Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 7
TPCEE 2024

Sentiment Analysis of Long-Term Care Insurance in China Using NLP

Yixuan Li *

School of Insurance, Central University of Finance and Economics, Beijing, China

* Corresponding author: 2023210901@email.cufe.edu.cn

Abstract. Long-term care insurance (LTCI) is important for addressing the needs of aging populations and increasing long-term care demand. China initiated its LTCI pilot in 2016 and expanded it in 2019. This study employs Natural Language Processing (NLP) to analyze the sentiment impact of policyholders on LTCI decisions. By classifying sentiments in comments from various Chinese provinces, this study trained a Long Short-Term Memory (LSTM)-based deep learning model using annotated data and validated its accuracy with a pre-extracted test dataset. The application of the trained model to LTCI comments revealed sentiment distributions across different provinces, highlighting the developmental progress of LTCI since its pilot implementation. This research not only demonstrates the effective use of NLP in understanding sentiment trends but also underscores the significance of transitioning from qualitative to quantitative analysis. The study proposes integrating NLP with behavioral finance and insurance disciplines to enhance the utility and value of insurance products. This approach can guide product development, refine target marketing strategies, and inform policy decisions, contributing to the future expansion and success of LTCI in China.

Keywords: Long-Term Care Insurance (LTCI); Natural Language Processing (NLP); Sentiment Trends; Deep Learning.

1. Introduction

Long-term care insurance (LTCI) is designed to address the needs of aging societies and long-term care. While LTCI systems emerged overseas in the late 1980s, such as in the United States (U.S.) with private products in 1988 followed by public plans, China only began its LTCI pilot program in 2016, expanding it fully by 2019. As a vital financial tool for managing population aging and long-term care needs, LTCI decisions are heavily influenced by policyholders. These decisions are shaped by factors like personal health, family support, economic status, expected long-term care needs, and insurance costs and coverage. Understanding these decision-making processes is crucial for insurers, policymakers, and researchers. Emotions significantly impact decision-making, as Lerner demonstrated in studies on the endowment effect, where sentiment biases lead to higher selling prices than purchase prices. Further research showed that emotions have lasting impacts on decisions, even those unrelated to their initial cause [1].

Daniel Kahneman and Amos Tversky introduced prospect theory explaining how individuals make decisions when faced with uncertainty. The theory suggests that people evaluate gains and losses relative to a reference point, exhibiting loss aversion [2]. Kahneman integrated economic and psychological theories to introduce Decision Utility and Experienced Utility. They argued that the utility typically used in current decision theories is decision utility, which also influences predictions for subsequent decisions [3]. Bracha and Brown analyzed decision-making behavior in the insurance market using the theory of Affective Decision Making (ADM). They introduced ADM as a framework that integrates rational and sentiment cognitive processes to model decision-making under risk [4]. Slovic emphasized the role of emotions in risk assessment. Emotions can distort perceptions of risk severity and cause individuals to deviate from conventional risk assessment models [5].

With the advancement and application of machine learning technologies, integration and analysis of large-scale medical data, they enable machine learning models to effectively identify and quantify individual probabilities and risks for future long-term care needs. For instance, models based on deep learning capture complex disease progression dynamics from multiple dimensions, thereby enhancing prediction accuracy and personalized assessment capabilities. Natural Language Processing (NLP) is a branch of artificial intelligence and linguistics which focused on computers' understanding and generating human language.

Early research was primarily focused on machine translation. With the introduction of statistical learning methods, NLP shifted from rule-based approaches to data-driven methods. Statistical machine translation utilized large bilingual text datasets to discover correspondences between languages [6]. The development of machine learning techniques brought significant advancements to NLP. Researchers began using algorithms to automatically learn features from text data for various language processing tasks [7]. The rise of deep learning techniques, particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory networks (LSTM), greatly propelled NLP forward as they are effective in handling sequential data [8]. The emergence of pre-trained language models like Bidirectional Encoder Representations from Transformers (BERT), trained on vast amounts of text, captured rich language features and significantly improved the performance of various NLP tasks [9]. Currently, trends in NLP include continued advancements in deep learning techniques and the application of NLP in specialized fields such as healthcare and law [10].

The field of emotion detection encompasses the analysis and recognition of emotional states derived from various modalities, including vocal intonations, facial cues, body movements, and written language. Sharma conducted an examination of discourses that blend Hindi and English and applied the Teaching-Learning-based Optimization (TLBO) technique to categorize messages in order of priority, taking into account the emotional content they express [11]. Seal introduced a proficient strategy for detecting emotions by identifying emotional lexicons from a curated list of sentiment keywords and further scrutinizing these words, along with verb phrases and negation constructs. This methodology has proven to outperform other contemporary approaches in terms of efficacy [12]. In studying the impact of policyholder emotions on decisions within LTCI, NLP serves as an excellent tool for quantification and data processing. This study utilizes NLP to classify sentiments in comments regarding LTCI from different province in China.

2. Methodology

2.1. Dataset Description and Preprocessing

Sina Weibo records a huge amount of information in the form of web pages, and the researcher can collect data by using the Weibo open platform or web crawler, i.e., by simulating the user's behavior of searching Weibo independently through simulated login. In order to build a model that can analyze the tendency of residents' attitudes toward LTCI in each province, the dataset used in this paper is divided into the following two parts:

First, the dataset used for model construction comes from the public dataset on Kaggle, which includes the content of comments and labeled two-dimensional sentiments (i.e., "positive" and "negative"), which is reflected in the Weibo comment data that contains 6,000 positive comments and 6,000 negative comments [13]. Second, the dataset used to analyze the sentiment tendency of each province consists of 2000 comments under the term "Long Term Care Insurance", "LTC", "Seven Insurances and Three Funds" and "The Sixth Social Insurance" on Weibo. For the purpose of provincial classification, the comments containing the address of the province and those with clear sentiment tendency (i.e., excluding sentiment-neutral information such as user name and keyword name) are filtered, and the final data obtained is 1665 comments, which are valid [14]. Table 1 shows examples of the comments data.

