Proceedings of 2014 IEEE
International Conference on Mechatronics and Automation
August 3 - 6, Tianjin, China

Use Context to Understand User's Implicit Intentions in Activities of Daily Living

Rui Liu, Xiaoli Zhang *IEEE Member* and Songpo Li
Department of Mechanical Engineering, Colorado School of Mines, Golden, CO 80401, USA
{rliu, xlzhang & soli}@mines.edu

***Abstract* – For an assistive robot, accurately understanding the user's intention with minimal user involvement is critical in human-robot interaction. Especially for the elderly in their Activities of Daily Living (ADL), implicit intention recognition using subtle cues could provide them great convenience as well as improving quality of life. Usually user intentions in ADL are closely connected with specific context. For example, the user's intention for a cup on a hot day is likely to be drinking. Therefore, the relation between intention and context should be emphasized. The goal of our research is to effectively infer implicit human intention using the object that the user attends to and its subtle context. The context helps to filter the coarse intentions. Towards this goal, a new context-specific implicit intention recognition (IR) model has been established based on a Bayesian Network (BN) algorithm. Using set of questionnaires, the strongly intention-related context features have been selected and their conditional values for different intentions have been calculated. Then another set of questionnaires were used to verify the model's effectiveness. The results showed the BN based context-specific implicit IR model could help the robot to proactively get a good understanding of a user's implicit intentions with merely subtle context cues.**

Index Terms – assistive robot; ADL; context-specific implicit intention recognition; BN; Human-Robot Interaction;

I. INTRODUCTION

A. Assistive Robot

Currently the world population is in the aging trend. It's estimated that about 10% of current population are over 60 years old; by 2050, the proportion of the old population would exceed 20% [1]. This situation prompts the intelligent technologies such as assistive robots to play a crucial role in assisting the aging people with some daily activities. A robot walker provides mobility assistance and navigation service [2]. The Kinect-based guidance robot could help the elderly with their neurorehabilitations [3]. Auto-reminder alerts the elderly to process the daily schedule [4].

B. Intention Recognition (IR)

Robot assistance in activities of daily living (ADL), especially for the elderly, should be executed in an intuitive and natural way [1][5][6][7]. This requires a robot can proactively understand the user's intention and automatically provide the desired service. Moreover the robot should be able to predict human intention and then choose the most reasonable actions to cooperate with humans [8]. So satisfied assistance is based on accurate and real-time intention recognition (IR) technologies. For example, a smart wheelchair could detect human's sight paths to adjust its direction and speed in a more cooperating and a less frustrating way [9][10][11]. Surveillance robots have memory of the past sub-plans to help them recognize intention [8][13]. Recognizing system in a fighter distinguishes the most dangerous hypotheses in a very short time to assist the pilot's manipulation [8][14]. IR generally relies on action recognition [15][16], voice recognition [16], biological motion recognition [17], plan recognition [11][18], and activity or behavior recognition [19]. These features could be extracted by using various image processing [20], gesture detecting [21], motion-calculating technology [20][22] and object affordance [23][24]. (The affordance is a relation of action/activity/intention and a specific object used to predict the next action/activity [25][26].)

C. Context-Aware IR Technologies

In recent years, context awareness has been introduced into IR models, which allows robots to use context information to infer human intention. The context information includes location, people, objects, time, weather, season, temperature, user's emotional states, focus, etc [18][19][27][28]. Based on analyzing the environment and situations, the auto reminder robot could calculate the proper reminding time to initiate a proper reminder manner [29]. By gathering information such as hospital worker location, activity, and artifacts being used, the robot could conduct the clinical case assessment [28]. Based on time and location, the monitory robot infers a moving user's intended destination from his/her spatial behaviors [30]. Combining a person's actions with the surrounding environment, whether the user desires to cook or not could be judged by the robot [31].

