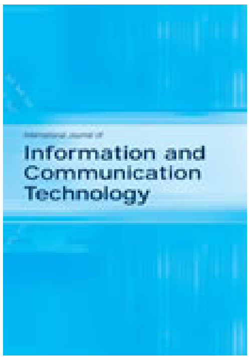


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Legal framework development through adversarial transfer learning for consumer grievance classification in algorithmic price discrimination contexts

Xiaoling Guo

DOI: 10.1504/IJICT.2025.10072531

Article History:

Received:	02 June 2025
Last revised:	16 June 2025
Accepted:	16 June 2025
Published online:	30 July 2025

Copyright © 2025 Inderscience Enterprises Ltd.

Legal framework development through adversarial transfer learning for consumer grievance classification in algorithmic price discrimination contexts

Xiaoling Guo

School of Economics and Finance,
Zhanjiang University of Science and Technology,
Guangdong 524084, China
Email: guoxiaoling808@163.com

Abstract: The escalation of algorithmic price discrimination necessitates systematic analysis of consumer grievance data. Addressing the scarcity of annotated datasets in regulatory research, this study proposes a hybrid neural framework integrating BERT-based semantic encoding with optimised convolutional architectures. The model employs bidirectional recurrent layers to capture sequential dependencies while applying multi-head attention mechanisms for contextual feature fusion. A domain adaptation strategy combining adversarial training and transfer learning bridges feature distribution gaps between source and target domains. Through iterative parameter optimisation, the framework achieves cross-domain knowledge transfer while maintaining discriminative classification capabilities. Empirical validation demonstrates significant performance improvements, with macro-F1 scores increasing by 13.29% compared to baseline models. The classification outcomes inform three regulatory proposals addressing dynamic pricing oversight, algorithmic transparency requirements, and consumer compensation mechanisms. This dual technical-legal approach provides implementable solutions for governing emerging digital market practices while advancing domain adaptation methodologies in computational legal studies.

Keywords: algorithmic price discrimination; complaint information categorisation; legal regulation; attention mechanisms; adversarial transfer learning.

Reference to this paper should be made as follows: Guo, X. (2025) ‘Legal framework development through adversarial transfer learning for consumer grievance classification in algorithmic price discrimination contexts’, *Int. J. Information and Communication Technology*, Vol. 26, No. 30, pp.81–96.

Biographical notes: Xiaoling Guo obtained her Master’s degree in Economic Law from Anhui University in 2007. She is currently an Associate Professor at the School of Economics and Finance, Zhanjiang University of Science and Technology. Her research directions include data mining, consumer protection law, product quality law, e-commerce law and digital economy.

Copyright © The Author(s) 2025. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

1 Introduction

As the China's internet giants rapidly rising, the platform economy, a new form of economy, has gradually come into the public eye and become a hotspot for people's attention. Big data kills familiarity precisely in the context of the rapid development of the platform economy, it is through the collection and processing of customer information, the use of algorithmic technology to accurately locate the customer's consumption preferences, and then realise the differentiation of pricing (Cao and Yang, 2023). In the face of big data killing behaviour, consumers often choose to file a complaint to protect their rights (Donoghue and De Klerk, 2009). If the cause of the complaint is not localised properly, it cannot be assigned to the appropriate support department in time to provide an appropriate solution, which will reduce customer satisfaction and may cause the complaint to escalate. Therefore how to effectively categorise this consumer complaint information has become a pressing issue (HaCohen-Kerner et al., 2019). Accurate categorisation of complaint information not only helps relevant departments to quickly understand the crux of the problem and improve handling efficiency, but also provides strong data support for subsequent supervision and legal regulation.

Consumer complaint information categorisation and legal regulation is to obtain the real evaluation of consumers through text categorisation method, so as to integrate the evaluation results to propose legal regulation (Choe et al., 2013). Bozyiğit et al. (2022) proposed a decision tree-based text categorisation algorithm for customer complaints, but there is a risk of overfitting. Yang et al. (2018) synthesised textual information in customer complaints and used evidential inference rules to construct a classification model for the integrated strategy to classify customer complaints. Ghazzawi and Alharbi (2019) feature select the complaint text by GRW model to capture effective characteristic words, and then construct a user complaint text categorisation approach relied on FastText model, and propose a series of legal regulation suggestions using the classification results, but since this method requires training a large number of complaint questions, it leads to low classification accuracy.

Machine-learning features need to be manually crafted for complex data analysis and feature engineering, resulting in inefficient text categorisation. Deep learning can automatically extract key characteristics from data, greatly reducing the workload of manually designing features. Khedkar and Shinde (2020) used convolutional neural networks (CNNs) to construct a text classification model for building quality complaints and achieved better performance than SVM models. Gupta et al. (2021) proposed a recurrent neural network (RNN)-based text categorisation method for railroad complaints with good results. Naik et al. (2023) modelled text representation through LSTM and attention mechanism and combined it with entity representation to further improve text classification performance. Wang et al. (2023) proposed a joint attention augmentation network based on the BERT model for text categorisation of citizen complaint reports, and used the categorisation results for the design of legal regulation methods. Liu et al. (2024) used BERT to extract contextual information, and then used the multiscale CNN-Inception module to extract more features to improve the model performance. Zhang et al. (2024) used BERT training word vectors as embedding layers, and used a two-layer LSTM network and attention mechanism to capture text contextual characteristics and key information, and achieved better classification results. With the advancement of transfer learning techniques, knowledge from related domains is applied to the target domain by means of knowledge transfer, thus enhancing the performance indicators of the target domain tasks. Hossain et al. (2022) used transfer learning to achieve classification of complaint opinion texts by preventing the introduction of specific information from a small corpus into a shared space, thus improving classification accuracy. Fan et al. (2023) used transfer learning to construct a deep learning model relied on the similarity of customer complaint texts, and used domain knowledge to assist text credibility analysis to classify false texts. The adversarial vulnerability of deep neural networks can mislead state-of-the-art (SOTA) classifiers to make incorrect predictions. To solve this issue, adversarial training essentially enhances the robustness of the model. Zhang et al. (2020) extended the adversarial training approach to the NLP domain and applied perturbation to the text categorisation task on word embeddings based on LSTM models, using the FGM algorithm for the computation of perturbation.