In the preprocessing stage, this paper first statistically analyzes the length of sentences to determine the specific input parameters of the input sequence in the subsequent steps. After plotting the distribution of sentence lengths, the cumulative distribution function of comment lengths is further calculated and plotted to identify the quartiles in the data. Based on Fig. 1 and Fig. 2, it can be seen that more than 90% of the comments in this dataset are less than 200 characters in length, then 200 is used below as the maximum length of the input sequence for the model.

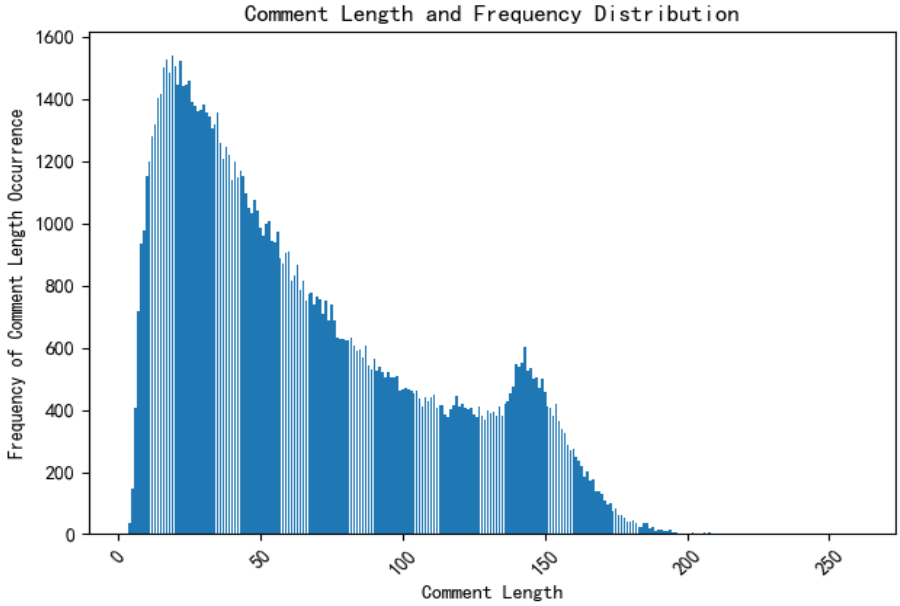

Figure 1. Comment length and frequency of occurrence

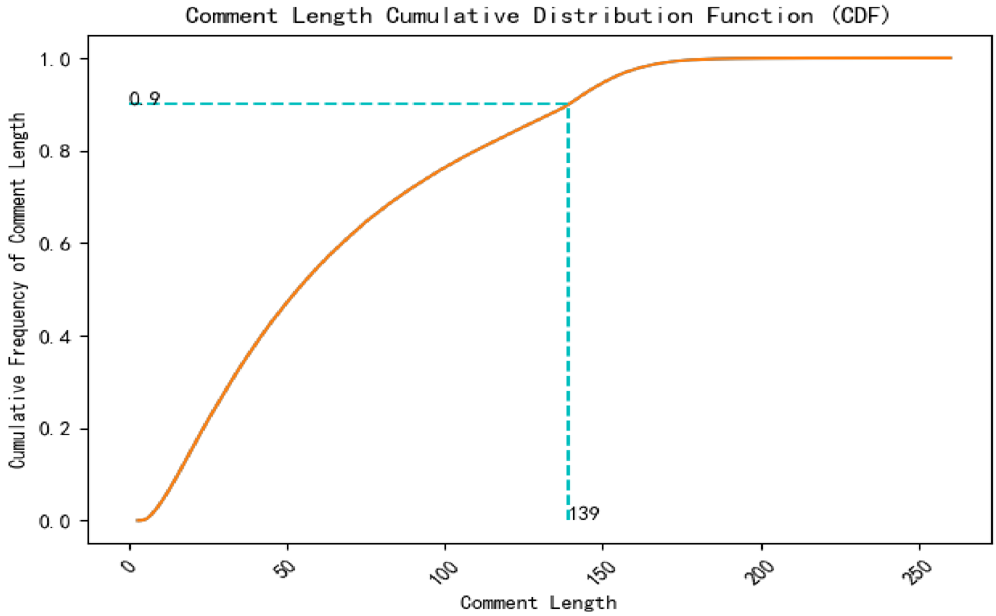

Figure 2. Cumulative distribution function of comment lengths

2.2. Proposed Approach

This paper designed a deep learning model based on LSTM for automatically learning and recognizing the sentiment tendency in text data, and the specific flowchart is shown in Fig. 3.

