2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

A Knowledge Acquisition Framework for Autonomous Decision Making in Service Robots

1st Hao Wu
School of Control Science and Engineering
Shandong University
Shandong Province, China
wh911@sdu.edu.cn

2nd Zhixian Zhao
School of Control Science and Engineering
Shandong University
Shandong Province, China
202234957@mail.sdu.edu.cn

3th Qing Ma*
School of Control Science and Engineering
Shandong University
Shandong Province, China
maqing@sdu.edu.cn

4rd Guohui Tian
School of Control Science and Engineering
Shandong University
Shandong Province, China
g.h.tian@sdu.edu.cn

***Abstract*—Service robots are expected to autonomously perform a wide range of service tasks to satisfy users' needs, but are limited in practice by their weak decision-making capabilities. This work introduces a Knowledge Acquisition Framework (KAFS) to help robots make autonomous decisions through this knowledge. This framework is divided into two parts: service knowledge acquisition and scene knowledge construction and uses a variety of intelligent methods to easily and accurately acquire a large amount of service and scene knowledge. We demonstrate the knowledge acquired by KAFS and validate the effectiveness of KAFS on robot service tasks.**

***Index Terms*—service robot, knowledge acquisition, decision making**

I. Introduction

As a major branch of robotics applications, service robots have made great progress in robot perception and motion with long-term research and development. There are also applications in real home scenarios, such as floor cleaning robots. Although these robots provide some convenience, they are still unable to perform generic service tasks autonomously to meet the needs of users due to the robots' limited decision-making capabilities. However, robots can guide motion planning based on their own knowledge base, just as humans. Therefore, acquiring a large amount and high quality of knowledge is essential to improve the intelligent decision making ability of service robots under complex tasks.

The knowledge required in the process of intelligent robot service includes service knowledge and scene knowledge. Service knowledge provides the robot with specific execution strategies for different service tasks. Scene knowledge is the spatial knowledge of scenes and objects learned by the robot in different service scenarios. By combining these two types of information, the robot can perform multiple service tasks in different service situations.

Therefore, we used advanced natural language processing algorithms, deep neural network models, and service robot research results to build KAFS (a **k**nowledge **a**cquisition **f**ramework for **s**ervice robots) for autonomously acquiring service knowledge and scene knowledge in the intelligent robot service process. The robot can guide its own service behavior based on the explicit knowledge acquired by KAFS to accomplish complex service tasks in a changing environment.

In this framework, the service knowledge acquisition starts with designing a VSM-based thematic web crawler for crawling service knowledge from the web, followed by dividing the clustered service knowledge into a task decision table and a task execution table so that the robot can use this service knowledge. Meanwhile, according to the actual service situation, Apriori algorithm is used to mine potential service rules and expand them into service task chains. The scene knowledge is constructed according to the scene knowledge framework, using ResNet-50 and Faster-RCNN models to identify areas and objects, and using Neural Motifs model to detect the location relationship between objects.

In our experiments, we first introduce the service knowledge base and scene knowledge map constructed from the knowledge acquired by KAFS, which can facilitate the robot's query. Then, we take the service task of ”make coffee” as an example and show how the robot can use the service knowledge and scene knowledge to complete the service task. The experimental results show the effectiveness and robustness of KAFS. Our contributions are as follows:

- Proposing KAFS (a **k**nowledge **a**cquisition **f**ramework for **s**ervice robots), which is an extensible, interpretable, knowledge acquisition framework aimed at improving the autonomous decision-making capabilities of service robots.
- A specific method of acquiring service knowledge and scene knowledge is given, by using various algorithms and models.
- A service knowledge base and scene knowledge graph available for use are constructed according to the methods provided by KAFS, and the effectiveness of the frame-

work is demonstrated by experimentally demonstrating the robot's use of knowledge to autonomously complete service tasks.

