelled without being aware of it, are concluded “unknowing customization question”. The cost query arising from unknowing customization is classified as “unknowing customization question”.

(3) information security question

Complaint orders that the customer’s personal information or privacy information is got damage should be annotated “information security question”, such as password stolen.

Annotation rule: complaint orders about crank call/short messages, telephone fraud/short messages, adverse website report, subscriber information revealing, card replication, monitoring, telephone poisoning, etc., should be classified “information security question”.

(4) service question

Complaint orders about service channels, business related service attitude and service quality, etc., should be annotated “service question”, such as slow process for solving problem.

Annotation rule: complaint orders related to various service channels in mobile customer service should be labeled as “service question”, which contains: service attitude, service quality, service aging, service channels can’t provide normal service, etc. For service flow regulation of service channels, service opening time, discontent with service boundaries, service problem from official business channel, etc. For example, the adjustment of service time during Spring Festival, scope of service at night.

(5) cost question

Complaint orders about tariff and consumption situation should be annotated “cost question”, such as accounting query, charge after cancelling.

Annotation rule: complaint orders are annotated as cost question should meet either of two situations: (a) the customer have the accounting query for the use of tariff or some products/businesses, including the disagree of telephone fees, without using so much, charge after cancelling, having free resources but charging, etc. (b) the customer clearly state that he do not use flow but generate fees, such as the flow can’t access to the Internet.

(6) business processing question

Complaint orders about the failure to provide specific business/products for the mobile business and results in fault problems in the process of handling, are annotated “business processing question”. such as involving products or business, the failure to

open, cancel, change, the suite is not successfully processed, the invoice can't be printed, the flow is not reached, the fixed tariff can't be cancelled, etc.

Annotation rule: When judging, confirm whether the object of complaint order is a product or business provided by Mobile Corporation. If so, we annotate complaint orders about products/business failed to open, change, cancel as "business processing question". Moreover, complaint orders about payment problems (such as fees recharged and fees reached is different) and unsuccessful registration of real-name system are also labeled "business processing question".

(7) network question

Complaint orders about going online, talking quality of speech, signal intensity, etc. are annotated "network question", such as no signal, weak signal.

Annotation rule: if complaint orders are about network and there are clear official notifications, which should be annotated "network question". If there is no reference to the surrounding people, it does not belong to the "network question". If the customer expresses the specific position in the complaint order, it is still classified as "network question", although it does not mention other people conditions.

(8) business use question

Complaint orders about a fault caused by the operation of the specific products/business should be annotated "business use question", such as integral reduction without reason, display error 678 of broadband connection, broadband can't access to Internet.

Annotation rule: when judging, we need to confirm whether the object of complaint order is a product or business provided by the Mobile Corporation. If so, as long as the product/business can't provide the service normally, it should be annotated as "business use question". Moreover, complaint orders about broadband use problems, and integral use problems (including exchange telephone fare, call duration, short messages, the failure of virtual products to reach and logistics problems of exchanging integral for material object, etc.) also should be annotated "business use question".

3.3 Annotation Process

This paper takes complaint order as the annotated unit (no matter how many sentences are included, each complaint order has only one annotated label). Annotation process is shown in Fig. 2. Firstly, we analyze characteristics of complaint order, and craft detailed annotation rules. Secondly, rule makers explain annotation rules to annotators from the annotation team, and annotators carry out annotation on complaint orders. Finally, rule makers check annotated results, and we carry on the repeated discussion and revision about bifurcation points with annotators, business experts without participation in annotating the corpus, to determine inconsistent labels of complaint orders, then COCC is construct finally.

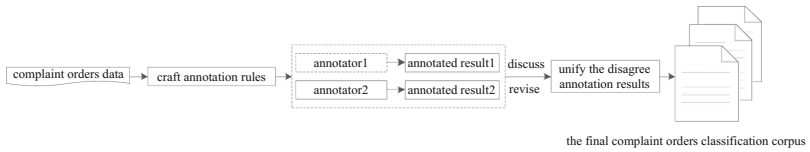


Fig. 2. Annotation process for the corpus

3.4 Basic Rules for Complaint Orders Classification Corpus

Understanding the customer's problem and taking customer's complaint intention as the main consideration is the premise of annotating the corpus. three kinds of complaint orders are discarded as follows:

- duplicate order: complaint order has no specific complaint content and only a number of the previous complaint order, which is directly marked as "duplicate order".
- template order: a single template built for the order, which has no free text information about customer's complaint intention, is labeled as "template order".
- invalid order: complaint order that customer's description is not clear, and it is hard to find the customer's complaint point, is labeled as "invalid order".

To ensure the accuracy and effectiveness of the annotated COCC, annotators must strictly obey the following annotation rules in annotation process.