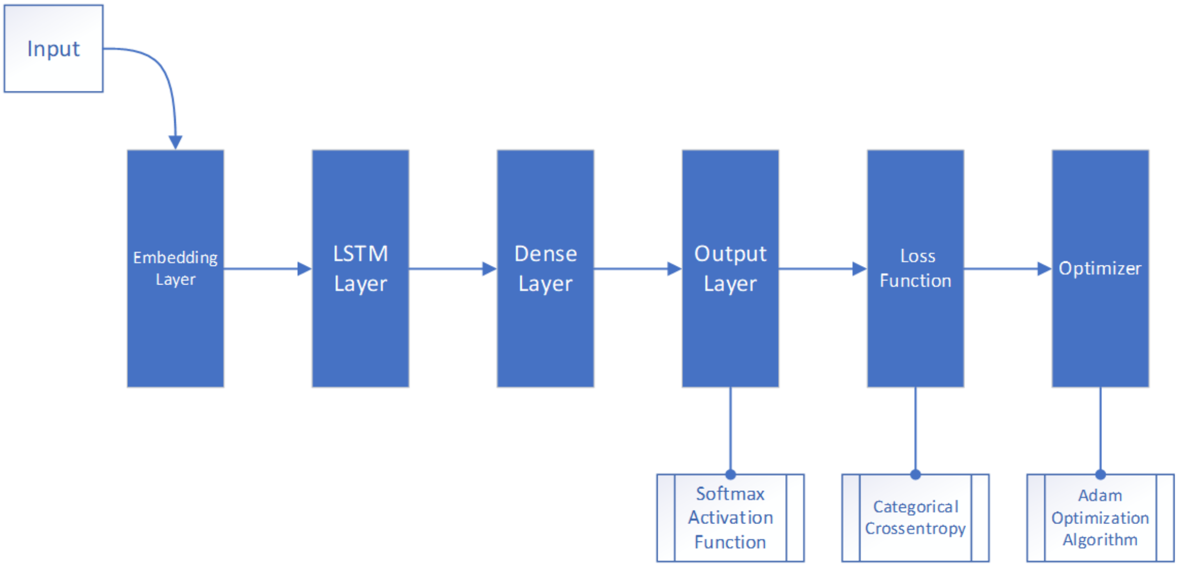

Figure 3. The pipeline of the model

The model's initial component is an Embedding layer, tasked with transforming textual words into numerical form while preserving their semantic associations. Subsequent to this is the LSTM layer, designed to handle sequential inputs and identify patterns indicative of long-range dependencies. To bolster the model's capacity to generalize from training data, a Dropout layer is interposed subsequent to the LSTM, serving to mitigate the potential for overfitting. The subsequent Dense layer is responsible for generating the probabilistic distribution across sentiment categories, employing a SoftMax activation to ensure the outputs sum to one. The model's loss function is the Categorical Cross-entropy, and it utilizes the Adam optimizer, recognized for its efficacy across a spectrum of deep learning applications.

2.2.1. Deep learning model.

Common deep learning models include Transformer, BERT, Convolutional Neural Networks (CNN), RNN, LSTM, and their derivatives. Considering that the text of Weibo comments is usually short and contains non-standard text elements such as emoticons, it is more appropriate to use a model that can capture contextual information and process sequential data, while LSTM is an improved version of RNN, which solves the gradient problem in long sequence training by introducing a gating mechanism (as shown in Fig. 4), especially solves the gradient disappearance and gradient explosion problems of RNN when processing long sequences, and performs well in the task of sentiment analysis also with excellent performance in capturing long semantic dependencies. Therefore, in this paper, author designs a deep learning model based on LSTM.