II. Related Works

A. Service Task Understanding and Mining

Task understanding transforms abstracted natural language instructions into knowledge that the robot itself can understand and execute. Misra et al. [1] formed a dataset of the collected tasks and the corresponding sequences of action execution. Along with the establishment of Web-based knowledge sharing sites, more and more work has begun to attempt to use public knowledge resources for task decomposition [2]. Perera et al. [3] utilized web resources to learn knowledge related to task execution. Tenorth et al. [4] dealt with knowledge gaps based on the KNOWROB knowledge system. Some recent studies have utilized end-to-end learning methods to train robots to move directly from verbal command inputs to executive action outputs [5], [6], but limited by the size of the neural network and the unavailability of high-quality data, this approach performs poorly in terms of semantic generalizability.

Rule patterns at the multitasking level require new rules to be inferred by means of pattern mining. Najafabadi et al. [7] captured multiple purchase records for each transaction based on association rules. Zhang et al. [8] proposed a fuzzy association rule mining algorithm to model the links between different types of crime rates. These applications show the realistic possibility of employing association mining algorithms to deal with hidden service patterns during robotic tasks.

B. Expression of Environmental Information

The representation of environmental information has been a key direction in the research field of intelligent robot services. Silberman et al. [9] modeled support relationships for typical object regions in the environment, reflecting the physical interactions of the regions. Shi Y et al. [10] proposed modeling environmental context through a data-driven approach. With the construction of large-scale datasets, it has become a common practice to express structure into a natural language expression through neural network modeling [11]- [14], which is gradually replacing the traditional methods. The current use of neural networks provides the basis for robots to construct knowledge from scenes.

III. Service Knowledge Acquisition

Our work uses natural language to characterize the service knowledge, which is easy for humans to understand and for robots to acquire autonomously. Since the Internet contains a large amount of information related to home services, it is possible to extract this information into the deep knowledge needed for service tasks. First design high performance thematic web crawlers to extract relevant knowledge from the web. After that, we use the text clustering method to integrate the massive task knowledge and construct the service knowledge base according to certain rules, so that the robot can obtain the strategy to complete the task by querying the knowledge base. To make the robot service more intelligent and humanized, the Apriori algorithm is used to mine potential service rules to expand a single service task into a multi-service task chain.

A. Designing a Thematic Web Crawler

Setting the initial link queue and initial topic description of a web page as the basis for crawling, the accuracy of both plays a vital role in obtaining high quality topic pages. The crawler simulates a browser client sending a request to the corresponding web server and obtaining the valid content of the web page. The similarity between the crawled page and the initial topic is calculated using the vector space model (VSM) to determine whether this page is relevant to the topic.

VSM treats a document as a vector of n-dimensional features, n depending on the number of document feature words. The text d_i can be expressed as $\vec{d_i} = (w_{1i}, w_{2i}, \ldots, w_{ni})$, w_{ki} denotes the weight of feature word k. The weights are calculated using the TF-IDF algorithm, which is often used in technical fields such as information retrieval and text mining. Based on the obtained word weights, the similarity between the crawled pages d_i and the initial topic d_j is measured using the cosine distance metric (1). If the similarity is greater than the set threshold, the text content of the page is stored, otherwise the page is discarded.

$$sim\left(d_i, d_j\right) = \frac{\sum_k w_{ki} \times w_{kj}}{\sqrt{\sum_k (w_{ki})^2}\sqrt{\sum_k (w_{kj})^2}} \tag{1}$$

B. Extraction and Storage of Service Knowledge

Since service knowledge is jointly extracted from multiple knowledge sources in the network to enrich the knowledge base, there is a large amount of knowledge redundancy. Text clustering algorithms are needed to merge similar service knowledge and distinguish tasks with different descriptions to save knowledge storage space and improve the efficiency of robots querying service policies. Service knowledge clustering using BIRCH algorithm, a hierarchical text-based clustering method. Using this method, the text closest to the center of mass in each clustering feature is extracted and stored as representative service task knowledge.

The service knowledge extracted from the network is expressed in the form of natural language, and this type of knowledge cannot be directly used by robots, so it is necessary to adopt a form of knowledge representation suitable for use in robot services. Considering that even for the same service task, the strategy for performing the task is never the same because the robot faces different service items in the environment, the method of deciding the service strategy based on the task semantics and object conditions is proposed. Task semantic forms such as: ”make coffee”, ”make tea”, and ”turn on the TV”. Depending on the object's conditions, there exist many different ways to perform each task. Use the extracted service knowledge to autonomously construct a task decision table (TABLE. I), and decide the execution method by querying this table.