According to the analysis of existing research, it is known that traditional research exists the problem of lack of labelled data and poor classification effect, for this reason, this paper offers a consumer complaint information classification approach relied on adversarial migration learning in the context of big data killing, and provides relevant and referable suggestions for legal regulation. The innovativeness of the proposed methodology is reflected in the following four main aspects.

1 The BERT model is adopted to learn the dynamic word vector representation of the text, and the CNN is optimised (RCNN) by using a bidirectional recurrent structure to obtain local features and a maximum pooling level to capture global characteristics to improve the feature extraction capability of the model.

2 A multi-head collaborative attention mechanism is designed to reflect the relationship among important characteristic vectors by computing the shared similarity matrix among characteristic vectors, which is adopted to complain about the interaction between local and global characteristics, and the interacted feature vectors are fused with local and global characteristics to gain the fused features.

3 Adversarial transfer learning is introduced to migrate the characteristics of the initial labelled data to achieve feature adaptation in the source and target domains, and the classification results are obtained by softmax. Adversarial training is used to continuously correct the classification results of the model, so as to improve the classification ability. The fusion of the classification results suggests the legal regulation of consumer protection.

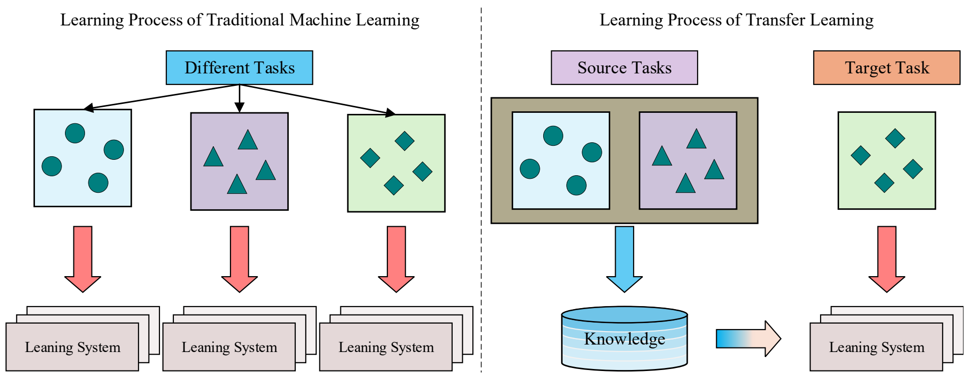
4 Simulation experiments were conducted on the TCCD and ChnSenti-Corp datasets, and the outcome implies that the proposed models exhibit excellent classification with average classification accuracies of 89.84% and 92.39%, respectively, which are better than the comparison models.

2 Related works

2.1 Transfer learning

Most of the existing machine learning algorithms are relied on supervised algorithms, which often encounter the problem of insufficient labelled data in the training process. To address this challenge, many researchers have focused on the development of transfer learning (Weiss et al., 2016). Machine learning automatically learns patterns and rules from data and builds models for prediction or decision making. Transfer learning utilises the knowledge of the source domain and transfers it to the target domain to solve the problems of data scarcity or domain differences. Transfer learning resolves the limitations of machine learning in scenarios of scarce data and domain differences, and enhances the adaptability and performance of the model, Figure 1 implies the traditional machine learning process on the left and the migration learning process on the right.

Figure 1 Traditional machine learning and transfer learning (see online version for colours)



To express the transfer learning definition more intuitively, first understand two basic concepts, one is called domain and the other is called task, the concepts are as follows:

1 Domain: a domain D, consisting of a characteristic space χ and an edge probability distribution $P(X)$, i.e., $D = \{\chi, P(X)\}$, where $X = \{X_1, X_2, \ldots, X_n\} \subset \chi$.

2 Task: given D, the corresponding task T is made up of a labelling space Y and a forecasting function $f(\cdot)$, i.e., $T = \{Y, f(\cdot)\}$, where $f(\cdot)$is learned using the feature vector x_i and the corresponding label y_i, i.e., the combination of the two, $\{x_i, y_i\}$, where $x_i \in X$, $y_i \in Y$.

Based on the above basic concepts and notations, assuming the existence of a source domain Ds and a target domain Dt, and the corresponding source and target tasks Ts and Tt in the domains, transfer learning is the process of improving the prediction function $f_t(\cdot)$of the target task Tt on the target domain Dt by adopting the correlation information obtained from Ds and Dt, where $Ds \neq Dt$ or $Ts \neq Tt$, and the number of the source domains can be generalised from one to many.

2.2 SENet attention mechanisms

Adversarial training (AT) is a training method that introduces noise into the model to generate adversarial samples, enabling the model to correctly categorise the adversarial and original samples, thus improving the robustness of the model (Andriushchenko and Flammarion, 2020). When training on supervised data, cross entropy is adopted as the loss function with the following equation:

$$\mathcal{L} = -\log p\left(y|x + r_{adv};\theta\right) \quad (1)$$

$$r_{adv} = \arg\min \log p\left(y|x + r;\hat{\theta}\right) \quad (2)$$

where θ is the model parameters, x is the input sample, and y is the true label.

The optimal adversarial perturbation is the perturbation generated in the worst case, i.e., the perturbation that maximises the model loss. Based on this idea, the process of adversarial training can be abstracted as equation (3).

$$\min_{\theta} E_{(x,y)\sim D} \max L\left(\theta, x + r_{adv}, y\right) \quad (3)$$

where r_{adv} is the computed perturbation, S is the perturbation space, L is the loss function, is the model parameters, and x and y correspond to the inputs and outputs, respectively. d is the sample space, and e is the empirical risk.

