Integrated modelling of automobile maintenance expert system based on knowledge graph

Gen Liu¹,², Ge Hong¹,², Mengdie Huang¹,², Tangbin Xia¹,²,³ and Zheng Chen²

¹Chinese Institute for Quality Research, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
²School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
³E-mail: xtbxtb@sjtu.edu.cn

Abstract. Using online information resources to build knowledge bases to provide knowledge answering services would help auto companies or third-party platforms to gain competitive advantages. Therefore, a construction plan of automobile maintenance expert system based on knowledge graph was proposed by integrated modelling. In terms of the entity recognition algorithm, the BM LSTM (Boyer-Moore Long Short-Term Memory) algorithm was proposed by integrating hidden Markov model, Conditional Random Field (CRF), Bi-directional Long Short-Term Memory (BiLSTM), BiLSTM-CRF and Lattice LSTM, which improved the accuracy index F1-score. In terms of the text quality evaluation algorithm, a secondary text quality evaluation system was designed. It evaluated the matching quality of the problem based on the word toolkit Synonyms and Levenshtein Distance algorithm. And it evaluated the quality of the answer text based on the TF-IDF (Term Frequency-Inverse Document Frequency) similarity algorithm and centered on completeness, accuracy, reliability, and argument strength. Finally, experiments are carried out on the proposed model and algorithm to prove its effectiveness.

1. Introduction
With the rapid development of technologies, expert systems have become a hot research field. It receives the text of the question, uses the knowledge base to find the answer, and returns it to the user [1]. It enhances the convenience for users to acquire knowledge, saves information screening time, and improves information quality [2]. But the research on the automobile maintenance expert system model is in its infancy. Zhang Jinming built a knowledge graph based on network resources and implements a question-and-answer system in the automotive field [3]. However, the system can only answer issues about brands, models, and prices. Based on the "ASP/SaaS-based manufacturing industry value chain collaboration platform", Zhang Qiang carried out research on the after-sales maintenance data of tens of thousands of auto companies in the past 10 years and proposed an automatic question-and-answer system solution [4]. But its method relies on high-quality after-sales maintenance data, and does not distinguish the importance of keywords in each category.

In general, there are some problems in the research of automobile maintenance expert system models, such as insufficient functions, rough models, poor scalability, and poor application effects. Therefore, this article carries out the above research based on knowledge graph, which will improve the knowledge base in the automotive field and help car users quickly and fully understand the relevant knowledge.
2. Framework and related algorithms

2.1. Framework
This article proposes a set of model design schemes. The framework consists of two modules: the knowledge graph module and the human-computer interaction module. (1) Knowledge graph module. First, 25,397 relevant valid data were crawled and processed. And then, the data files were imported into the graph database Neo4j. (2) Human-computer interaction module. The core function of this module is to receive questions and return answers. The framework is shown in Figure 1:

![Framework Diagram](image)

**Figure 1.** The framework of the expert system model.

2.2. Name entity recognition algorithm
The accuracy of the entity recognition algorithm is critical to the model. Classical statistical models have almost all been used, such as Hidden Markov Model (HMM), Maximum Entropy Model (ME), Conditional Random Field (CRF) and Support Vector Machines (SVM) [5]. Among them, HMM algorithm and CRF algorithm perform well. With the rise of deep learning, many scholars try to apply deep learning techniques to entity recognition, such as Bi-directional Long Short-Term Memory (BiLSTM). In fact, adding CRF layer can add constraints, which can improve its rationality, such as BiLSTM-CRF. In addition, adding word segmentation information can also promote the improvement of recognition accuracy, such as Lattice LSTM (Lattice Long Short-Term Memory) [6].

