Research on E-Commerce Automatic Question Answering System Model Based on Data Mining

Research on E-Commerce Automatic Question Answering System Model Based on Data Mining

Xiaohui Pan[1] and Tao Zhang[2,3]

1 Computer Course Teaching Department, Shanghai University of Political Science and Law, Shanghai, China 201701
2 Department of Computer Engineering and Science, Shanghai University, Shanghai, China 200444
3 Shanghai Institute for Advanced Communication and Data Science, Shanghai, China 200444
Email: panxiaohui@shupl.edu.cn; taozhang@shu.edu.cn

Abstract. With the rapid development of the Internet and artificial intelligence, intelligent question answering systems have become current research hotspots because they can provide users with accurate answers and intelligent services. They are gradually entering the e-commerce field to replace some manual work. In this paper, we proposed several models including a user intent recognition model, a pattern of association rule mining model and a model of the entire e-commerce auto answering system. In addition, the application effects of these models were analysed. The models and analysis in this work are useful for constructing e-commerce automatic question answering system or making personalized recommendations and related services.

1. Introduction

E-commerce is the use of the Internet, electronic data exchange, electronic funds transfer and other information technology in the country, companies and individuals to carry out economic businesses. In these businesses, buyers and sellers do not meet each other and all activities are conducted online. This introduce some problems since buyers are not able to get the product information such as quality, function and usage through physical contact, while the merchants are struggling to cope with numerous buyers with the same or different problems and consultations, which consumes a lot of energy and time. Therefore, the use of data mining [1] technology to design a high-intelligence e-commerce online automatic question answering system makes it possible for a large number of customers in e-commerce operations to randomly ask questions without the need for online answers from e-commerce customer service people.

1.1. E-Commerce Automatic Question Answering System

In recent years, people want quick access to effective information, prompting question answering systems to become a hot topic in the field of natural language processing. Both industry and academia are actively researching and exploring related theories and techniques for constructing automatic question answering systems [2]. The e-commerce question answering system is proposed for the universality and commonness of consumer issues [3]. In simple terms, the e-commerce customer service first saves a large number of frequently asked questions and answers in the system problem database. When the user randomly asks questions online, the system looks for the same problem or highly relevant problems in the problem database and submits it to the user for reference. When the

user thinks that the question and answer are consistent, the online consultation is completed; when the user thinks it is inconsistent, the question will be sent to the electricity supplier customer service, and the customer service answers the question and adds it to the system question bank [4].

For example, when a user issues a question to a merchant through an e-commerce platform on the Internet (such as Taobao, Jingdong, etc.), the system first understands the problem, and then calls the problem-matching algorithm to calculate the matching degree of the relative new problem for all questions in the problem bank. The most relevant matching result sequence is sent to the client. The user selects and submits the answers to the questions from the answers provided. If there is an answer that matches the new question in the recommended question answer sequence, the recommended question visit frequency value is automatically increased by 1; if it does not exist, the system considers that there is no similar problem in the question bank. The system sends the new question to the e-commerce customer service. The customer service answers the question to the user who asked the question in the system, and saves the answer to the question bank, so that the specific question answer is also guaranteed.

The question raised by the user is called the question to be solved Q; after the system analyses the question Q, the question set that is related to it is called the related question (Q1... Qn). The process of dispelling the traditional e-commerce question answering model is mainly as follows: (1) knowledge point and semantic word decomposition of the problem to be solved Q; (2) process of finding related problems (Q1... Qn); (3) calculation of correlation degree of related problems (Q1...Qn) and sorting.

1.2. Key Technologies of Automatic E-Commerce Question Answering System

The ultimate goal of the automated question answering system is to understand the questions raised by customers in natural language and to make correct answers quickly. It involves many aspects of technology, including knowledge representation technology, knowledge storage technology and information extraction technology. In general, the automatic question answering system has the key technical components of problem analysis, information retrieval, and answer extraction [5].

The key technology in the problem-solving section can be used to analyse the customer's question semantically to determine the intention of the user's question [6]. It is very difficult for a computer to understand a language. For example, the word "apple" has a variety of meanings, including the "Apple" of fruits, and the famous "Apple" technology company. So sometimes we need to judge with contextual context. The problem-solving part needs to analyse the syntax, lexical and semantic information to analyse the semantic types of the problem and extract keywords.