In adversarial training, external empirical risk minimisation can be achieved by using methods such as gradient descent, but internal loss maximisation is difficult to obtain, i.e., the perturbation cannot be calculated directly and can only be solved by approximation. Therefore, how to find the optimal adversarial perturbation is a difficult point in adversarial training (Ganin et al., 2016).

3 Consumer complaint information preprocessing and feature extraction

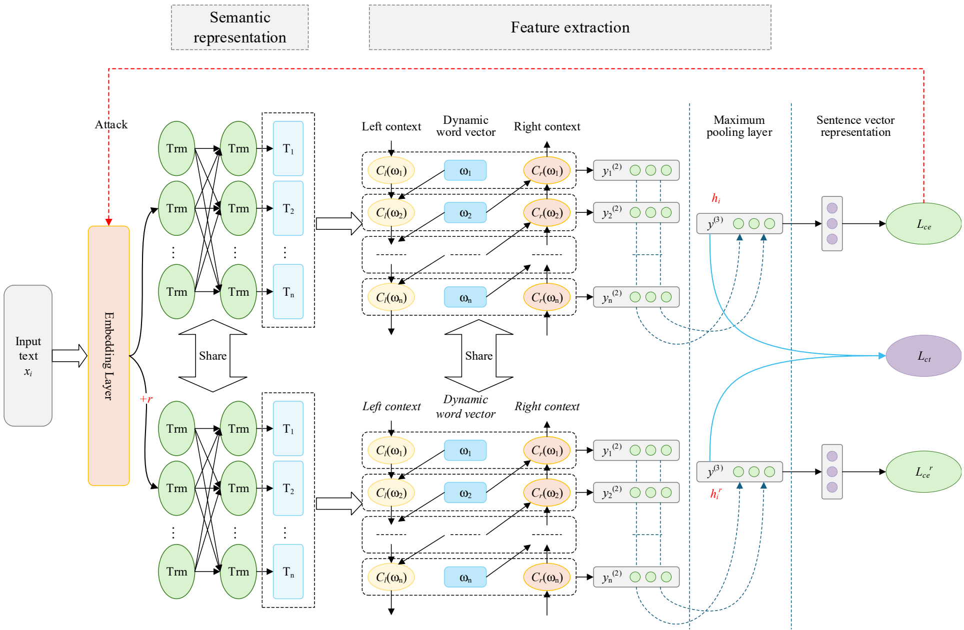
3.1 Text embedding representation of consumer complaints based on BERT modelling

Most of the information of consumer complaints is text data, before text classification, the complaint text needs to be preprocessed and feature extraction, as shown in Figure 2, in this paper, the vectorised representation of the text is obtained through the BERT model, and the global and local features are captured by the improved CNN.

Conventional text vectorisation representation models such as Word2Vec, Doc2Vec, LDA, etc. (Abubakar et al., 2022) cannot cope with the issue of multiple meanings of a word well. The BERT model not only addresses the issue of multiple meanings of a word efficiently. For the text sequence $S = [x_1, x_2, x_3, \ldots, x_n]$ to be trained, it is input into the BERT model for pre-training. Firstly, the initialised word vector embedding representation is generated for each input sentence by superimposing the features using word embedding, paragraph embedding and position embedding. Then, the transformer's own self-attention mechanism is utilised to link the contextual semantic information for feature extraction.

In this way, the BERT model finally generates a sequence of word vectors as $T = [T_1, T_2, T_3, \ldots, T_n]$, where T_i is the vector representation of the i^{th} word in the text S, and the dimensionality of T is $n \times 768$. These word vectors combine contextual and global semantic information, and can provide more accurate and rich feature representations for subsequent classification tasks.

Figure 2 Consumer complaint information pre-processing and feature extraction process (see online version for colours)



3.2 Text feature extraction based on improved CNN

In the feature extraction stage, the current mainstream method is to utilise CNN or RNN for feature extraction, but when analysing document semantics with RNN, the later words are more dominant than the earlier ones, which reduces the model efficiency (Du and Huang, 2018). Whereas CNNs lose more feature information when they choose a smaller window size. For this reason, this paper adopts the bidirectional recurrent structure of LSTM for CNN optimisation (RCNN), as shown in Figure 3.

First for the dynamic word vector T_i, denoted as $e(w_i)$, which is output from the encoding of the BERT model, and then the dynamic word vector is fed into the RCNN model in order to extract the contextual information to obtain $c_l(w_i)$, $c_r(w_i)$, where $c_l(w_i)$ is the semantic representation of the left side of w_i, and $c_r(w_i)$ is the semantic representation of the right side of w_i.

$$c_l(w_i) = f\left(W^{(l)} c_l(w_{i-1}) + W^{(sl)} e(w_{i-1})\right) \tag{4}$$

$$c_r(w_i) = f\left(W^{(r)} c_r(w_{i+1}) + W^{(sr)} e(w_{i+1})\right) \tag{5}$$

Subsequently, $c_l(w_i)$, $c_r(w_i)$ and $e(w_i)$ are stitched together to gain a localised characteristic representation of the text, which is fed into the fully connected network for integration. To enhance the nonlinear expression capability of the model, the ReLU activation function is used in the fully connected network for nonlinear mapping, which can effectively alleviate the issue of gradient vanishing, and at the same time, the computation speed is faster.