(1) On the basis of customer requirement

When annotating, it is necessary to fill in the customer requirement from the perspective of customers. As long as the customer expresses the intention whether the requirement is reasonable or not, the intention of complaint order should be annotated.

(2) Abandon business experience

The purpose of annotation is to let the system learn the rules of COCC from massive data. As the system does not have business knowledge and reasoning ability, in annotation process, the business experience can't be considered and result annotated can't be obtained based on reasoning.

(3) Avoid speculation

When annotating complaint orders, annotators can't dig into the subjective conjecture, and can't speculate on customer's complaint intention beyond the intention expressed in the complaint order.

(4) Purely lean on literals

To ensure the effectiveness and accuracy of the annotated data, annotators should purely on literals in the complaint order, and do not make associations and experience judgments beyond literals.

3.5 Annotation Explanation of Bifurcation Points

Although clear labeling rules have been formulated, there are still bifurcation points due to some subjectivity in annotation process, and the flexibility and complexity of complaint orders recorded, which lead to the disagreement of annotation results. Bifurcation points are as follows:

- Whether the broadband problem is network question. We think that broadband is a product provided by Mobile Corporation, rather than a network service, so the broadband problem is annotated “business use question”.
- Query fees problem generated by unknowing customization is prone to confusion. If the customer has clearly expressed that he did not subscribing the related business, and the fee is only caused by unknowing customization, the order should be annotated “unknowing customization question”.
- Unknowing customization question and marketing activities question are easy to confuse. If we have verified that the customers subscribing the business from marketing activities, according to the special situation that “subscribing marketing activities without being aware of it should be labeled as marketing activities question” to unify them.
- Service question and business processing question are easy to confuse. Business processing question is only appearance of the customer’s complaint. If the customer’s final complaint intention is that some businesses fail to be processed because of the mistake made by the staff, this order should be annotated as “service question”.

4 Statistics and Agreement Analysis

4.1 Statistics for Complaint Orders Classification Corpus (COCC)

COCC consists of 130044 complaint orders annotated⁵. Table 1 shows the statistical information about COCC, it indicates that:

- (1) The three highest proportion of complaint orders are business use question, marketing activities question and cost question. Analyzing the Top3 questions of complaint orders are of great significance for quickly locating the categories of customer complaints and improving the quality of service. Moreover, information security question is only 2.29%, the main reason is that: the specialty of the mobile customer, the Mobile Corporation is strict with the management of customer’s privacy information, which leads to the number of complaint orders belonging to information security question is very less.
- (2) The average length of complaint orders from 8 categories is between 200 and 800 characters, among them the average length of the complaint orders from information security question is the longest, reaching 772.31 characters. This may due

⁵ Available at <https://github.com/zhng1200/COCC>.

Table 1. Statistical information about COCC

Classified label names	#complaint orders	%complaint orders	Avg. length of complaint orders
Network question	10262	7.89	413.13
Cost question	17901	13.77	274.14
Marketing activities question	22702	17.46	418.38
Business use question	46438	35.71	452.63
Business processing question	11831	9.10	329.49
Service question	10004	7.69	383.91
Unknowing customization question	7925	6.09	344.82
Information security question	2981	2.29	772.31

(#complaint orders: the number of complaint orders. Avg. Length: The average number of Chinese characters.)

to that the complaint order's template from information security question is too long, and more detailed information is added by the customer service staff when they record the content of complaint order.

4.2 Agreement Analysis

Each complaint order is annotated the only question label, so the recall is 100%. The accuracy is used as the agreement (the number of complaint orders annotated by two annotators is same/the total number of complaint orders \times 100%.) to measure the annotated effect. Table 2 shows the agreement analysis on COCC. Result_{1&2} is the agreement between the first annotation results and the second annotation results. Result_{1&final} is the agreement between the first annotation results and the final annotation results, and result_{2&final} is the consistency between the second annotation results and the final annotation results. It shows that:

- (1) The agreements of 8 question labels reach over 91%. Artstein and Poesio [18] show that the agreement of annotated corpus reaches 80% can be considered as trustworthy. So the corpus we construct is reliable.
- (2) The first column's agreements are generally lower than that the second column and the third column. This may due to that: the annotated results are influenced by subjective factors, and the final COCC is obtained by unifying the difference between the first annotated results and the second annotated results.
- (3) The agreement of cost question and service question are relatively low, and the agreement of marketing activities question, network question and unknowing customization question are relatively high. Because the characteristics and distinctions of each question is different. For example, compared with other classes, cost question and service question annotation are more complicated.