TABLE I
TASK DECISION TABLE

Task	Objects	Topic
make coffee	cup coffee water	A0001
make coffee	kettle coffee cup	A0002

Then use the service knowledge to construct a task execution table (TABLE. II) to give the specific execution strategy for each execution method, so that the robot can follow the steps to complete the service task. The literature states that service knowledge consists of four types of key information: instruction action object (V_{action}), action operation object (V_{obj}), service object (V_{per}), and location object (V_{loc}), so the service knowledge is disassembled into these four types of information.

TABLE II
TASK EXECUTION TABLE

Topic	Step	$\mathbf{V_{action}}$	$\mathbf{V_{obj}}$	$\mathbf{V_{per}}$	$\mathbf{V_{loc}}$
A0001	1	take	cup	null	null
	2	put	coffee	null	cup
	3	put	water	null	cup

C. Expansion of Service Knowledge

In fact, in advanced task scenarios for personalized services, the various service operations are interconnected, so that a single service task can be expanded into a multi-service task chain. For example, when the robot performs the service of turning on the TV, it may also need to perform the service task of fetching fruits, which is an implicit service association knowledge. Knowledge expansion is achieved at the service rule level by extracting the implicit service knowledge generated by the robot in the service. The Apriori algorithm [15] is a frequent itemset algorithm for mining Boolean association rules, which is widely used in various research fields. Based on the Apriori algorithm, a service rule expansion algorithm for robots can be designed (TABLE. III).

TABLE III
SERVICE RULE EXPANSION ALGORITHM

Algorithm: Service rule expansion algorithm
Input: robot service items I, transaction data set TD, cycle count k. Output: new service rules. 1: Generate the candidate 1 item set $C_1, k = 1$. 2: Scan the transaction dataset TD, and count the number of occurrences of each item in C_1, filter out the least supported items, and generate the candidate frequent 1 item set L_1. 3: Associate L_1 with itself to generate the candidate 2-item set C_2, $k = 2$. 4: When $k >= 2$, the number of occurrences of each item in C_k is scanned in the data set TD, and the set of items meeting the minimum support is filtered out as the candidate frequent k-item set L_k. 5: Generate the candidate set C_{k+1} by associating L_k with itself. 6: Repeat step 4 until the final set of frequent items is found. 7: The set of items satisfying the support and confidence requirements is used as a new rule for intelligent robot services R_{new}. 8: Return R_{new}.

IV. Scene Knowledge Construction

The advanced cognitive ability of service robots relies on effective modeling of environmental structures, but there are numerous scenes and objects in the home service environment, and there are complex and variable spatial connections between service objects and scenes. Therefore, modeling only the overall space is not only inefficient but also challenging to extract the semantic information required for advanced service instructions. Through scene knowledge construction, the attribution relationship between objects and environment and the relative position relationship between objects and objects are abstracted using semantics, and the objects are recognized at the instance level. A unified knowledge construction is formed.

A. Building a Scene Knowledge Framework

The scene knowledge in the service environment contains the spatial attribution relationship between objects and areas and the location relationship between objects. According to the characteristics of the environment, a progressive spatial attribution relationship of ”functional area - sign object - service object” can be drawn, as shown in Fig. 1. Functional areas describe high-level scenes, such as bedrooms, living rooms, etc. Service objects refer to objects that the robot needs to perform operations on, such as cups, phones, etc. The sign objects often have more obvious visual characteristics and have a strong location invariance, the location of such objects and service objects are closely linked, with space ”sign” roles, such as tables, coffee tables, etc. The spatial hierarchy is reflected by such a chained semantic description that conceptualizes the robot’s workspace.