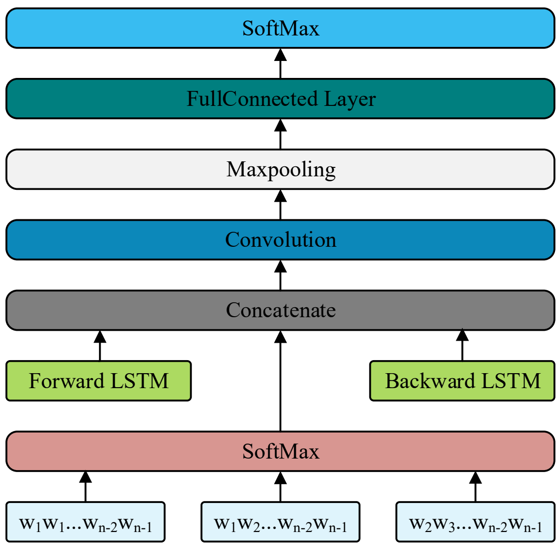
$$x_i = \left[c_l\left(w_i\right); e\left(w_i\right); c_r\left(w_i\right)\right] \tag{6}$$

$$y_i^{(2)} = \text{ReLU}\left(W^{(2)} x_i + b^{(2)}\right) \tag{7}$$

Finally the output of the fully connected network is maximally pooled to gain a global characteristic representation of the text.

$$y^{(3)} = \max_{i=1}^{n} y_i^{(2)} \tag{8}$$

Figure 3 The structure of RCNN (see online version for colours)



3.3 Feature fusion based on multi-headed collaborative attention mechanism

After obtaining the global and local feature representations of the text, this paper innovatively proposes the multihead collaborative attention mechanism (MCAM), which reveals the relation among important characteristic vectors by computing the shared similarity matrix among the characteristic vectors, and is used to complain about the interaction between the local and global characteristics. The input features are first downscaled by h linear layers to acquire the expression of the identical characteristic vector in various vector spaces as shown in equations (9) and (10).

$$x_1, x_2, \ldots, x_h = Linear_{1,2,\ldots,h}^{a}(x) \tag{9}$$

$$y_1, y_2, \ldots, y_h = Linear_{1,2,\ldots,h}^{b}(y) \tag{10}$$

After that, the shared similarity matrix $C_i = \tanh(x_i^T \cdot y_i)$ between the feature vectors is calculated inside MCAM, and the similarity matrix is extrapolated to the original feature vector matrix to realise the feature interaction, and the extracted features f_i^x and f_i^y are shown as follows:

$$\begin{cases} f_i^x = \tanh\left(x_i C_i\right) \\ f_i^y = \tanh\left(y_i C_i^T\right) \end{cases} \tag{11}$$

The outputs of MCAMs are spliced together to gain the characteristic outputs F_{xy} and F_{yx} of the relationship between important feature vectors, which are calculated as shown in equations (12) and (13).

$$F_{xy} = concat\left(f_1^x, f_2^x, \ldots, f_h^x\right) \tag{12}$$

$$F_{yx} = concat\left(f_1^y, f_2^y, \ldots, f_h^y\right) \tag{13}$$

This paper borrowed the residual linkage operation similar to ResNet (Xu et al., 2023) and the MCAM operation in transformer (Han et al., 2021), and added the feature vectors F_{xy} and F_{yx} after the interaction, and the local characteristic x_i and the overall characteristic y_i of the text of the consumer's complaint before the interaction, and then inputted them into the full linkage level. The fusion feature vector F, which contains the relationship information among the characteristic vectors, is gained as follows:

$$F = concat\left(F_{xy}, F_{yx}, x_i, y_i\right) \tag{14}$$

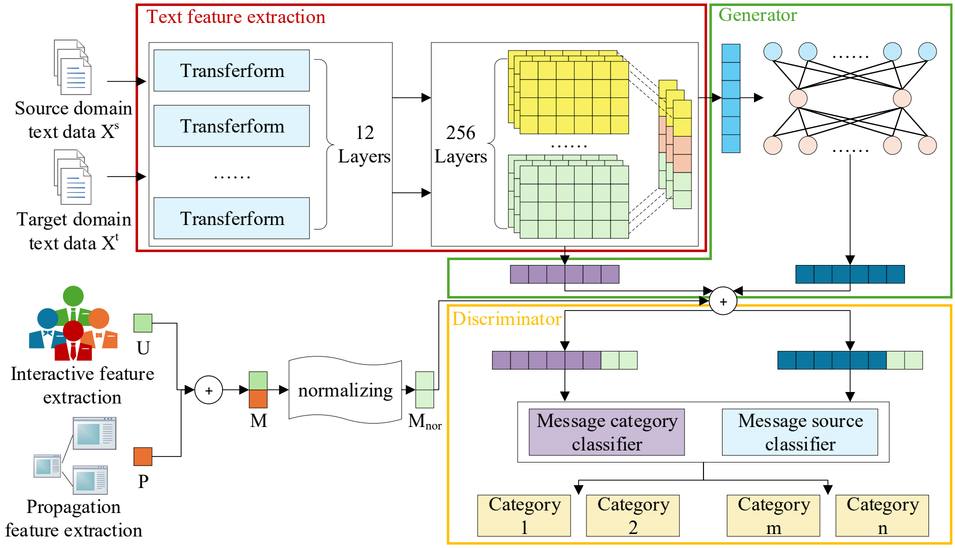
4 Classification of consumer complaint information based on adversarial transfer learning in the context of big data killers

4.1 Classifying consumer complaint information based on adversarial transfer learning

To cope with the issue of lack of labelled data and insufficient classification discrimination ability in the current research, this paper migrates the features of the original labelled data through adversarial migration learning, so that its feature distribution converges to the characteristic distribution of the new text, and realises the feature adaptation of the source and target domains, so as to obtain the accurate classification results through softmax. Finally the ATL-CCI model keeps on correcting its own parameters through adversarial training and the ability of the model to classify different consumer complaint information keeps on improving. The proposed classification model is shown in Figure 4.