In most of existing context-aware IR models, context features would firstly be combined with a BN model to get a set of coarse intentions [32]; and then human actions would be observed [31]; next mapping procedure is performed: if the sequence of these actions is coincident with the stored standard patterns in the robot's knowledge database, these actions would be recognized; lastly combining with coarse intentions and recognized actions, actual human intention is inferred [11][32][33][34].

D. BN Based Context-Specific Implicit Intention Recognition

Although robot could assist human in a more intelligent way by inferring a person's intention from observing his/her actions or activities, this kind of IR relies on the user's explicit inputs. For the elderly who are suffering from the physical impairment and even fully lost motion ability, it's unrealistic for them to perform such explicit motion tasks in the ADL assistances. Therefore implicit intention recognition with minimal human activity requirements is in an urgent need.

To meet this requirement, a new context-specific implicit IR model is proposed, shown as Fig.1. In this model the context

information has been emphasized to filter the coarse intentions. The user would no longer need to take actions/behaviors. The user merely needs to give a concise and ambiguous command which could just be a vocabulary, a nod or even a gaze to indicate the intention-related object. The object affordances, which are the possible intentions related to the object, are known from the database. Then the intention-related context features would be extracted from the surrounding environment. Based on a BN algorithm and these specific context features, the most likely intention would be inferred. Taking the windows for example, when "window" is mentioned in an ambiguous command, the possible intentions would be closing or opening some windows; combining with the context information that the noise level is very high only in windows A and air in other widows is fresh, the robot would infer that the user's intention is to close the window A. This new context-specific implicit IR model would provide a relative simple and effective way to infer user intentions.

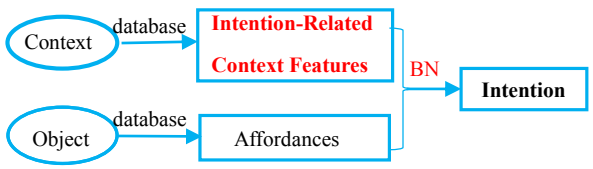


Fig.1 BN based context-specific implicit IR model in this paper

Research on integrating object affordances and context information into IR technologies is very rare. Based on the object that the user wants to manipulate, the likely human actions such as eating or drinking could be predicted [26]. However, the context features were not considered. Combining the context features such as the object shape and status with object affordances, the hand manipulation actions could be recognized [25]. However, its goal was limited to the simple action recognition level, not the intention level. In addition, it needs complex image processing work. In this paper, our objective is to effectively understand human implicit intentions using more subtle cues, such as the relation of user implicit intention related to an object and in what kind of context the user creates that specific intention. Our goal is to design a natural and intuitive Human-Robot Interaction Interface with the minimal effort from the user.

II. BN BASED CONTEXT-SPECIFIC IMPLICIT IR MODEL

In [30], the context is defined as any context information that could be used to characterize the situation of the entity. As Fig. 2 shows, there are various context features ranging from environment, user, to involved objects. In our model only the object intention-related context features would be taken into account. Selecting context features based on the intention-related object has three advantages: (1) The IR accuracy would be ensured. The selected context features are more targeted at each specific intention, so the correlation between context and intention is strong; (2) The IR computation efficiency would be improved. Only correlated context features are used for recognition; (3) The IR process could be simplified as the irrelevant context features will be eliminated and the

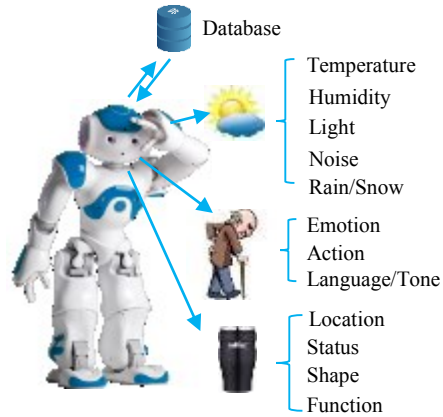


Fig. 2 Context features used in general IR models

dimensionality of the IR model is reduced.