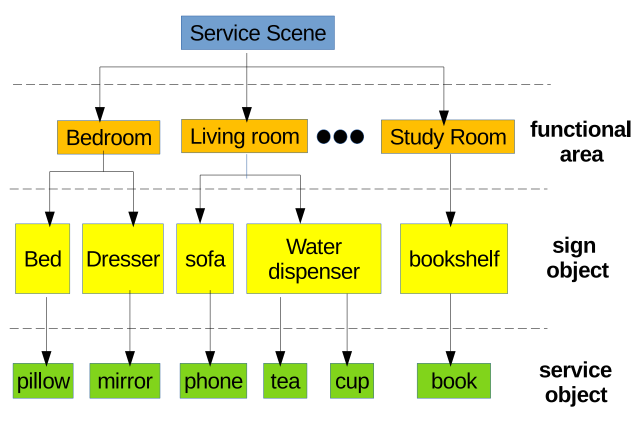


Fig. 1. Scene knowledge framework

B. Identification of Areas and Objects

1) Functional area Identification: The different functional areas directly affect the robot’s judgment of the services required by humans, and by acquiring the semantics of different areas, the service robot can make more accurate decisions. Using a classical deep residual network (ResNet-50) [16], the information from the environment pictures collected by the robot is fed into the neural network for classification and recognition to determine the functional properties of each area.

2) Sign Objects Identification: The distance between the objects and the ground can be used to determine whether the object is a sign object or not because it is placed on the ground due to its properties. Therefore, in this paper, RGB-D images are used to obtain the 3D coordinates of the objects and the target detection frame, and the two are combined to obtain the sign object. Detect object position using the Faster-RCNN algorithm [17] and transform pixel coordinates to camera coordinate system using depth information.

3) Service Object Instantiation Identification: The complex background of objects in the service scenario can confuse object instance detection, so a two-stage instance detection algorithm is used to obtain higher performance. After the robot acquires the pictures containing objects in the scene, the target detection frame is obtained by the Faster-RCNN algorithm to extract the objects from the background to reduce the noise brought by environmental information in the object instance recognition. Then the extracted objects are re-fed to the neural network to extract high-dimensional features, and the service objects instance cognition is completed by the feature similarity judgment of the item and the instance.

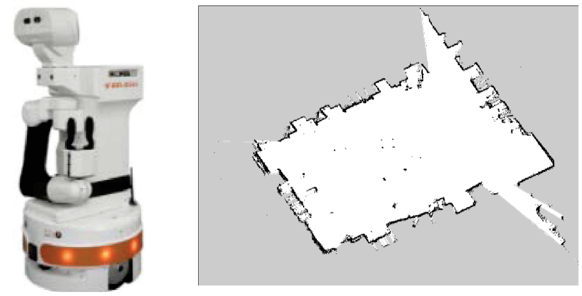
C. Detecting the Positional Relationship between Objects

The Neural Motifs model [18] is used to detect the positional relationships between objects. The model uses global prior knowledge in the scenario to guide the target relationship prediction task to improve accuracy. The model first generates item features and location vectors by Faster-RCNN. After that, the context framework is used to incorporate contextual information into target detection by coding and decoding two LSTMs to re-predict the target labels. Finally, it is combined with visual prediction to achieve location relationship prediction between pairs of items.

V. Experiments

In our experiments, by randomly giving some service commands to the service robot, we test whether the robot can autonomously reason, decide and execute the service tasks using the service knowledge and scene knowledge extracted from KAFS. Here is an example of the ”make coffee” command, showing the entire experiment process of the robot using the KAFS to complete the service requirements autonomously.

The selected real robot model is TIAGO, whose modules mainly include: TIM561 LIDAR sensor, movable base, manipulator, and other hardware. The robot platform was used to model the home environment, represented by a 2D raster map, which was used to assist the robot in performing its tasks, and the experimental platform is shown in Fig. 2:



(a) TIAGO (b) 2D raster map of the home environment

Fig. 2. Experimental platform

A. Service Knowledge Base

The service knowledge base consists of a task decision table (TABLE. I) and a task execution table (TABLE. II). A large expansion of the service rules is implemented in accordance with the thematic web crawler and extraction and storage of service knowledge. To ensure that the robot can efficiently query the service knowledge, these data are stored using a MySQL database. A total of 2521 task execution methods corresponding to 247 tasks are stored in the task decision table, with an average of 10 different execution methods for each task. The task execution table, on the other hand, provides specific execution steps for each execution method in the decision table, with an average of three steps required for each method.