However, different methods have their advantages and disadvantages. Therefore, this article selects five Chinese named entity recognition algorithms that perform well, including: HMM, CRF, BiLSTM, BiLSTM-CRF and Lattice LSTM. The BM LSTM (Boyer-Moore Long Short-Term Memory) algorithm is then designed. The label score calculation formula is as follows:
\[
\max_{i}(s_{ik}) = \frac{1}{n} \sum_{j=1}^{n} c_{ijk} + \alpha p(w_i | w_{k-1})
\]

where \(i\) represents the label number, \(j\) represents the algorithm number, and \(k\) represents the character number. The value of \(c_{ijk}\) is in \{0,1\}. When it is 1, it means that the \(k\)-th character is predicted by the \(j\)-th algorithm as the \(i\)-th type label. \(p(w_i | w_{k-1})\) means the probability that the \(k\)-th character is predicted to be the \(i\)-th type label when the label of the \((k-1)\)-th character is \(w_{k-1}\). \(\alpha\) is the constraint strength coefficient, which represents the influence strength of the prediction result of the previous character. Its temporary value is 0.5. After obtaining the prediction scores of various tags of the \(k\)-th character, the label with the highest score is selected as the prediction result.

2.3. Text quality assessment algorithm

Because there are some problems in user-generated data, such as high redundancy, low credibility, incompleteness, and uneven quality levels. This paper designs a two-level text quality evaluation system. The overall process of text quality evaluation is shown in Figure 2:

![Text quality assessment flowchart.](image)

**Figure 2.** Text quality assessment flowchart.

First, five types of entities, including objects, conditions, attributes, parameters and questions, are designed. The weights of objects, conditions, attributes, and parameters are set to 0.3, 0.5, 0.7, and 0.9 respectively. Secondly, considering that there are different ways of expressing the same entity, this article uses Synonyms Toolkit to expand the entities. Next, this paper uses the Levenshtein Distance algorithm to extend the user input entity. An example is shown in Figure 3:

![Entity recognition and entity extension example chart.](image)

**Figure 3.** Entity recognition and entity extension example chart.
The question scoring formula is:

$$\max(q_i) = \sum_{j=1}^{k} w_{ij} \cdot n_{ij} \cdot s_{ij} \quad i = 1, \ldots, m; \ j = 1, \ldots, k$$  \hspace{1cm} (2)$$

where \(i\) represents the matched question number, \(j\) represents the entity number pointing to the matched question. \(q_i\) represents the score of the \(i\)-th matched question, \(w_{ij}\) represents the weight of the entity category pointing to the question, \(n_{ij}\) represents the similarity of the extended entity of the corresponding knowledge graph entity node, and \(s_{ij}\) represents the similarity of the extended entity of the corresponding user input entity. The similarity of the entity itself is recorded as 1.

Since the number of entities participating in the matching is small, the number of highest-scoring questions is big, and there are multiple answers to each question, it is necessary to grade the answer text quality. First, we need to examine the factors affecting text quality. The information adoption model assumes information seekers will adopt the perceived useful information [7]. In fact, the key factors can be classified into review characteristics, reviewer characteristics, and reader characteristics [8]. In addition, other factors that influence text quality include the length of the comment, the release time, and the emotional tendency of the comment [9]. But in fact, the emotional intensity of reviews has no significant effect on review quality [10]. Therefore, this article designed the answer text quality evaluation system around four aspects of completeness, accuracy, reliability, and discussion intensity.

The index system is shown in Table 1:

| First-level index | Second-level index | Measurement method | Index nature | Remarks |
|-------------------|--------------------|--------------------|--------------|---------|
| Completeness      | The depth of the answer | the length of the text | Positive diminishing marginal effect | |
| Focus             |                     | the number of entities in the answer text | Neutral | assume that 2 is best |
| Accuracy          | Text similarity     | the similarity between user question text and answer text | Positive | using TF-IDF algorithm |
| Readability       |                     | the number of words in sentences | Neutral | assume 15 is best |
| Reliability       | Reviewer characteristics | the number of fans | Positive diminishing marginal effect | |
| Discussing intensity | popularity           | the number of views and release time | Positive diminishing marginal effect marginal effect unchanged | |
|                   | Recognition         | the number of likes | Positive | unchanged |

And then answer text quality evaluation calculation formula is designed, where \(\alpha, \beta, \chi, \delta, \varepsilon\) are undetermined parameters.