In the real living environment, to build the relationship between the object-related intentions and correlated context is complex and challenging. That is because one object-based intention may be caused by more than one context features. On the other hand, a context feature cloud be correlated to more than one intentions. To solve this problem, a BN algorithm [31] is used to build the model. The BN algorithm could accurately describe complex probabilistic dependence relationships among many parameters and supply a natural method to solve problems under uncertainties.

We used a probability value to represent the correlation strength of a context feature to a specific intention. A higher probability value describes a stronger correlation, and vice versa. When multiple intentions share context features, the intention probabilities could be calculated to compare the relative probability degree.

The BN based context-specific implicit IR model M_{O_j} is shown in Fig.3. It builds a relationship between context $C_{O_j}(cf_1 \sim cf_s)$ (s is the number of total context features) and the intentions $I(i_1 \sim i_N$, N is the number of the total possible intentions). When an object O_j is attended by the user, the affordances $A_{O_j}(a_1 \sim a_N)$ would be evaluated by a specific context features c_i. Then the probability $f_i(i_i|c_i)$ that the intention i_i happens in context c_i(i=1~N) is calculated. The intention with the greatest probability value would be considered as the user's actual intention I_i.

The specific context c_i consists of s_i features $cf_{1i} \sim cf_{s_i i}$ which are strongly related to the intention i_i. $f(cf_{pi})$ means how likely would the context feature cf_{pi} appear in set c_i. As equations (1)-(5) show, I_i means a single IR process of the context-specific implicit IR model M_{O_j}. When information of O_j, a_i, c_i has been provided, the corresponding intention i_i would be identified.

$$M_{O_j} = \left[O_j, A_{O_j}, C_{O_j}, I\right] = \left[O_j, a_1 \sim a_N, c_1 \sim c_N, i_1 \sim i_N\right] = \sum_{i=1}^{N}\left[O_j, a_i, c_i, i_i\right] = \sum_{i=1}^{N} I_i \quad (1)$$

$$I_i = \left[O_j, a_i, c_i, i_i\right] \quad (2)$$

$$c_i = \{cf_{1i}, cf_{2i}, \ldots, cf_{s_i i}\} \quad (3)$$

$$f_i(i_i|c_i) = \sum_{p=1}^{s_i} f(i_i|cf_{pi}) f(cf_{pi}) \quad (4)$$

$$f(i_i|c_i, O_j, a_i) = \max(f_i(i_i|c_i)) \quad (5)$$

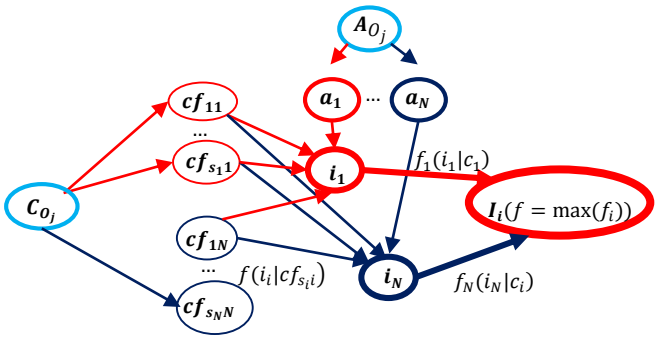


Fig. 3 BN based context-specific implicit IR model

III. EXPERIMENT

To build and validate our context-specific implicit IR model, conceptual experiments were preformed to recognize human implicit intentions using object affordances and context information. The simulated experimental environments are typical scenes in a daily living space with two task-related objects: cup and window, respectively. The two scenes are shown in Fig.4 and Fig.5.