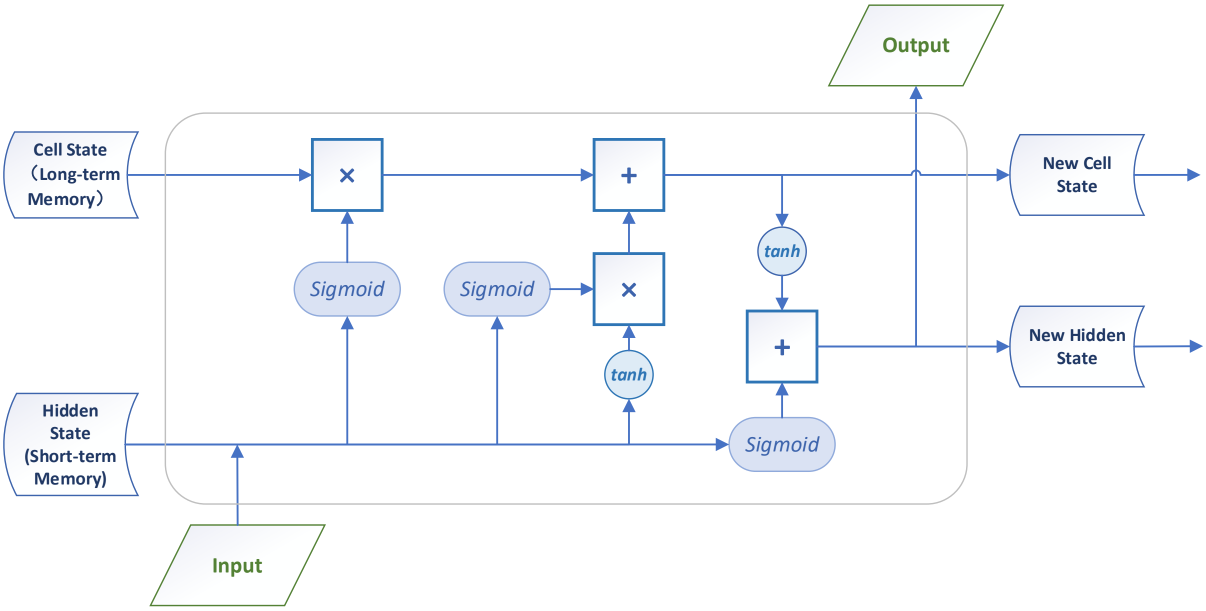

Figure 4. The pipeline of LSTM

2.2.2. Loss function.

The Categorical Cross-Entropy Loss function is an algorithmic metric predominantly employed for addressing classification challenges, particularly in scenarios where the target labels are encoded in a one-hot format. This function assesses the divergence between the predicted probability distribution generated by the model and the actual probability distribution associated with the correct label. The underlying mathematical rationale can be articulated as follows: Given a random variable *X* and two probability distributions *P* and *Q*, where *P* is the true distribution and *Q* is the distribution predicted by the model, the cross- entropy is defined as: $H(P,Q) = -\sum_x P(x)\, log\, Q(x)$, where summation is done for all possible events *x*, *P(x)* is the probability of event *x* occurring (true distribution) and *Q(x)* is the probability that the model predicts event x to occur (the predictive distribution).

In classification problems, the cross-entropy loss function is typically used to measure the difference between the probability distribution of the model output and the one-hot coded representation of the target label. For a multi-category classification problem, the loss function can be expressed as:

$$L = -\sum_{c=1}^{C} y_c \log p_c \tag{1}$$

Where *C* is the total number of categories, y_c is the one-hot coding of the target label and p_c is the probability that the model predicts that the sample belongs to category c. Sentiment analysis typically involves categorizing textual data into multiple sentiment categories, such as positive, negative, or neutral. Cross-entropy loss is designed for multi-category classification problems and can measure the difference between the probability distribution predicted by the model and the actual labels. Also, the cross-entropy loss function is continuously fictitious, which allows the gradient-based optimization algorithm Adam to be used to train the model for the subsequent parts of this paper.

2.2.3. Optimization algorithm.

Adaptive Moment Estimation (Adam) is a popular gradient descent optimization algorithm that combines the advantages of Momentum and Adaptive Learning Rate Optimization (AdaGrad). The Adam optimizer is widely used for its effective adaptive learning rate adjustment and robustness to non-smooth objectives. The specific principle is: firstly, the exponentially weighted average of the gradient is calculated by first-order moment estimation, which is equivalent to the momentum term and reduces the effect of small fluctuations. Secondly by second-order moment estimation: the exponentially weighted average of the squared gradient is computed, similar to the root mean square prop (RMSProp), for adjusting the learning rate.

$$k_t = \gamma_1 * k_{t-1} + (1-\gamma_1) * \tau_t \tag{2}$$

$$j_t = \gamma_2 * j_{t-1} + (1-\gamma_2) * (\tau_t * \tau_t) \tag{3}$$

$$\delta_{t+1} = \delta_t - \alpha * \frac{k_t}{\sqrt{j_t} + \varepsilon} \tag{4}$$

Where τ_t is the gradient at time step t, k_t is the first order moment estimate of the gradient, j_t is the second order moment estimate of the gradient, γ_1, γ_2 are hyperparameters controlling the exponential decay rate, α is the learning rate and ε is a very small constant for numerical stability. For the purpose of sentiment analysis in this paper, sentiment analysis datasets may have different feature scales, which can be handled more effectively by Adam's adaptive learning rate. And for huge datasets in natural language processing, Adam's computational complexity is linearly related to the number of parameters, which is suitable for large-scale parameter optimization problems, and it can usually converge to the optimal solution faster than standard gradient descent. As well as having better robustness to non-smooth objectives and noisy data.

3. Result and Discussion

3.1. Model Training Performance

Using the previously described deep learning model based on LSTM, author divided the entire dataset into training and testing sets with a 9:1 ratio. According to the test results, the Receiver Operating Characteristic (ROC) chart (Fig. 5) illustrates the classification performance of the model in detail. The ROC curve starts at (0.0, 0.0), indicating no classification ability, and as the false positive rate increases, the true positive rate gradually increases, ultimately approaching (1.0, 1.0), which represents perfect classification. The area under the curve (AUC) is 0.98, a very high value, indicating that the model has excellent classification ability and can distinguish between different categories effectively.

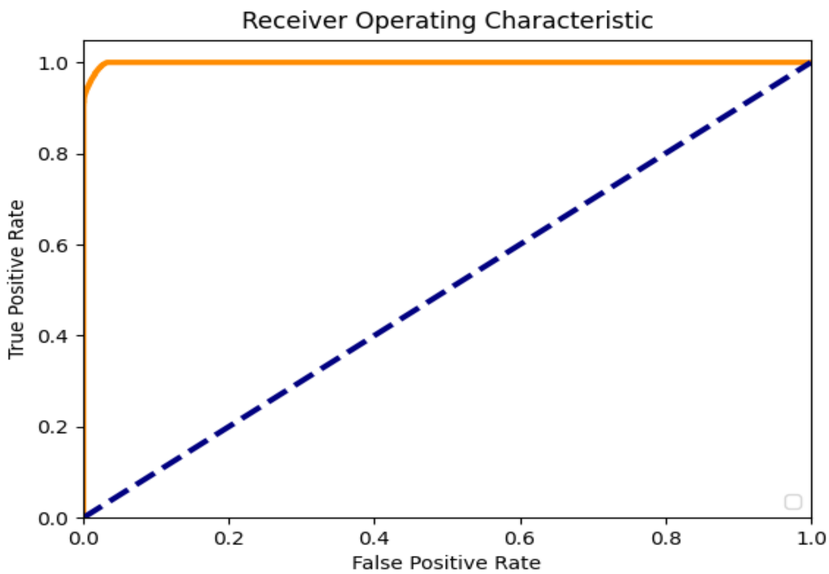

Figure 5. Receiver operating characteristic

Simultaneously, the Model Loss chart (Fig. 6) shows the change in loss values during the training process through a line chart. Both training loss and validation loss decrease from 0.11 and gradually reduce as the number of epochs increases, eventually stabilizing around 0.05. This indicates that the model training is effective and has strong generalization capabilities. The Model Accuracy chart (Fig. 7) displays the accuracy changes during the training and validation processes. Both training accuracy and validation accuracy are very high, approaching 0.984, and both increase with the number of epochs, indicating that the model performs excellently on both the training and validation sets.