B. Scene Knowledge Graph

The scene knowledge graph uses the graph form to integrate the scene knowledge obtained by scene knowledge construction, as shown in Fig. 3. The nodes of the knowledge graph are divided into three categories, which correspond to functional area, sign object, and service object in the scene knowledge framework. And respectively by the corresponding neural network algorithm in the identification of areas and objects to complete the recognition. The edges of the knowledge graph have different labels depending on the nodes. One category is the relationship labels (BeTo, Rel) that indicate spatial belonging, where BeTo is the belonging of the sign object and functional area, and Rel is the belonging of the sign object and the service object. The other category is the location relationship labels between objects, such as ”on”, ”in”, ”near”, etc., which can be accurately determined by the model in detecting the positional relationship between objects.

C. Complete Service Process

After receiving the ”make coffee” command, the robot extends its service knowledge through the expansion of service knowledge. Specifically, (make coffee, serve fruit, serve trash-can) constitutes a frequent triple set that is a strong association rule satisfying the confidence threshold and support threshold. That is, in the coffee-making task, the user’s implied service requirements include ”serve fruit” and ”serve trash-can”, so the latter is also added to the service task, and this kind of autonomous reasoning is very humanized, as shown in Fig. 4(a). Query the most appropriate execution steps for each

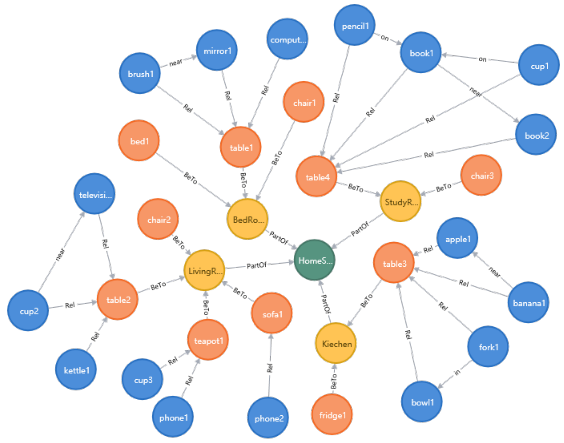


Fig. 3. Part of scene knowledge graph

task from the task decision table, as shown in Fig. 4(b). There are 10 rows for ”make coffee” and 1 row for each of the other two tasks. The most suitable step topics for ”make coffee”, ”serve fruit” and ”serve trash-can” are found to be ”A0034”, ”A0026” and ”A0014” respectively. The detailed steps of the corresponding execution methods are then obtained from the task decision table, as shown in Fig. 4(c). The final composition of the complete service task chain is shown in Fig. 5.

Input service message:	Get services:
make coffee	*{make coffee,serve fruit,serve trash-can}*
Loading service log …	Scanning decision table …
Successfully!	2521 rows, 247 tasks
Inquire possible service …	make coffee --10 rows
Having 2 results.	serve fruit --1 rows
--------------------------	serve trash-can --1 rows
serve fruit	Best match topic --make coffee --A0034
serve trash-can	Best match topic --serve fruit --A0026
---------------------------	Best match topic --serve trash-can --A0014
(a) Service knowledge expansion results	(b) Task decision table search results

ID	Topic	Step	Action	obj	per	loc
63	A0034	1	take	cup	null	null
64	A0034	2	put	coffee	null	cup
65	A0034	3	take	kettle	null	null
66	A0034	4	put	hot water	null	cup
		4 rows in set (0.06 sec)				
51	A0026	1	find	fruit	null	null
52	A0026	2	take	fruit	null	null
		2 rows in set (0.00sec)				
26	A0014	1	find	trash-can	null	null
27	A0014	2	take	trash-can	null	null
		2 rows in set (0.00 sec)				

(c) Task execution table search results

Fig. 4. Service knowledge query results about ”make coffee”