A. Build the Context-Specific Implicit IR Model

The context-specific implicit model including objects cup and window with the correlated context information was first built. The frequently related affordance set of a cup is A_{Cup}; the frequently related affordance set of a window is A_{window}. The context features strongly related to intentions of cup are C_{cup}; the context features strongly related to intentions of window are C_{window}. Based on life experience, A_{Cup}, A_{window} could be defined as {drink, wash, transfer} and {open, close}, respectively; C_{cup}, C_{window} could be initially defined as { hot day, low drinking frequency, cup was used, near sink, in cabinet, medicine taking time } and {nice weather (sunny, air is fresh), poor weather (snowy, rainy), outside is quiet, outside is noisy, low room temperature, high room temp } respectively.

To establish the IR model M_{O_j}, the model's basic element I_i should be firstly created. According to these intentions, their related context feature sets $c_1 \sim c_N$ should be identified from the whole context feature set C_{O_j}. To calculate the relative possibility of each intention with a specific context, the conditional possibility of each context feature cf_{pi} would also be gained. The standard forms of cup and window IR models are as equations (6)~(9).

$$C_{Cup}=\{c_1,\ c_2,\ c_3\}=\{\textstyle\sum_{p=1}^{s_1} cf_{p1}, \sum_{p=1}^{s_2} cf_{p2}, \sum_{p=1}^{s_3} cf_{p3}\} \quad (6)$$

$$M_{Cup}=\textstyle\sum_{i=1}^{3} I_i \quad (7)$$

$$C_{window}=\{c_1,\ c_2\}=\{\textstyle\sum_{p=1}^{s_1'} cf_{p1}, \sum_{p=1}^{s_2'} cf_{p2}\} \quad (8)$$

$$M_{window}=\textstyle\sum_{i=1}^{2} I_i \quad (9)$$

B. Selection of strongly Intention-Related Context Features and Their Conditional Possibility Values

In our preliminary studies, we adopted the questionnaire method to identify the strongly related context features and gain their conditional possibility values. 60 copies of questionnaires have been filled by 60 participants with different ages ranging from 19~60 and different genders. The selection of the strongly intention-related context features and gaining of the probability values were based on the first set questionnaires.