The key technology of the information retrieval part can search for relevant data from an off-line or on-line database according to the keywords extracted from the problem-solving link, so as to obtain a candidate answer to the question. For example, you can search keywords in an online search engine and then obtain the first few results in the returned results as candidate results.

The key technique of the answer extraction part is to extract the answer from the preliminary candidate results, that is, to calculate the weight of each candidate result according to a certain strategy, such as calculating the user's problem and the similarity of each candidate answer, and then according to the weights, the candidate results. Sorting, and finally according to the selection strategy such as the type of problem to filter out the final answer.

1.3. Main Contributions

There are three main contributions in this paper: (1) we analysed the core technology of an automated question answering system and summarized recent research progress in this area; (2) It proposed several models including a user intent recognition model, a pattern of association rule mining model and a model of the entire e-commerce auto answering system; (3) the application effects of these models were analysed. The models and analysis in this work are useful for constructing e-commerce automatic question answering system or making personalized recommendations and related services.

2. E-Commerce Question Answering System Model

2.1. User Intention Recognition Model

In the automatic question and answer in the field of e-commerce, the system first needs to identify the user's intention through a single round or multiple rounds of dialogue, and then use different subsystems (such as customer service, assistant, chat) to deal with. Therefore, the architecture abstraction can be implemented by dividing the scene into sub-fields, and then different machine learning methods can be used to design the technology according to different layers and scenes. The e-commerce user intent recognition model can identify the real intention of the user's language, classify the intent and extract the intent attribute. This model is a process of inferring intentions based on contextual data and domain data models in the database, as shown in Fig. 1. In common e-commerce scenarios, there are three common types of user problems: question and answer, task and chat. Question-and-answer examples include "What if the username and password are forgotten?" This type of question can be answered by constructing a knowledge map and retrieving the model match. Examples of tasks include "I want to book a bouquet of 99 red roses and send it to my home at noon tomorrow". Such questions can be answered based on intention-based decision making and deep reinforcement based learning. Examples of chat types include “How do you feel good?” You can use a combination of search models and deep learning to get answers.

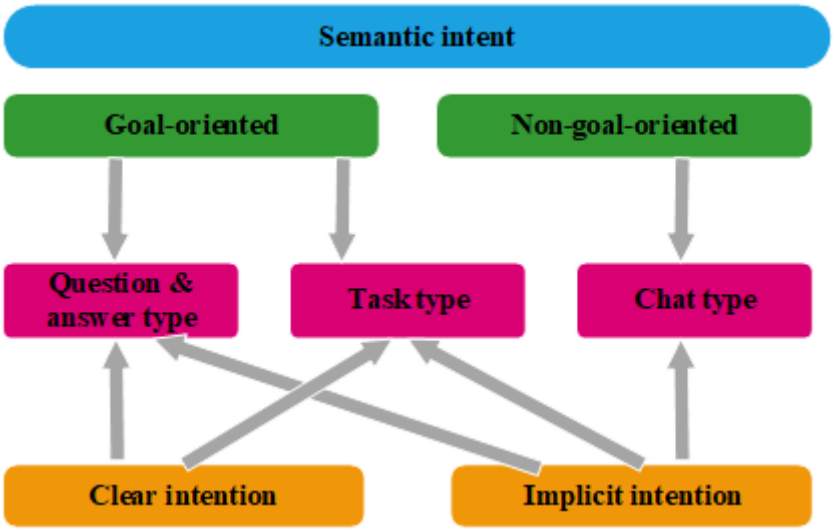

Figure 1. User intention recognition model.

2.2. Pattern Association Rules Mining Model

2.2.1 Mining association rules. The way to answer the question is to give the user some related questions. The user chooses to view the most relevant questions. There is a default premise: from the overall point of view, the problem that the user first chooses is generally related to the problem to be solved Q; the order of viewing reflects the order of the degree of relevance of the problem. It reflects the knowledge of dependence or association between an event and other events. If there is an association between two or more attributes, the attribute value of one of the attributes can be predicted based on other attribute values. The most famous discovery method for association rules is the Apriori algorithm proposed by R. Agrawal.