After obtaining the fusion feature F, this paper realises the correction of the parameters of the consumer complaint information classification model by means of adversarial migration. In this section, the labelled data of the original complaint platform is defined as the source domain data: $S = \{[X_1^s, B_1^s], [X_2^s, B_2^s], \ldots, [X_M^s, B_M^s]\}$, where X_i^s is the text information in the i^{th} data and B_1^s is the category of the i^{th} information. The labeled data in the newly occurred complaint event is defined as the target domain data: $T = \{[X_1^t, B_1^t], [X_2^t, B_2^t], \ldots, [X_N^t, C_N^t]\}$, where X_i^t is the text information in the i^{th} data and B_i^t is the category of the information. ATL-CCI recodes the fusion feature F in S by generator G. The ATL-CCI model uses an autoencoder (AE) as generator G. During adversarial training, G and D play with each other, and in the process, the model realises the modification of its own parameters so as to improve the classification accuracy.

Figure 4 The structure of the proposed classification model (see online version for colours)



1 Generator: the AE can learn the characteristics of the shared portion of the cross-domain data as it refactors the data. In summary, the text features of the data in *S* are recoded by AE during the adversarial migration process, and the distribution of the text features of the data in *T* is adjusted during the training process, so that the distribution of the features tends to converge to the distribution of the text features of the data in *T*. The recoded feature *F*′ is shown below, where θ^G is the full parameter of AE. Distinct from the traditional AE, this section updates the parameter θ^G of the AE by D determining the error in the source of complaint information.

$$F' = G\left(F, \theta^G\right) \tag{15}$$

2 *Discriminator:* the input to *D* consists of *F*′ and *F*. *F*′ and *F* are mapped by a fully-connected level and fused with the corresponding local and global features, respectively. Then the normalisation operation is carried out, and finally fused with the text features after mapping through a fully connected layer to obtain the recoded fused feature F^*, which is classified by softmax classifier as follows, where *W* is the weight and *b* is the bias.

$$z = \text{softmax}\left(WF^* + b\right) \tag{16}$$

4.2 Confrontation training process

There are two objective functions set in ATL-CCI, which are the objective function L_D of *D* and the objective function *B* of L_G. Discriminatory loss is $L_D = L_{ds} + L_{db}$. L_{ds} is to determine the loss of the source of the complaint information, as shown in equation (17), where when $S_i = 1$ is the data from the target domain *T*, and when $S_i = 0$ represents that the data comes from the source domain *S*. L_{db} is the loss to determine whether the complaint is true or not, as shown in equation (18), where when $B_i = 1$ indicates that the information is true news and when $B_i = 0$ indicates that the information is false.

$$\begin{aligned} L_{ds} &= E\left(\log P\left(\hat{S} = target \middle| S_i = 1\right)\right) + E\left(\log P\left(\hat{S} = source \middle| S_i = 0\right)\right) \\ &= E\left(\log\left[S_i * P\left(\hat{S} = target \middle| X_i\right)\right]\right) \\ &+ E\left(\log\left[(1 - S_i) * P\left(\hat{S} = source \middle| X_i\right)\right]\right) \end{aligned} \tag{17}$$

$$\begin{aligned} L_{db} &= E\left(\log P\left(\hat{B} = true \middle| B_i = 1\right)\right) + E\left(\log P\left(\hat{B} = false \middle| B_i = 0\right)\right) \\ &= E\left(\log\left[B_i * P\left(\hat{B} = true \middle| B_i\right)\right]\right) + E\left(\log\left[(1 - B_i) * P\left(\hat{B} = false \middle| B_i\right)\right]\right) \end{aligned} \tag{18}$$

The generating loss is $L_G = L_{gs} - L_{gb}$. L_{gs} is the loss of judging the source of the data, as shown in equation (19), where $S_i = 0$ is that the data comes from S. L_{gb} is the loss of judging the truth of the data, as shown in equation (20), where $B_i = 0$ stands for the complaint that the information is untrue.

$$\begin{aligned} L_{gs} &= E\left(\log P\left(\hat{S} = source \middle| S_i = 0\right)\right) \\ &= E\left(\log\left[(1 - S_i) * P\left(\hat{S} = source \middle| X_i\right)\right]\right) \end{aligned} \tag{19}$$

$$\begin{aligned} L_{gb} &= E\left(\log P\left(\hat{B} = false \middle| B_i = 0\right)\right) \\ &= E\left(\log\left[(1 - B_i) * P\left(\hat{B} = false \middle| B_i\right)\right]\right) \end{aligned} \tag{20}$$

The ATL-CCI model uses adversarial training to correct its own parameters, and the parameters of the discriminator and generator are updated separately during the training process in the following steps:

Step 1 Fix the parameters of G, train D, update the parameters θ^D according to the error of D, and maximise the value of L_D as shown in equation (21).

$$\theta_D = \arg\max\left(L_{db} + L_{ds}\right) \tag{21}$$

Step 2 Fix the parameters of D, train G, update the parameters θ^G according to the error of D, and maximise L_G as shown in equation (22).

$$\theta_G = \arg\max\left(L_{gc} - L_{gs}\right) \tag{22}$$

During the training process, G and D play with each other so that the distribution of original text features recoded by G converges to the distribution of text features in the new text data.

5 Integration of legal regulation of categorisation of information on consumer complaints

Consumer complaint information is an important window to reflect market problems, consumer needs and satisfaction. The classification results of consumer complaint information obtained through the ATL-CCI model can not only respond to consumer concerns in a timely manner, but also provide strong support for government regulation and business improvement. However, there are problems such as unclear categorisation and unstandardised handling of consumer complaint information, which urgently need to be improved through legal regulation. With regard to the handling of consumer complaints, China has initially established a legal regulatory system centred on the Consumer Rights and Interests Protection Law. However, in practice, there are still problems such as imperfect legal regulation and insufficient enforcement. In order to improve the mechanism of consumer complaint information processing, improve processing efficiency, this paper puts forward the following suggestions.