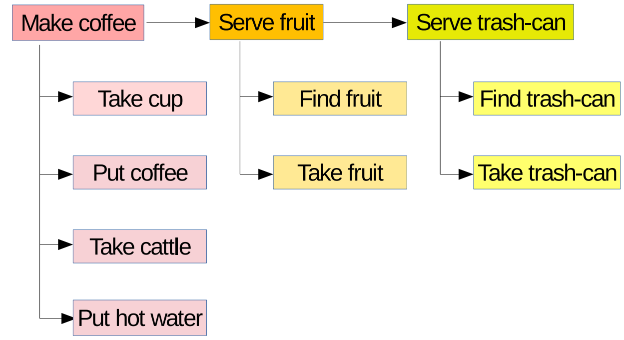
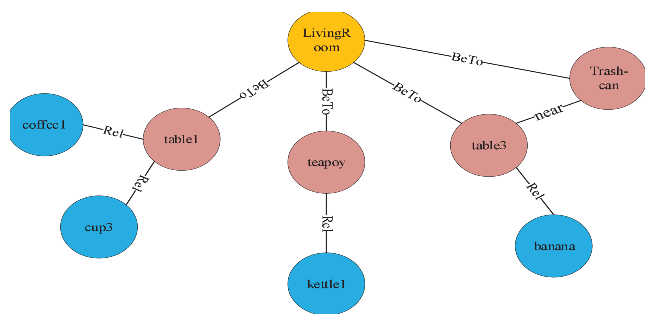
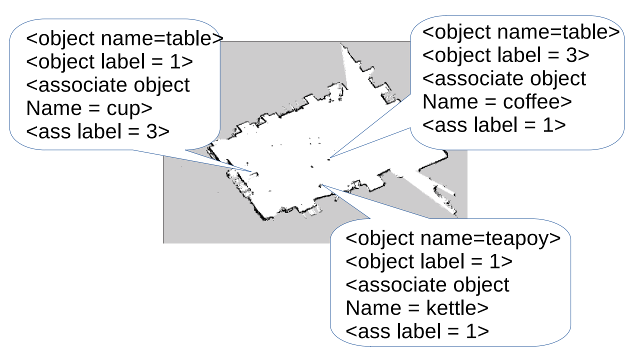


Fig. 5. Service task chain generated by the ”make coffee” command

In this task, the cups and coffee are on Table 1 in the living room, the kettle is on the coffee table in the living room, the fruit is on Table 3 in the living room, and the trash can is next to Table 3, as can be obtained from Fig. 6(a). Through spatial location attribution, the robot can quickly locate service objects based on the more visually obvious ones. In contrast, the relative location relationships between objects enable the robot to accurately understand the user's possible service needs. The acquired environment information and the 2D map of the environment are combined to generate a map of the home scene with semantics, as shown in Fig. 6(b). Fig. 7 shows the execution route of the robot in the home environment.



(a) Scene knowledge graph related to the task "make coffee"



(b) Semantic map generated from scene knowledge graph

Fig. 6. Using scene knowledge to generate semantic maps

We chose three different realistic home scenes, each with 10 experiments similar to the ”make coffee” instruction task. The

KAFS framework first constructs three scenario knowledge graphs, and then combines the service knowledge base and the experimentally given instructions to reason that the final service task chain has a success rate of 100%, with an average of five subtasks contained in each command task. According to the manual evaluation, the service task chain reasoned by the KAFS framework has some unreasonable task planning in the current scenario, and the average reasoning accuracy of the overall experiment is 84%. The final average task success rate for the 10 commands across the three scenarios was 76% due to the many interfering factors in the actual robots performing the tasks.

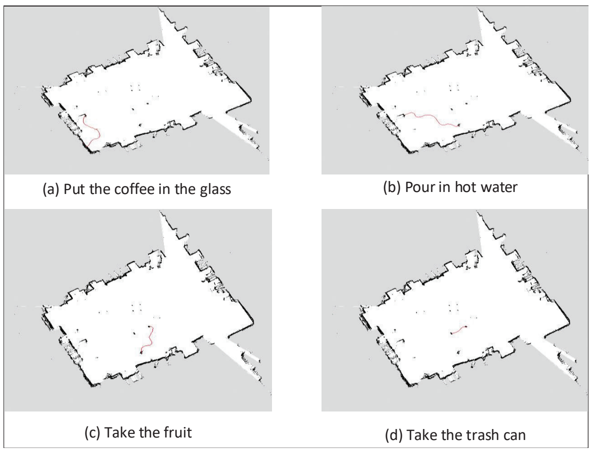


Fig. 7. The execution route of the robot

The experimental results show that the service knowledge in KAFS enables the robot to make task decisions autonomously based on service instructions and actual scenarios. The scene knowledge in the KAFS allows the robot to maintain semantic recognition of spatial relationships between things in complex and changing unstructured scenes. The experimental success rate of the robot autonomously completing service tasks under different service commands and scenarios is very high. This demonstrates the effectiveness and the superior robustness of the framework.