The discovery of association rules can be divided into two steps. The first step is to iteratively identify all frequent item sets, requiring that the support rate for frequent item sets is not lower than the user-set minimum value; the second step is to build confidence from frequent item sets to be no lower than the user-set minimum value. The rule of identifying or discovering all frequent item sets is the core of the association rule discovery algorithm and the most computational part. The mining of

association rules can be described formally as following: let $I=\{i_1,i_2,i_3,..,i_m\}$ be a set of m different attributes, also called items; a set of items is called an item A set, and a set of items containing k items is called a k-item set. Each transaction T in database D has a unique identifier TID and contains a set of items $T \in I$. An association rule can be represented as A=>B, where A, $B \in I$ and $A \cap B=NULL$. The support degree of an item set is s, if s% of transactions in D contain the item set; a set of items is said to be frequent, if the item set's support degree is not less than the frequency threshold determined by the user, that is, the minimum support degree; say c is the credibility of an association rule. If a transaction that contains A's transaction contains c%, it also contains B, that is, $c=support(A \cap B)/support(A)$.

2.2.2 Answer data preprocessing of resource mining - from related issues to related model conversions. For the initial question Q raised by the user, a series of related questions ($Q_1...Q_m$) will be generated after the system analysis. In turn, the Q and one of the related problems Qp are decomposed into a sequence of knowledge points and semantic words ($K_1...K_m$, $Y_1...Y_n$) and ($K_{p1}...K_{p(mi)}$, $Y_{p1}...Y_{p(ni)}$). The sequence ($K_1...K_m$, $Y_1...Y_n$) is converted to m*n pairs of patterns (K_i,Y_j) consisting of knowledge points and semantic words. Decomposing and transforming each pair of {Q, Q_p}, respectively can get the related pattern item set $T_i = \{(K_i, Y_j), (K_{pi}, Y_{pj})\}$. This translates the problem into a pattern that constitutes a targeted source of data mining.

2.2.3 A pattern tree consisting of a knowledge tree and a semantic tree. The structure of the knowledge tree is a tree-like network structure. As a node in the knowledge network, the connection between knowledge points is very complicated. It is generally believed that there are mainly the following types of links between knowledge points: forerunner/successor; inclusions; brothers; citations. The knowledge point relationship applied in the generalization of the pattern is a knowledge point tree concerning the containment relationship of knowledge points. Similarly, the point relationships of the semantic tree used are also based on the inclusion relationship between semantics, similar to the structure of the knowledge tree. In the construction of the pattern tree, the partial order relationship between one pattern and other patterns is determined according to the following principle: If for $M_1(K_1,Y_1)$ and $M_2(K_2,Y_2)$, $K_1 \in K_2$ and $Y_1 \in Y_2$, then $M_1 \in M_2$. If $K_1 \in K_2$ but does not satisfy $Y_1 \in Y_2$, there is only a hierarchical relationship instead of a inclusion relationship between M_1 and M_2. Their hierarchy level is judged by the hierarchy sum of K1 and Y1 and the hierarchy sum of K_2 and Y_2.

The e-commerce auto answering system knowledge tree specifically refers to the knowledge of various commodities (including brands, features, parameters, utilities, etc.), depending on the technical experts of the commodity or marketing experts to participate in the construction.

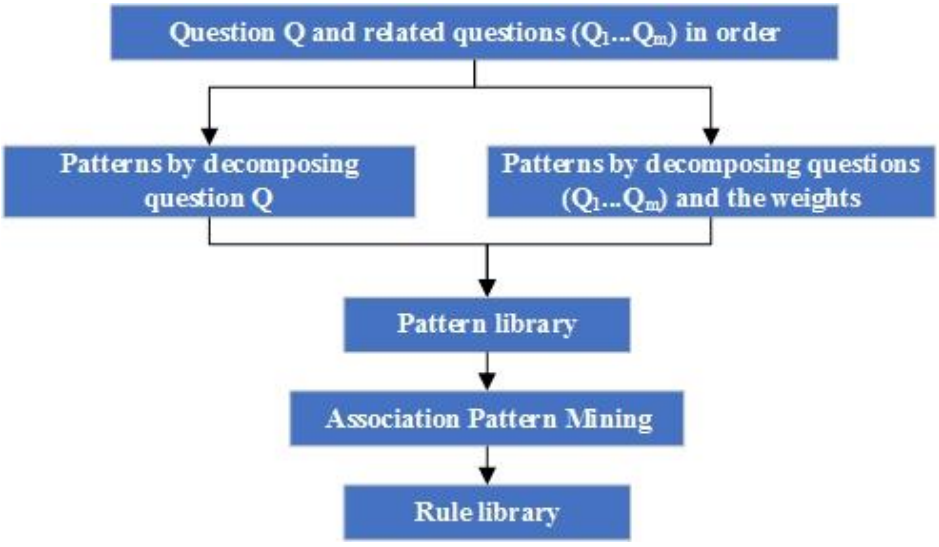

Figure 2. Mining multi-level association rules based on pattern tree generalized correlation patterns.