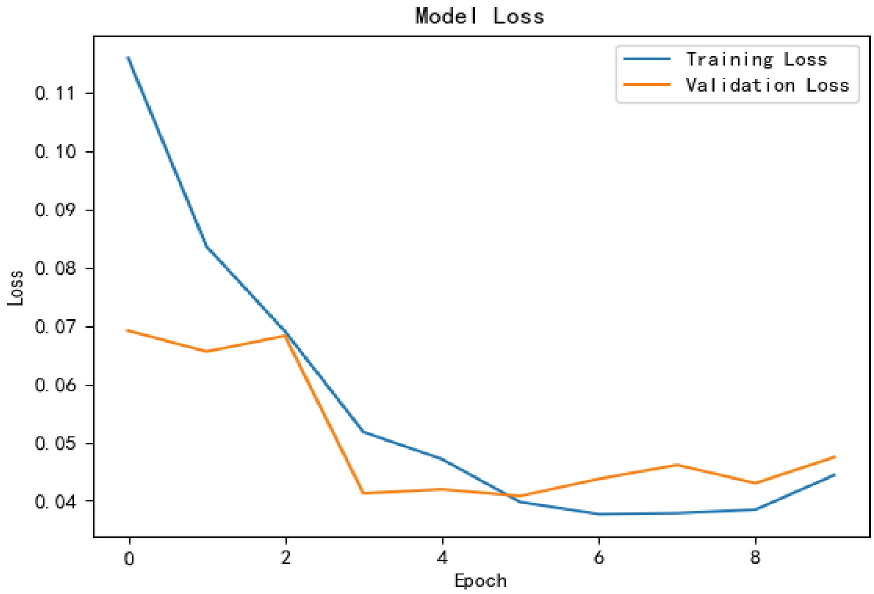

Figure 6. Model loss

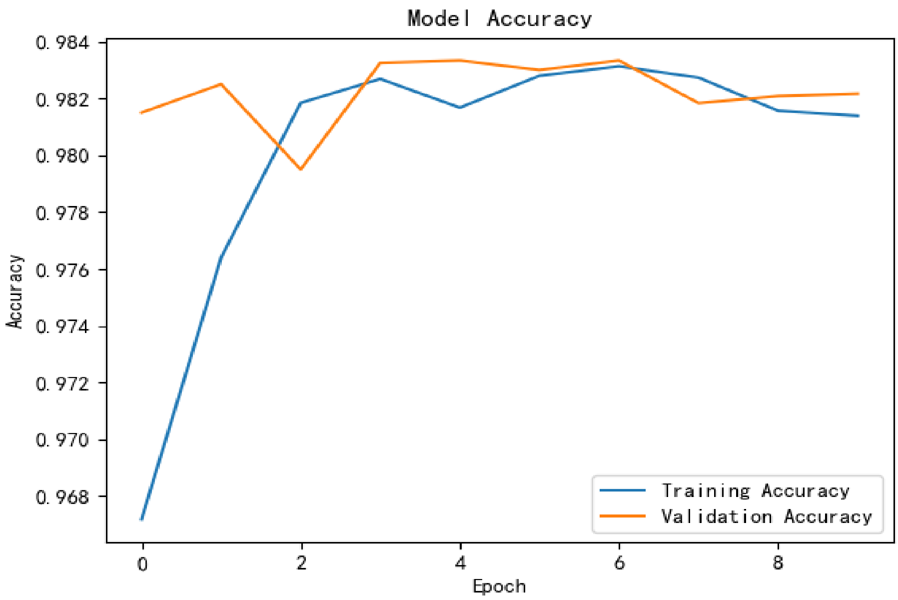

Figure 7. Model accuracy

Overall, according to Table 1, for the "positive" category, the precision is 1.00 (i.e., all samples predicted as "positive" are indeed "positive"), the recall is 0.97, the F1 score is 0.98, and the support is 6077. For the "negative" category, the precision is 0.97, the recall is 1.00 (i.e., all "negative" samples are successfully identified), the F1 score is 0.98, and the support is 5923. These metrics show that the model has high classification precision and recall for both categories, with an F1 score close to 1.00, indicating a good balance between precision and recall.

Table 1. Performance Measures

	Positive	Negative
precision	1.00	0.97
recall	0.97	1.00
f1-score	0.98	0.98
support	6077	5923

3.2. Sentiment Classification Results of LTCI Dataset

In the realm of data, author utilized a total of 1,665 valid comments pertaining to the terms "Long Term Care Insurance", "LTCI", "Seven Insurances and Three Funds" and "The Sixth Social Insurance" on the Weibo platform. Fig. 8 illustrates the distribution of these comment data across various provinces, where regions with larger populations or higher levels of development, such as Beijing, Shanghai, Shandong, and Jiangsu, exhibit a greater degree of attention and discussion regarding LTCI. Concurrently, areas such as Xinjiang, Ningxia, and Hong Kong, due to their usage habits of the Weibo platform, as well as factors like the extent of policy dissemination and economic development, have an insufficient number of comments. In subsequent analyses, this paper will take care to differentiate the data discrepancies introduced by these provinces.