1 Formulate unified classification standards and data specifications, clarify the definition, scope and classification methods of various types of complaint information, and ensure data comparability and interoperability among different organisations and platforms. It is necessary to clarify the responsibilities and obligations of the main parties involved, including consumers, enterprises, platforms and regulators, and to standardise the collection, processing and use of complaint information. The core mechanism of the FedAG algorithm is to optimise the global model update in federated learning through gradient correction and dynamic aggregation strategies, especially when dealing with non-independent co-distributed (Non-IID) data, to enhance the convergence and generalisation ability of the model. The FedOpt algorithm achieves dynamic step size adjustment and gradient heterogeneity response in federated learning by introducing adaptive optimisers (such as Adam and Yogi), significantly improving the convergence and stability of the global model under Non-IID data. Compared with the traditional FedAvg, FedOpt is more suitable for complex tasks and scenarios with non-independent and identically distributed data. Compared with FedProx, FedOpt has advantages in communication efficiency and computational complexity. However, in extreme Non-IID cases, it may be necessary to combine other technologies (such as Fedprox-Adam) to further improve performance.

2 In terms of data protection and privacy protection, a sound legal system should be established to clarify the data security requirements and privacy protection measures in the process of handling complaint information. For example, the principle of data minimisation should be stipulated to limit the unnecessary collection of personal information; and a system for managing data access rights should be established to prevent the leakage and misuse of information. Mechanisms for sharing information on complaints should also be established to promote information sharing and collaborative governance among relevant departments and agencies while protecting privacy. The design of the data layer is of vital importance in federated learning and must simultaneously meet the two core requirements of data security and privacy protection as well as efficient data management and scheduling. Data privacy is protected by adopting technologies such as differential privacy and homomorphic encryption. Design reasonable data partitioning and load balancing strategies to prevent the client from becoming a bottleneck.

3 Innovations in legal regulation adapted to new types of consumer complaints. Special laws and regulations should be formulated to deal with complaints brought about by new consumption models such as sharing economy and live streaming. In the field of sharing economy, the sharing economy supervision regulations can be

formulated to clarify the main responsibilities of sharing economy enterprises, operation norms, and specific measures to protect the rights and interests of consumers. In terms of live banding, the measures for the administration of live banding can be introduced to provide detailed provisions on the qualifications of the main body of live banding and the content of live broadcasting. Clarify the joint and several liability of anchors and merchants in live streaming with goods, strengthen the supervision of live streaming platforms, and require platforms to establish a sound complaint handling mechanism to deal with the violations of merchants and anchors in a timely manner.

6 Experimental results and analyses

In this paper, experiments have been conducted on two datasets from different domains, TCCD, a customer complaint text dataset provided by telecommunication companies, and ChnSenti-Corp, a hotel complaint text dataset, where TCCD contains four categories labelled as 0–3 and ChnSenti-Corp contains ten categories labelled as 0–9. The specifics of each dataset are shown in Table 1, and the distribution of complaint text lengths in each dataset is shown in Figure 5.

Figure 5 The distribution of complaint text lengths in each dataset, (a) TCCD dataset (b) ChnSenti-Corp dataset (see online version for colours)

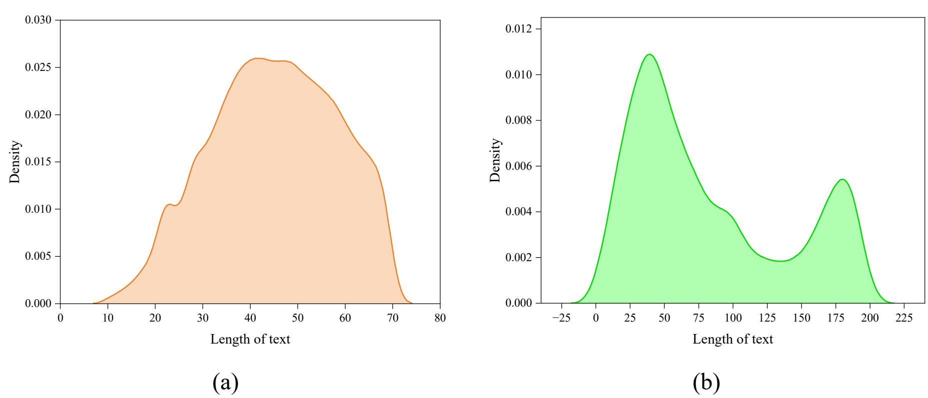


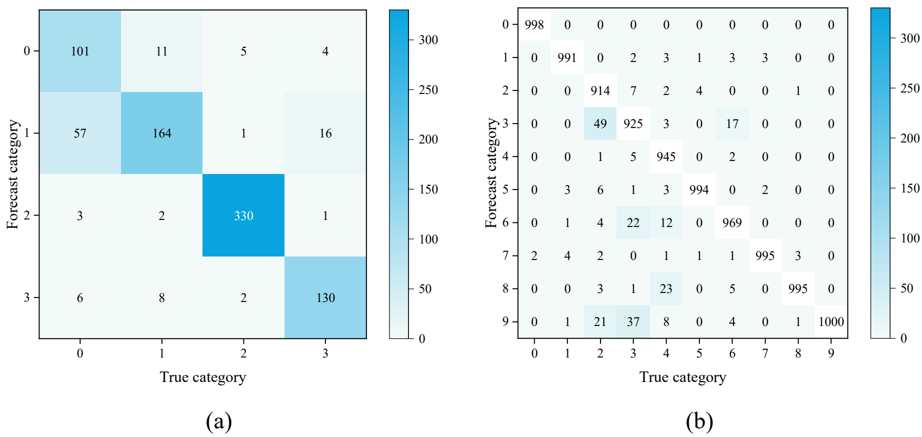
Table 1 Specifics of the dataset

Dataset	*Number of categories*	*Training set*	*Validation set*	*Test set*
TCCD	4	16,800	2,100	2,100
ChnSenti-Corp	10	14,400	1,800	1,800