2.2.4 Mining multi-level association rules based on pattern tree generalized correlation patterns. The application pattern tree performs data generalization processing for all patterns to a certain level and obtains generalized related pattern data. Because the access order of the questions in each model is different and the obtained patterns has a corresponding degree of relevance after the generalization of the order of the questions, the categories of the relation degree are related to the category coefficients. This results in a series of patterns containing correlation coefficients. By applying the correlation rules mining on the generalized relational model, we can get some rules listed in a certain level of the model. More ideal rules can be obtained by applying the adjustment algorithm to merge or delete some rules, as shown in Fig. 2.

2.3. A Complete Model of E-Commerce Question Answering System

The automatic question answering model is briefly summarized as follows: (1). decompose knowledge points and semantic words of the initial problem Q; (2). generate relevant problems (Q_1... Q_m) and related patterns using rule bases, problem databases, etc.; (3). use rule bases, problem banks, etc. to calculate and rank related issues (Q_1...Q_m); (4). match the most relevant questions or store them in the problem bank without related problems; (5). e-commerce customer service maintenance library in non-real-time.

Using the rule results of multi-level association rule mining based on correlation patterns of decision tree (pattern tree) generalization, we propose a complete model of e-commerce question answering system, as shown in Fig. 3.

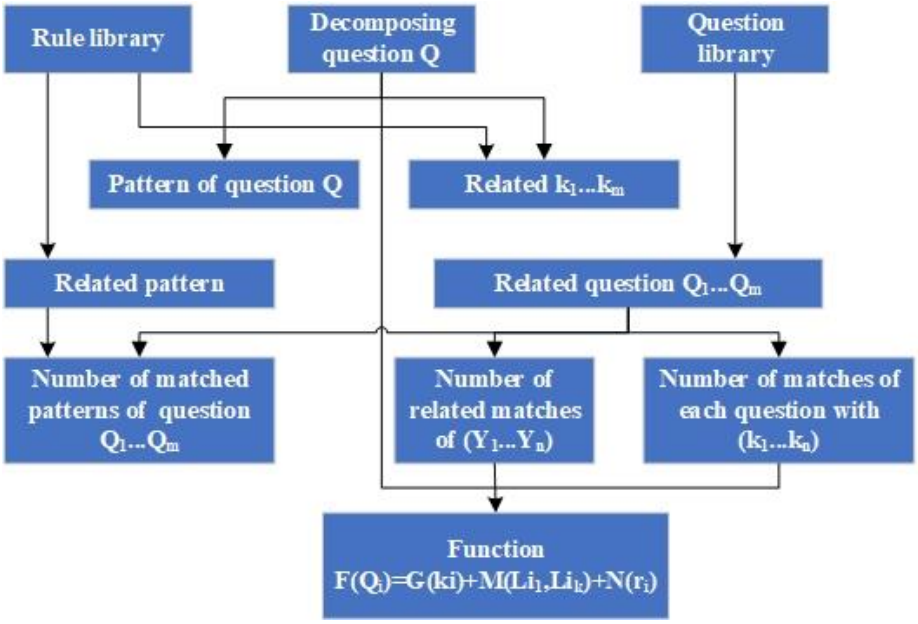

Figure 3. Automatic question answering system model.

The product knowledge points in the better rule pattern are extracted and added as knowledge points to the database of knowledge points. In word segmenting, these knowledge points are given higher priorities. In this way, the word segmentation is not only fast, but also highly accurate.