$$Score = \alpha \cdot Z[\ln(\text{Len}(a))] \cdot Z[100 - (\text{Enti}(a) - 2)^2] + \beta \cdot Z[500 - \frac{\text{WordC}(a)}{\text{SenC}(a)} - 15]^2$$

$$+ \chi \cdot Z[\ln(\text{Re Fan}(a)) \cdot (10 + \text{Like}(a)) \cdot \text{BroN}(a)] \cdot Z[e^{\frac{1}{\text{Dis}(a)}}] + \delta \cdot \text{Sim}(a) + \varepsilon$$  \hspace{1cm} (3)$$

where \(z[]\) indicates that the data is standardized to eliminate the influence of the difference in data dimensions and value ranges.
In terms of text similarity calculation, the text similarity algorithm based on TF-IDF (Term Frequency-Inverse Document Frequency) is selected. In terms of parameter determination, the multiple regression algorithm and Back Propagation neural network are compared and analyzed.

3. Analysis of results

3.1. Entity recognition algorithm experimental results

In order to select an algorithm for entity recognition in the field of automobile maintenance, this paper compared the accuracy of five commonly used entity recognition algorithms, HMM, CRF, BiLSTM, BiLSTM-CRF and Lattice LSTM, and integrated them to improve accuracy. Among them, the labeled training set is composed of 500 questions, with a total of more than 29,000 labels, and the ratio of training data, development data, and test data is 7:1.3:1.7. Evaluation indexes include accuracy, precision, recall, and F1-score. The formula is shown in Table 2:

| Name | Accuracy | Precision | Recall | F1-score |
|------|----------|-----------|--------|----------|
| TP + TN | TP + FP | TP | \( \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \) |

where TP means identifying a positive example as a positive example, TN means identifying a negative example as a negative example, FP means identifying a negative example as a positive example, and FN means identifying a positive example as a negative example. The F1-score is calculated based on the entity. Figure 4 shows the entity recognition results. Among them, the learning rate of BiLSTM and BiLSTM-CRF is set to 0.002, the batch size is 10, the iteration period is 30, the number of model steps is 5, and the word vector dimension and the hidden layer vector dimension are both 128. The initial learning rate of Lattice LSTM is 0.015, the decay rate is set to 0.05, the iteration period is 30 times, the batch size is 10, the hidden layer dimension is 200, and the dropout is 0.1.

![Figure 4. Entity recognition algorithm accuracy chart.](image)

It can be seen from Figure 4 that the entity recognition effect of the integrated algorithm BM LSTM proposed in this paper is better than that of a single entity recognition algorithm. After the algorithm is integrated, it is found that F1-score has been improved, from 0.81 to 0.83, and the recall...
rate is increased from 0.84 to 0.87. It can be seen that the integrated learning effect is significant, which helps to improve the application effect of the model.

3.2. Text quality evaluation algorithm experiment results

Based on the purpose of returning high-score answers to users, this article trains the parameters of the answer scoring formula. A total of 33 questions were constructed, and the text quality of the returned 2,366 answers was evaluated. The score interval was set between 0 and 1. The training set and the test set are randomly divided at a ratio of 8:2, and 1891 pieces of training data and 473 pieces of test data are obtained. Numerical experiment results showed that the Root Mean Squared Error (RMSE) is 0.1783. In order to ensure the rationality of the training algorithm selection, this paper also uses the machine learning method BP (Back Propagation) neural network to train the model. Set the number of hidden layer neurons to 50, the training period to 1000, and the learning rate to 0.001. The prediction error distribution of the multiple regression model and the BP neural network model is shown in Figure 5 and Figure 6:

![Figure 5. Multiple regression prediction error distribution chart.](image1)

![Figure 6. BP neural network prediction error distribution chart.](image2)

It can be seen from Figure 5 and Figure 6 that the RMSE of the multiple regression prediction is smaller and the prediction accuracy is higher. Among them, the prediction error refers to the absolute value of the difference between the prediction score and the annotation score. Compared with the multiple regression prediction models, the BP neural network model has a larger prediction error. In addition, considering its low interpretive performance of neural networks, high risk of overfitting, and difficulty in formulating expressions, it is reasonable to use multiple regression models as training models. Finally, after parameter estimation is performed through model training, the answer text quality evaluation formula is determined as:

\[
Score = 0.375 \times Z[\ln(Len(a))] \times Z[100 - (Enti(a) - 2)^2] - 0.078 \times Z[500 - \left(\frac{WordC(a)}{SenC(a)} - 15\right)^2] \\
+ 0.153 \times Z[\ln(Re Fan(a)) \times (10 + Like(a)) \times BroN(a)] \times Z[e^{\frac{1}{Len(a)}}] + 1.159 \times Sim(a) + 0.158
\]  