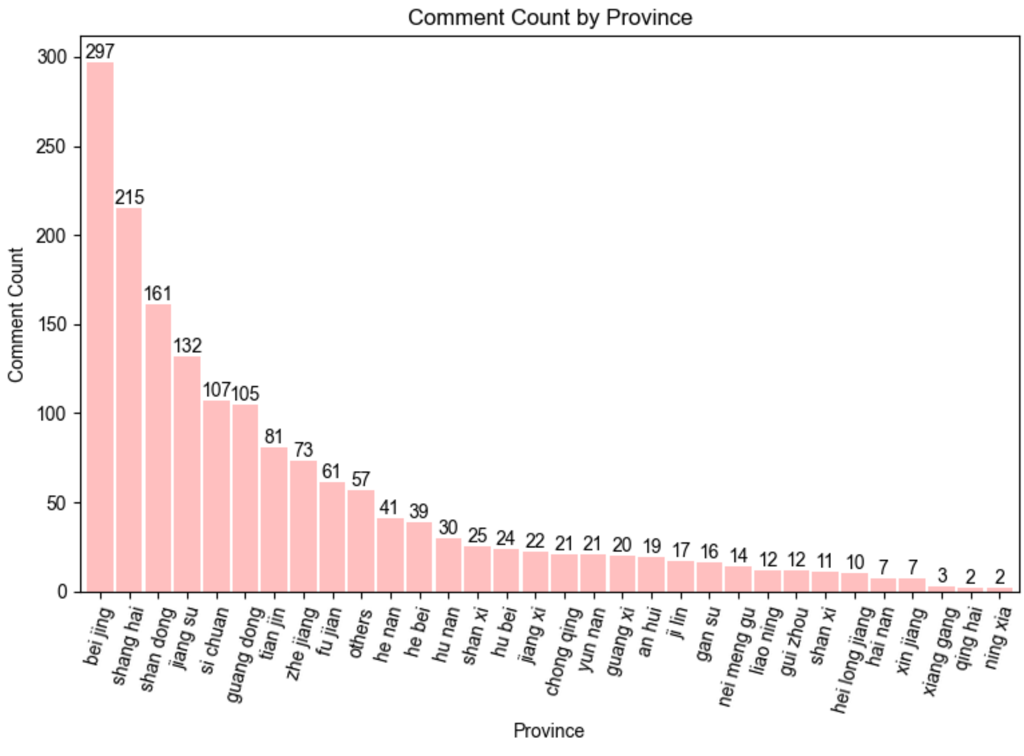

Figure 8. Regional distribution of comment data

Utilizing an LSTM model to predict sentiment from Weibo comment data, one can observe the sentiment tendencies of comments in different regions as depicted in Fig. 9, where the overall sentiment leans predominantly towards the positive. Additionally, Fig. 10 presents the top 15 areas with a positive sentiment tendency. It is important to note that regions such as Xinjiang, Qinghai, and Hong Kong, due to a scarcity of comment data, exhibit an excessively high proportion of positive sentiment. Beyond these exceptions, areas like Jiangxi, Hubei, and Tianjin demonstrate a relatively high proportion of positive sentiment within the normal range of data.