In calculating the degree of correlation based on the splitting the original problem, we analyse the existing problems and the patterns on the left and right sides of the existing rules in the rule base to match, and the calculation of the correlation of the original problems is obtained by the following formula:

$$\mathrm{F}(Q_i) = G(k_i) + M(L_1, L_k) + N(\mathrm{r_i}) = \mathrm{k_i} + L_1 \ldots\ldots L_k + 5 * \mathrm{r_i} \quad (1)$$

3. Analysis of Application Effects of E-Commerce Auto Answering Model

The adoption of e-commerce auto answering model can provide consumers with fast and efficient consulting services in real-time. Compared with current e-commerce customer services, online communication software often causes interruptions or delays in online response, which wastes valuable time of consumers. An e-commerce auto answering system is expected to save 60% of the time and more than 80% of customer service and customer service equipment. When a large number of consumers consult at the same time, it will result in insufficient customer service and increased manpower. At the same time, it also restricts customer service to be on-line at all times, which wastes human resources. Therefore, the practical application of the automatic question answering model will greatly improve the efficiency of e-commerce operations.

However, after all, the intelligence of the online automated Q&A model cannot be completely comparable to that of the customer service. When encountering some difficult-to-understand or more specific problems, it may fail to get a satisfactory answer. At this time, customer service is still required. Manually complete the response, so I believe that in the operation of e-commerce using automatic question and answer system model to deal with customer-based consulting, supplemented by customer service manual answer is the most ideal model.

4. Conclusion

E-commerce has gotten rid of the limitations of time and space, and saved a lot of manpower costs, realized the global sharing of various resources, and is rapidly developing with unparalleled advantages. The automated question answering system can save manpower and can therefore be used in e-commerce. This article analysed the core technology of an automated question answering system and summarized recent research progress in this area. It then proposed several models including a user intent recognition model, a pattern of association rule mining model and a model of the entire e-commerce auto answering system. Finally, the application effects of these models were analysed. The models and analysis in this work are useful for constructing e-commerce automatic question answering system or making personalized recommendations and related services. As of future work, we are going to explore how to use optimization algorithms [7] [8][9] to search valid solutions for user intention understanding and answer extraction. We'll also design more efficient models for e-commerce automatic question answering system.

5. Acknowledgements

The authors would like to thank anonymous reviewers for their fruitful feedback and comments that have helped them improve the quality of this work. This research is partially supported by Natural Science Foundation of Shanghai under grant No. 17ZR1409800.

6. References

[1] Wu X, Zhu X, Wu G Q and Ding W 2014 Data mining with big data *IEEE transactions on knowledge and data engineering* **26 (1)** 97-107

[2] Zhang M, Huang T, Cao Y and Hou L 2015 Target Detection and Knowledge Learning for Domain Restricted Question Answering *Proc. of the 4th CCF Conf. on Natural Language and Chinese Computing* (Verlag Berlin, Heidelberg: Springer) pp 325-336

[3] Yu J, Qiu M, Jiang J, Huang J, Song S, Chu W and Chen H 2018 Modelling Domain Relationships for Transfer Learning on Retrieval-based Question Answering Systems in E-commerce *In Proc. of the Eleventh ACM Int. Conf. on Web Search and Data Mining* (New York: ACM) pp 682-690

[4] Turban E, Outland J, King D, Lee J K, Liang T P and Turban D C 2018 Intelligent (Smart) E-Commerce *Electronic Commerce 2018: Springer Texts in Business and Economics* (Charm: Springer) pp249-283

[5] Jin X L, Zhou Z, Lee M K and Cheung C M 2013 Why users keep answering questions in online question answering communities: A theoretical and empirical investigation *Int. J. of Information Management* **33 (1)** 93-104

[6] Dwivedi S K and Singh V 2013 Research and reviews in question answering system *Procedia Technology* (Atlanta, USA: Elsevier) **10** 417- 424

[7] Harish G 2016 A hybrid PSO – GA algorithm for constrained optimization problems *Applied Mathematics and Computation* (Atlanta, USA: Elsevier) **274** 292 – 305
[8] Harish G 2014 Solving structural Engineering Design Optimization Problems using an artificial bee colony algorithm *J.l of Industrial and Management Optimization* **10(3)** 777 – 794
[9] Harish G 2014 A hybrid GA-GSA algorithm for optimizing the performance of an industrial system by utilizing uncertain data *In Handbook of Research on Artificial Intelligence Techniques and Algorithms* ed P Vasant (USA: IGI Global) Chapter **20** 625 – 659