VI. Conclusion

We presented KAFS, a knowledge acquisition framework for autonomous decision making in service robots. With this framework, it is possible to obtain the services knowledge from the network and the scene knowledge from real environments. Our experiments demonstrate that the framework can autonomously acquire knowledge and construct it into a form that robots can use to make decisions in service tasks. The framework is extensible, which means that more knowledge can be added to cope with other problems in robotic service tasks. Therefore in future work, we will try to acquire a wider range of knowledge and experiment in unfamiliar environments to improve the generalization of the framework.

Acknowledgment

This work was supported by the following projects: National Natural Science Foundation of China 61973192, U1813215, 61973187, 91748115 and 62273203.

References

[1] D. K. Misra, J. Sung, K. Lee, and A. Saxena, "Tell me dave: Context-sensitive grounding of natural language to manipulation instructions," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 281–300, 2016.

[2] J. Xie and X. Chen, "Understanding instructions on large scale for human-robot interaction," in *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 3, pp. 175–182, IEEE, 2014.

[3] V. Perera, R. Soetens, T. Kollar, M. Samadi, Y. Sun, D. Nardi, R. Van de Molengraft, and M. Veloso, "Learning task knowledge from dialog and web access," *Robotics*, vol. 4, no. 2, pp. 223–252, 2015.

[4] M. Tenorth and M. Beetz, "Knowrob: A knowledge processing infrastructure for cognition-enabled robots," *The International Journal of Robotics Research*, vol. 32, no. 5, pp. 566–590, 2013.

[5] M. Shridhar, L. Manuelli, and D. Fox, "Cliport: What and where pathways for robotic manipulation," in *Conference on Robot Learning*, pp. 894–906, PMLR, 2022.

[6] S. Huang, Z. Jiang, H. Dong, Y. Qiao, P. Gao, and H. Li, "Instruct2act: Mapping multi-modality instructions to robotic actions with large language model," *arXiv preprint arXiv:2305.11176*, 2023.

[7] M. K. Najafabadi, M. N. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Computers in Human Behavior*, vol. 67, pp. 113–128, 2017.

[8] Z. Zhang, J. Huang, J. Hao, J. Gong, and H. Chen, "Extracting relations of crime rates through fuzzy association rules mining," *Applied Intelligence*, vol. 50, pp. 448–467, 2020.

[9] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V 12*, pp. 746–760, Springer, 2012.

[10] Y. Shi, P. Long, K. Xu, H. Huang, and Y. Xiong, "Data-driven contextual modeling for 3d scene understanding," *Computers & Graphics*, vol. 55, pp. 55–67, 2016.

[11] C. Lu, R. Krishna, M. Bernstein, and L. Fei-Fei, "Visual relationship detection with language priors," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 852–869, Springer, 2016.

[12] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, *et al.*, "Connecting language and vision using crowdsourced dense image annotations," *Visual genome*, 2016.

[13] Y. Liang, Y. Bai, W. Zhang, X. Qian, L. Zhu, and T. Mei, "Vrr-vg: Refocusing visually-relevant relationships," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10403–10412, 2019.

[14] Y. Liu, T. Wang, X. Zhang, and J. Sun, "Petr: Position embedding transformation for multi-view 3d object detection," in *European Conference on Computer Vision*, pp. 531–548, Springer, 2022.

[15] M. Al-Maolegi and B. Arkok, "An improved apriori algorithm for association rules," *arXiv preprint arXiv:1403.3948*, 2014.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[17] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[18] R. Zellers, M. Yatskar, S. Thomson, and Y. Choi, "Neural motifs: Scene graph parsing with global context," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5831–5840, 2018.