Table 2. Agreement analysis on COCC

Classified label names	Result _{1&2} (%)	Result _{1&final} (%)	Result _{2&final} (%)
Network question	96.22	97.89	97.76
Cost question	91.21	93.76	92.83
Marketing activities question	96.06	97.01	96.94
Business use question	94.32	94.67	95.42
Business processing question	92.81	93.11	94.02
Service question	92.23	92.36	92.51
Unknowing customization question	96.11	97.27	97.15
Information security question	93.24	94.69	94.57

The agreement of COCC reaches over 91%. It indicates that COCC is effective and could provide a great support for classifying complaint orders, mining semantic information to carry on the specialized analysis.

5 Conclusion and Future Work

This paper constructs complaint orders classification corpus which consists of 130044 annotated complaint orders. Firstly, we combine K-means and professional knowledge to determine classification labels of complaint orders. Secondly, we craft detailed annotation rules and annotate the complaint orders. Finally, we statistically analyze the constructed corpus, and the agreements of each question class reach over 91%. As mentioned above, the purpose of corpus constructed is to study complaint orders classification, and to enrich the semantic information for thematic analysis. Our future work will focus on studying complaint orders classification on COCC, and to extract and integrate semantic information in mobile customer service field.

References

1. Lowe, R., Pow, N., Serban, I.V., Pineau, J.: The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-turn Dialogue Systems. arXiv preprint [arXiv:1506.08909](https://arxiv.org/abs/1506.08909) (2015)
2. Hu, B.T., Chen, Q.C., Zhu, F.Z.: Lcsts: A Large Scale Chinese Short Text Summarization Dataset. arXiv preprint [arXiv:1506.05865](https://arxiv.org/abs/1506.05865) (2015)
3. Yang, L.Y.: Research on the establishment and applications of public sentiment corpus based on micro-blog information. *Office Inform.* **22**, 015 (2016)
4. Xi, X.F., Zhu, X.M., Sun, Q.Y., Zhou, G.D.: Corpus construction for chinese discourse topic via micro-topic scheme. *J. Comput. Res. Develop.* **54**(8), 1833–1852 (2017)
5. Xue, N.W., Chiou, F.D., Palmer, M.: Building a large-scale annotated Chinese corpus. In: *Proceedings of COLING*, pp. 1–8. ACL, Taipei (2002)
6. Aksan, Y., Aksan, M., Koltuksuz, A.: Construction of the Turkish National Corpus (TNC). In: *Proceedings of LREC*, pp. 3223–3227. European Language Resources Association, Istanbul (2012)

7. You, Z.Y., Wang, Y.Q., Shu, H.P.: A corpus-based TCM symptoms of speech tagging. *Electron. Technol. Softw. Eng.* **21**, 177–178 (2017)
8. Brockett, C., Dolan, W.B.: Support Vector Machines for Paraphrase Identification and Corpus Construction. In: *Proceedings of IWP*, pp. 1–8. IWP, Iowa (2005)
9. Dolan, W.B., Brockett, C.: Automatically constructing a corpus of sentential paraphrases. In: *Proceedings of IWP*, pp. 9–16. IWP, Iowa (2005)
10. Vincze, V., Szarvas, G., Farkas, R.: The bioscope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* **9**(Suppl. 11), S9 (2008)
11. Zou, B.W., Zhu, Q.M., Zhou, G.D.: Negation and speculation identification in Chinese language. In: *Proceedings of ACL-IJCNLP*, pp. 656–665. ACL, Beijing (2015)
12. Zhou, H.W., Yang, H., Xu, J.L., Kang, S.Y.: Construction of Chinese hedge scope corpus. *J. Chin. Inf. Process.* **31**(3), 77–85 (2017)
13. Yin, P.F., Liu, Z., Xu, A.B.: Tone analyzer for online customer service: an unsupervised model with interfered training. In: *Proceedings of CIKM*, pp. 1887–1895. ACM, Singapore (2017)
14. Quan, C.Q., Ren, F.J.: Construction of a blog emotion corpus for chinese emotional expression analysis. In: *Proceedings of EMNLP*, pp. 1446–1454. ACL, Singapore (2009)
15. Chen, J., Nie, J.Y.: Automatic construction of parallel English-Chinese corpus for cross-language information retrieval. In: *Proceedings of ANLP*, pp. 21–28. ACL, Stroudsburg (2000)
16. Feng, G.J., Yu, L., Tian, S.W.: Auto construction of uyghur emotional words corpus based on CRFs. *Data Anal. Knowl. Discov.* **27**(3), 17–21 (2011)
17. Yang, J.F., Guan, Y., He, B., Qu, C.Y., Yu, Q.B., Liu, Y.X., Zhao, Y.J.: Corpus construction for named entities and entity relations on chinese electronic medical records. *J. Softw.* **27**(11), 2725–2746 (2016)
18. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. *Comput. Linguist.* **34**(4), 555–596 (2008)