The experiments were conducted using an Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz CPU, NVIDA GTX2080Ti with 12 GB of video memory, Ubuntu 7.4.0 as the operating system, Python 3.9 as the programming language, and Pytorch 1.11.0 as the deep learning framework. In the experiment, the batch size is set to 32, the learning rate is 2e-5, the optimiser is Adam, the dropout is set to 0.1, and the epoch is set to 4. The test set confusion matrix of the proposed model ATL-CCI on TCCD and ChnSenti-Corp datasets is shown in Figure 6. The classification accuracy of ATL-CCI in classes 2, 3 and

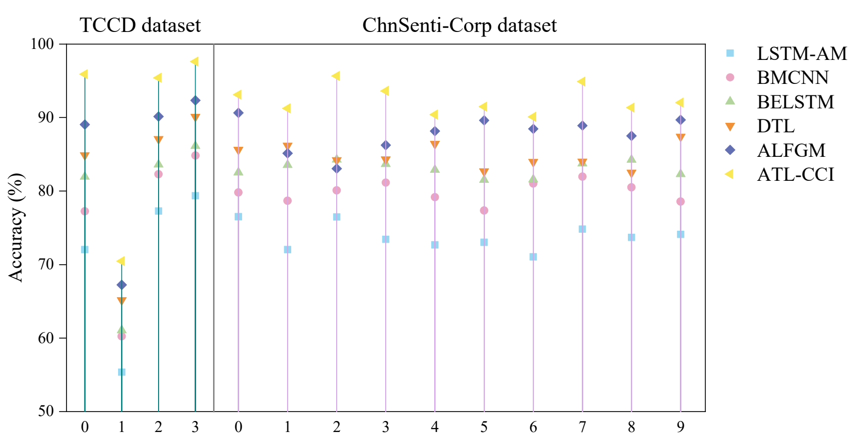
4 is relatively high, among which class 3 has an accuracy of 97.62%. The data in category 1 contains noise, which affects the classification accuracy of this kind of complaint text, resulting in a low classification accuracy of 70.47% for category 1. ATL-CCI performs well on the ChnSenti-Corp dataset, with accuracies of more than 90% in all categories, all of which show a high level of classification prediction.

Figure 6 Confusion matrix of ATL-CCI on two datasets, (a) confusion matrix on the TCCD dataset (b) confusion matrix on the ChnSenti-Corp dataset (see online version for colours)



To verify the classification effect of ATL-CCI model more comprehensively, ATL-CCI was compared with LSTM-AM model (Naik et al., 2023), BMCNN model (Liu et al., 2024), BELSTM model (Zhang et al., 2024), DTL model (Hossain et al., 2022) and ALFGM model (Zhang et al., 2020) were used for comparative experiments with accuracy (Acc), macro-precision rate (macro-P), macro-perfection rate (macro-R), and macro-F1 value (macro-F1). The classification accuracies of different categories of text on the two datasets are shown in Figure 7.

Figure 7 The classification accuracies of different categories (see online version for colours)



On the TCCD dataset, the average accuracy of ATL-CCI is 89.84%, which is an improvement of 18.82%, 13.68%, 11.64%, 8.02%, and 5.14% compared to LSTM-AM, BMCNN, BELSTM, DTL, and ALFGM, respectively. On the ChnSenti-Corp dataset, ATL-CCI has an average accuracy of 92.39%, which is an improvement of at least 4.65% compared to the other five models. ATL-CCI not only innovatively utilises RCNN to capture local and overall characteristics of text, but also achieves feature adaptation between source and target domains through adversarial migration learning, which greatly improves the classification accuracy.

Comparisons of macro-P, macro-R and macro-F1 of different models on the two datasets are implied in Table 2. Macro-F1 is the combined evaluation value of macro-P and macro-R, which best reflects the classification effect of each model intuitively. On the TCCD and ChnSenti-Corp datasets, the macro-F1 of ATL-CCI is 90.39% and 93.24%, respectively, which is at least 4.4% and 13.29% improvement compared to other models, respectively, indicating that ATL-CCI has better classification performance. LSTM-AM only obtains local semantic features of text through RNN without considering global features and adversarial training, resulting in poor classification results. Although BMCNN considers the multi-scale characteristics of the text, the local characteristics are not extracted and the model is not trained against them, and the classification accuracy is low. BELSTM also suffers from insufficient feature extraction, but enhances the important features through the attention mechanism, so the classification performance is better than LSTM-AM and BMCNN. Although DTL takes into account the problem of sample scarcity through migration learning, it does not enhance the robustness of the model by adversarial training, so the classification accuracy is not as good as that of ALFGM and ATL-CCI. The accuracy of DTL is worse than that of ATL-CCI because it considers the global features of the text. To summarise, ATL-CCI has excellent classification results.

Table 2 Comparison of classification performance on two datasets

Model	*TCCD dataset*			*ChnSenti-Corp dataset*		
	macro-P (%)	*macro-R (%)*	*macro-F1 (%)*	*macro-P (%)*	*macro-R (%)*	*macro-F1 (%)*
LSTM-AM	72.94	74.06	73.50	72.06	75.21	73.60
BMCNN	76.39	78.24	77.30	77.05	79.25	78.13
BELSTM	78.04	81.43	79.69	81.06	82.68	81.86
DTL	82.51	83.67	83.09	85.36	88.36	86.83
ALFGM	84.98	87.03	85.99	90.69	89.23	79.95
ATL-CCI	89.19	91.62	90.39	94.16	92.35	93.24

7 Conclusions

To solve the issue of poor classification effect of current consumer complaint information classification methods, this paper firstly uses BERT to realise the text vectorised representation, and optimises the CNN by using the bidirectional recurrent structure, and obtains the local and overall characteristics of the complaint text through RCNN. After that, MCAM is designed for the interaction among local and global characteristics of the complaint text, and the feature vectors after the interaction are spliced with the local and global characteristics to obtain the fusion features. Then adversarial migration learning is introduced to migrate the features of the original labelled data to achieve feature adaptation in the source and target domains, and softmax is used to obtain classification results. The classification results of the model are continuously corrected through adversarial training, so as to improve the classification ability. Finally, the results of the classification of consumer complaint information are integrated to propose the legal regulation of consumer rights protection. The experimental outcome on two real datasets show that the proposed model has a high classification accuracy and can provide a valuable reference for the legal regulation of consumer protection.