(4)

It can be seen from the above formula that text similarity has the greatest impact on the quality score of text answers, and its coefficient reaches 1.159, which is much higher than other factors’ coefficients. The second most affected factor is the completeness, with its coefficient of 0.357. In addition, the readability of the text has little effect on the quality of the answer.

3.3. Expert system model experiment results

In order to test the accuracy of the expert model, this paper extracted 90 automobile maintenance questions from the web page. The results showed that the replies to 67 questions basically met
expectations. The accuracy reached 73.6%. In terms of expert system model research, the accuracy of most systems is between 60% and 80%. For example, the accuracy of Answer-Bus system proposed by Zheng Zhiping is 64.18%, and the accuracy of START, LCC, IONAUT and QuASM systems is not more than 50% [11]. “Belt and Road” question-and-answer system proposed by Chen Jinghao et al has an accuracy of 70% for the “other questions” category [2]. In addition, most articles do not provide experimental accuracy. In summary, the model in this article is effective and has certain application values.

4. Conclusions
With the technique innovations, enterprises increasingly value service-oriented manufacturing to maintain competitiveness and meet customer needs [12]. Providing knowledge services is conducive to enhancing user stickiness, thus realizing service-oriented manufacturing. Therefore, this article uses knowledge graph and natural language processing to explore the fusion between network information resources and the automobile knowledge bases. This paper innovated the entity recognition algorithm for user-generated short texts by integrating HMM, CRF, BiLSTM, BiLSTM-CRF and Lattice LSTM, which increased the F1-score of entity recognition by 2%. In addition, a text quality evaluation system is designed and the experimental results proved the rationality of the system. Finally, the experimental results of the proposed model show that the model can basically meet the expectation, with an accuracy of 73.6%. In the future, we can continue to increase the data volume of the knowledge graph and iteratively optimize the model to improve the application effect of the model.

Acknowledgements
This research is supported by National Natural Science Foundation of China (51875359), Natural Science Foundation of Shanghai (20ZR1428600), Shanghai Science & Technology Innovation Center for System Engineering of Commercial Aircraft (FASE-2021-M7), Ministry of Education-China Mobile Research Foundation (CMHQ-JS-20190003).

References
[1] Guo T Y, Peng M, Yi M L, Mao W Y, Hu X C and Wei G G 2019 J. Wuhan Univ.(Nat. Sci. Ed.). 65 417-426
[2] Chen J H, Zeng Z and Li G 2020 Library And Information Service 62 95-105
[3] Zhang J M 2019 Design and Practice of Automatic Question-answering in Automobile Field Based on Knowledge Graph (Wuhan: Huazhong University of Science & Technology)
[4] Zhang Q 2019 Study of the Case-matching Model Supporting the Cars’ Maintenance with Question-answering Method (Chengdu: University of Electronic Science and Technology of China)
[5] Liu L and Wang D B 2018 Journal of The China Society for Scientific and Technical Information 37 329-340
[6] Zhang Y, Wang Y and Yang J 2020 IEEE-ACM. T. AUDIO. SPE. 28 1506-19
[7] Zhao L N and Han D M 2015 Statistics and Decision 16 108-111
[8] Zhang X and Zhu Z Z 2017 Journal of Commercial Economics 17 82-84
[9] He Y S and Li N 2016 Journal of Intelligence 35 202-206+201
[10] Yan J Y, Zhang L and Zhang L 2012 Information Science 30 713-716+719
[11] Zheng Z P 2002 Proc. 2nd. Int. Conf. on Human Language Technology Research (San Francisco: Margan Kaufmann) pp 399-404
[12] Xia T B, Dong Y F, Xiao L, Du S C, Pan E S and Xi L F 2018 Reliability Engineering & System Safety 178 255-268