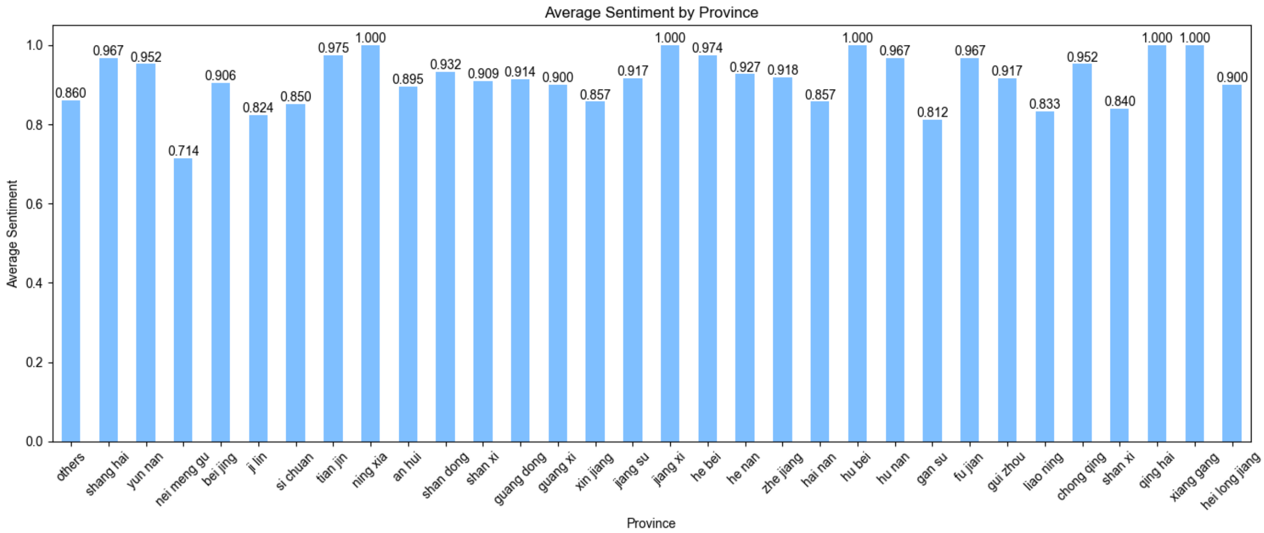

Figure 9. Sentiment tendencies of comments in different areas

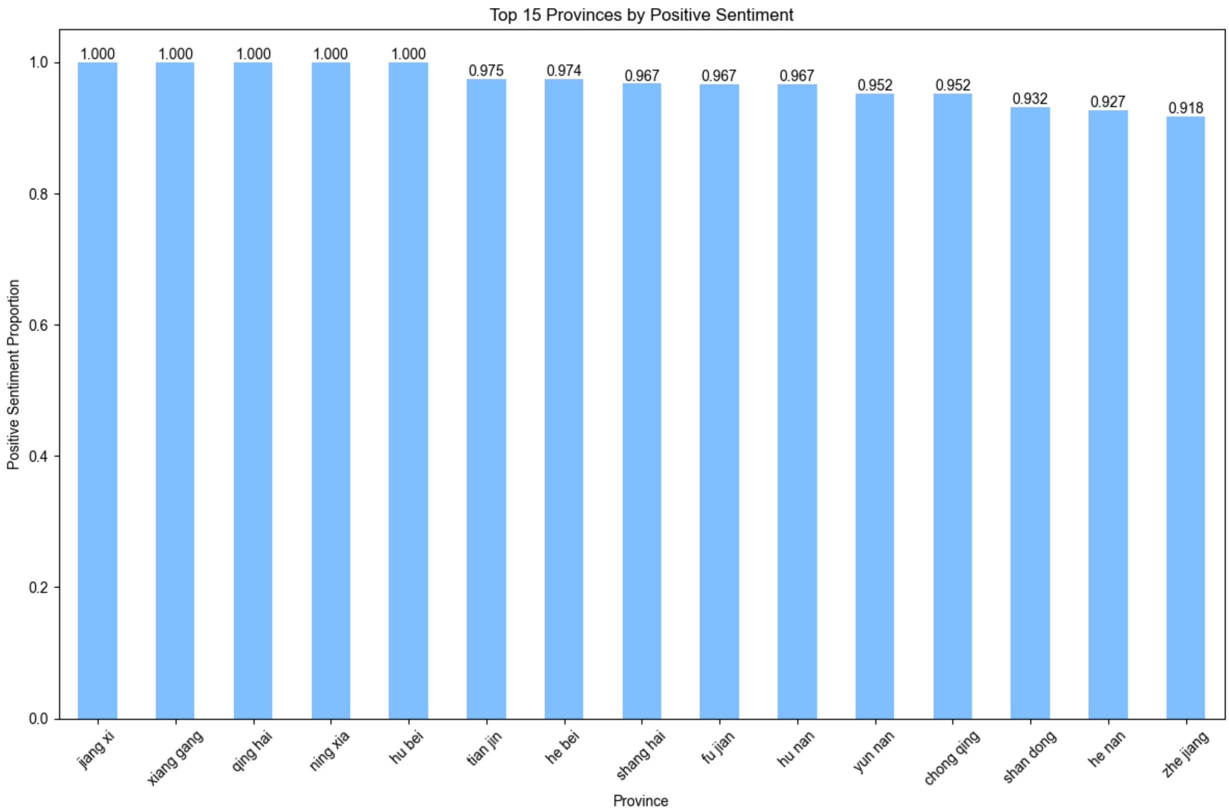

Figure 10. Top 15 areas with positive sentiment tendencies

In 2016, China proclaimed the commencement of the inaugural cohort of pilot cities for the LTCI scheme. This group of 15 cities includes Chengde, Changchun, Qiqihar, etc.