Declarations

This work is supported by the Research on Governance of 'Big Data Price Discrimination' in the Platform Economy of Guangdong Province under the Background of 'Dual-Zone' Construction: A Case-Based Empirical Analysis Perspective (No. 202212608005).

The author declares that she has no conflicts of interest.

References

Abubakar, H.D., Umar, M. and Bakale, M.A. (2022) 'Sentiment classification: review of text vectorization methods: bag of words, TF-IDF, Word2vec and Doc2vec', *SLU Journal of Science and Technology*, Vol. 4, No. 1, pp.27–33.

Andriushchenko, M. and Flammarion, N. (2020) 'Understanding and improving fast adversarial training', *Advances in Neural Information Processing Systems*, Vol. 33, pp.16048–16059.

Bozyiğit, F., Doğan, O. and Kılınç, D. (2022) 'Categorization of customer complaints in food industry using machine learning approaches', *Journal of Intelligent Systems: Theory and Applications*, Vol. 5, No. 1, pp.85–91.

Cao, Y. and Yang, M. (2023) 'Legal regulation of big data killing: from the perspective of 'Personal Information Protection Law'', *Journal of Education, Humanities and Social Sciences*, Vol. 7, pp.233–241.

Choe, P., Lehto, M.R., Shin, G.C. and Choi, K.Y. (2013) 'Semiautomated identification and classification of customer complaints', *Human Factors and Ergonomics in Manufacturing & Service Industries*, Vol. 23, No. 2, pp.149–162.

Donoghue, S. and De Klerk, H.M. (2009) 'The right to be heard and to be understood: a conceptual framework for consumer protection in emerging economies', *International Journal of Consumer Studies*, Vol. 33, No. 4, pp.456–467.

Du, C. and Huang, L. (2018) 'Text classification research with attention-based recurrent neural networks', *International Journal of Computers Communications & Control*, Vol. 13, No. 1, pp.50–61.

Fan, Q., Han, H. and Sun, X. (2023) 'Environmental false complaint detection based on adversarial transfer learning model', *Journal of Beijing University of Technology*, Vol. 49, No. 9, pp.999–1006.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M. and Lempitsky, V. (2016) 'Domain-adversarial training of neural networks', *Journal of Machine Learning Research*, Vol. 17, No. 59, pp.1–35.

Ghazzawi, A. and Alharbi, B. (2019) 'Analysis of customer complaints data using data mining techniques', *Procedia Computer Science*, Vol. 163, pp.62–69.

Gupta, M., Singh, A., Jain, R., Saxena, A. and Ahmed, S. (2021) 'Multi-class railway complaints categorization using neural networks: RailNeural', *Journal of Rail Transport Planning & Management*, Vol. 20, p.100265.

HaCohen-Kerner, Y., Dilmon, R., Hone, M. and Ben-Basan, M.A. (2019) 'Automatic classification of complaint letters according to service provider categories', *Information Processing & Management*, Vol. 56, No. 6, p.102102.

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C. and Wang, Y. (2021) 'Transformer in transformer', *Advances in Neural Information Processing Systems*, Vol. 34, pp.15908–15919.

Hossain, G., Rashid, M.H.O., Islam, M.R., Sarker, A. and Yasmin, M. (2022) 'Towards mining public opinion: an attention-based long short term memory network using transfer learning', *Journal of Computer and Communications*, Vol. 10, No. 6, pp.112–131.

Khedkar, S. and Shinde, S. (2020) 'Deep learning and ensemble approach for praise or complaint classification', *Procedia Computer Science*, Vol. 167, pp.449–458.

Liu, J., Long, R., Chen, H., Wu, M., Ma, W. and Li, Q. (2024) 'Topic-sentiment analysis of citizen environmental complaints in China: using a Stacking-BERT model', *Journal of Environmental Management*, Vol. 371, p.123112.

Naik, P.K., Prashanth, T., Chandru, S., Jaganath, S. and Balan, S. (2023) 'Consumer complaints classification using machine learning & deep learning', *International Research Journal on Advanced Science Hub*, Vol. 5, No. 5, pp.116–122.

Wang, Y., Zhou, Y. and Mei, Y. (2023) 'A joint attention enhancement network for text classification applied to citizen complaint reporting', *Applied Intelligenc*e, Vol. 53, No. 16, pp.19255–19265.

Weiss, K., Khoshgoftaar, T.M. and Wang, D. (2016) 'A survey of transfer learning', *Journal of Big Data*, Vol. 3, pp.1–40.

Xu, W., Fu, Y.-L. and Zhu, D. (2023) 'ResNet and its application to medical image processing: Research progress and challenges', *Computer Methods and Programs in Biomedicine*, Vol. 240, p.107660.

Yang, Y., Xu, D-L., Yang, J-B. and Chen, Y-W. (2018) 'An evidential reasoning-based decision support system for handling customer complaints in mobile telecommunications', *Knowledge-Based Systems*, Vol. 162, pp.202–210.

Zhang, W., Chen, Q. and Chen, Y. (2020) 'Deep learning based robust text classification method via virtual adversarial training', *IEEE Access*, Vol. 8, pp.61174–61182.

Zhang, Z., Lu, Z., Liu, J. and Bai, R. (2024) 'Medical chief complaint classification with hierarchical structure of label descriptions', *Expert Systems with Applications*, Vol. 252, p.123938.