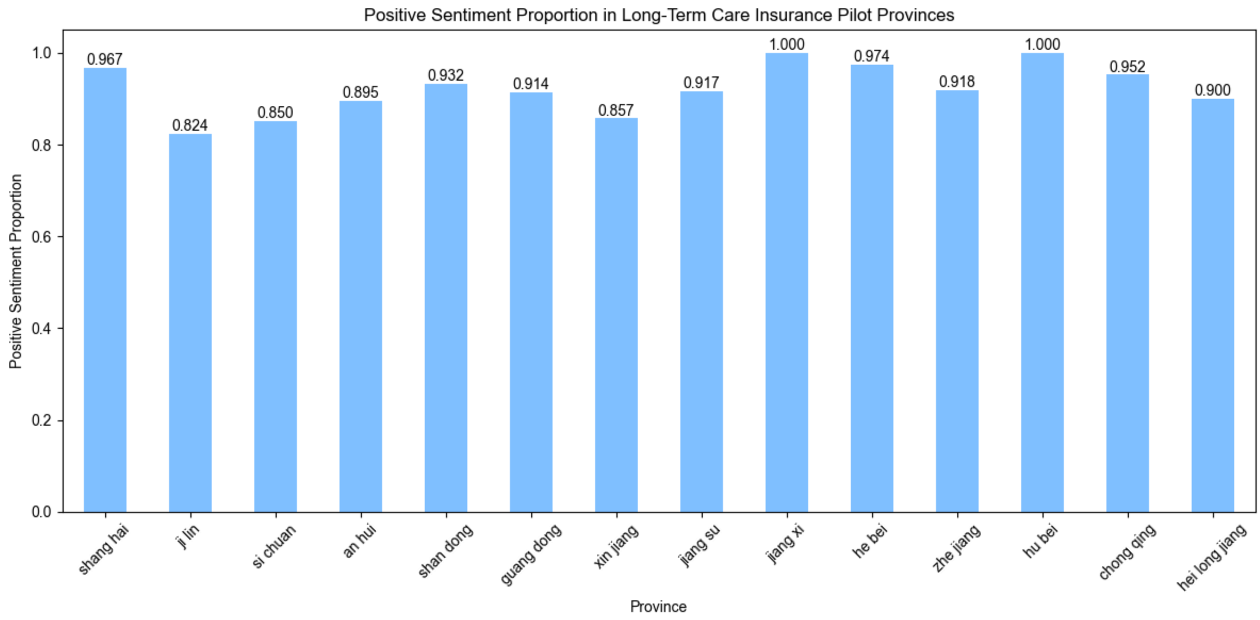

Figure 11. Sentiment tendencies in LTCI pilot areas

Examination of Fig. 11 reveals that among the regions where the 15 pilot cities are located, seven regions are categorized within the top 15 areas exhibiting positive sentiment tendencies. This observation intimates that since the inception of the LTCI system, these pilot cities have been proactively engaged in the exploration and innovation of tailored approaches to the system. Such initiatives have, to a significant extent, bolstered the populace's favorable perception and endorsement of the LTCI. Consequently, this has established a solid groundwork for the system's expanded adoption throughout the entirety of China.

4. Conclusion

This study successfully leverages NLP to analyze the sentiment tendencies of residents in the pilot provinces and cities participating in LTCI. Initially, the author trained a LSTM based deep learning model using annotated experimental data. The model's classification accuracy and recall were then evaluated using a pre-extracted 10% of the annotated test data. Finally, the trained model was applied to predict sentiments from comments on LTCI, revealing sentiment tendencies with typical distribution characteristics across different provinces. This qualitative analysis underscores the developmental achievements of LTCI in China since its pilot implementation. However, this study only preliminarily correlates the pilot implementation of LTCI with the sentiment tendencies of residents without detailed quantification or exploration of underlying causes. Transitioning from qualitative to quantitative analysis is necessary, and integrating NLP with behavioral finance or behavioral insurance disciplines is a crucial development direction.

In the existing field of behavioral finance, most empirical research based on big data or artificial intelligence is limited. Traditional behavioral finance often relies on small-sample surveys or experimental games to investigate risk and decision factors. Artificial intelligence can augment traditional behavioral finance by leveraging vast amounts of real data to enhance theoretical advancements. For insurance product developers, the utility of insurance products, especially those with low short-term payout probabilities, can largely be reflected in the emotions of policyholders or insured individuals. NLP can assist developers in understanding two key areas: the sentiment feedback different products receive in the market, guiding product development and strategy formulation, and the sentiment characteristics of high-expenditure customers, aiding in user profiling and target marketing. Whether in the study of financial theories or the formulation of macro policies, sentiment analysis is a powerful tool for addressing irrational decision-making and quantifying subjective indicators. It provides significant guidance for both insurance companies and individual consumers.

References

[1] Lerner J.S. Small D.A. Loewenstein G. Heart strings and purse strings: Carryover effects of emotions on economic decisions. Psychological science, 2004, 15 (5): 337 - 341.

[2] Kahneman D. Tversky A. Prospect theory: An analysis of decision under risk. Handbook of the fundamentals of financial decision making, 2013: 99 - 127.

[3] Kahneman D. Sugden R. Experienced utility as a standard of policy evaluation. Environmental and resource economics, 2005, 32: 161 - 181.

[4] Bracha A. Brown D.J. Affective decision making: A theory of optimism bias. Games and Economic Behavior, 2012, 75 (1): 67 - 80.

[5] Slovic P. Finucane M.L. Peters E. et al. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk and rationality. The feeling of risk. Routledge, 2013: 21 - 36.

[6] Tillmann C. Vogel S. Ney H. et al. Accelerated DP based search for statistical translation. Eurospeech, 1997: 2667 - 2670.

[7] McCallum A. Nigam K. A comparison of event models for naive bayes text classification. In AAAI-98 workshop on learning for text categorization, 1998, 752: 41 - 48.

[8] Hochreiter S. Schmidhuber J. Long short-term memory. Neural Comput, 1997, 9 (8): 1735 – 1780.

[9] Devlin J. Chang M.W. Lee K. Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. 2018, arXiv preprint: 1810.04805.

[10] Rospocher M. van Erp M. Vossen P. Fokkens A. Aldabe I. Rigau G. Soroa A. Ploeger T. Bogaard T. Building event-centric knowledge graphs from news. Web Semantics: Science, Services and Agents on the World Wide Web, 2016.

[11] Sharma S. Srinivas P.Y.K.L. Balabantaray R.C. Emotion Detection using Online Machine Learning Method and TLBO on Mixed Script. In Proceedings of Language Resources and Evaluation Conference, 2016: 47 - 51.

[12] Seal D. Roy U.K. Basak R. Sentence-level emotion detection from text based on semantic rules. Advances in intelligent Systems and computing, 2020, 933.

[13] Kaggle dataset, “Weibo comment data”, 2023, https://www.kaggle.com/datasets/yisywanlee/weibo-comments.

[14] Kaggle dataset, “final data”, 2023, https://www.kaggle.com/datasets/yisywanlee/weibo-comments-regarding-long-term-care-insurance.