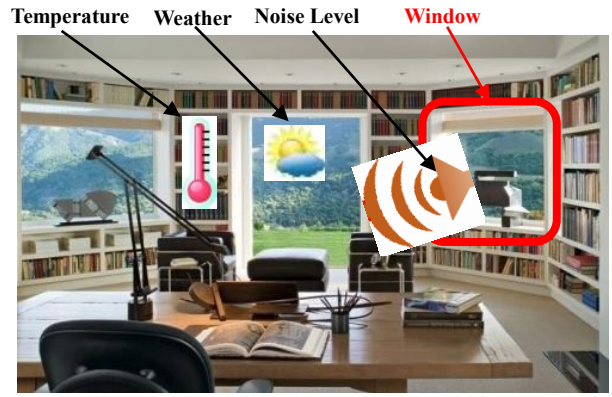


Fig. 4 Scene of daily living for the cup-related IR

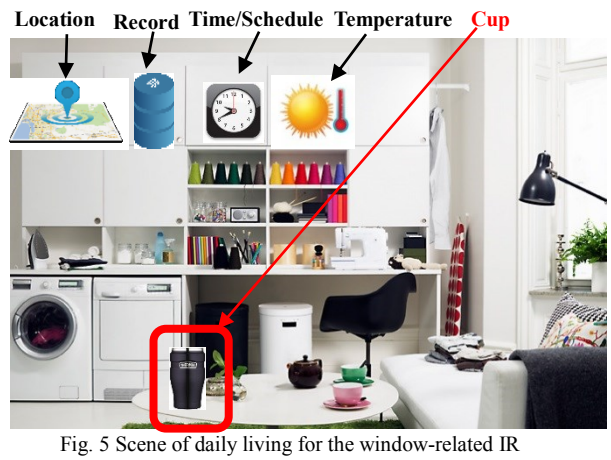


Fig. 5 Scene of daily living for the window-related IR

In the two scenes, for an intention, each context feature has four different probability levels 'highly likely', 'likely', 'not sure', 'impossible' and their corresponding possibility values are '0.9', '0.7', '0.5', '0.1'. A bigger value represents a stronger relationship between a context feature and an intention. The final conditional possibility value is obtained by training the BN model using the questionnaire results. More frequently a context feature is selected to indicate an intention with a bigger possibility value, more closely the context feature is connected with that intention. The training process could be expressed by equation (10). n_{levle} means the number of people choosing a specific probability level for the context feature cf_{pi} related to intention i_i; f_{levle} stands for the corresponding possibility value towards a specific level. When the model training procedure has been finished, the strongly intention-related context feature set would be gained.

$$f(i_i|cf_{pi}) = \frac{\sum_{k=1}^{4} n_{levle} * f_{levle}}{\sum_{k=1}^{4} n_{levle}} \quad (10)$$

C. Validation for Model Effectiveness

The second set of questionnaires (10 copies) was used to verify the effectiveness of the established models. Based on the context features and their intention-related conditional possibility values, the user intentions (cup $i_1 \sim i_3$, window

$i_1 \sim i_2$) have been inferred by our IR model. Only when the inferred intention is the same as that in the questionnaire, the intention recognition could be considered successful.

IV. RESULTS

Figures 6 and 7 show the conditional possibility values for each context feature related to cup-related intentions and window-related intentions, respectively. The context feature whose average possibility value is bigger than 0.6 was selected as the strongly intention-related context feature. They are marked by different color respectively.

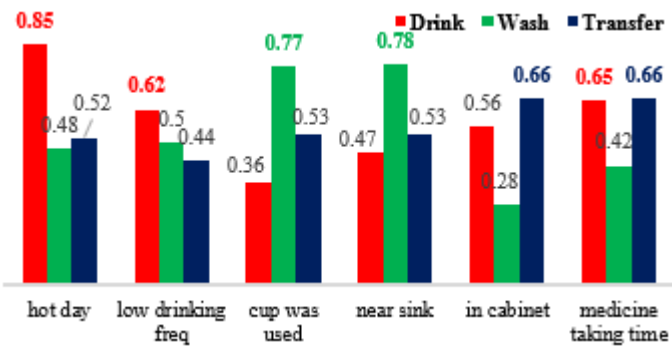


Fig. 6 Conditional probability values of context features for cup-related IR

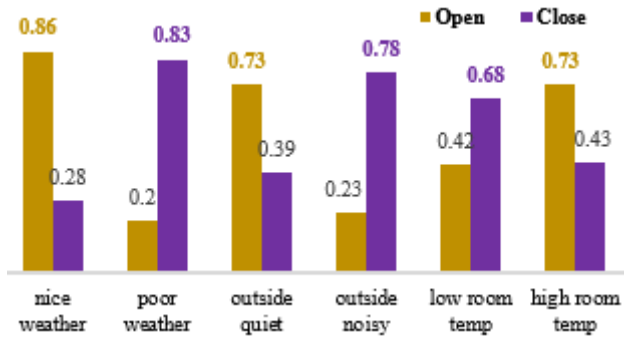


Fig. 7 Conditional probability values of features for window-related IR

We conclude that, (1) as to cup intentions, c_1={ hot day, low drinking frequency, medicine taking time}, c_2={ cup was used, near sink}, c_3={ in cabinet, medicine taking time}; (2) as to window intentions, c_1={ nice weather (sunny, air is fresh), outside is quiet, high room temp}, c_2={ poor weather(snowy, rainy), outside is noisy, low room temperature}. Then the completed BN based context-specific implicit IR models M_{cup} and M_{window} (shown in equations 6-9) have been successfully established. Figures 8 and 9 show the processes to infer 'drink', 'close' intentions in these two situations.

As Figs.10 and11 show, they are the comparison of the user's real intentions and the inferred intentions by the model

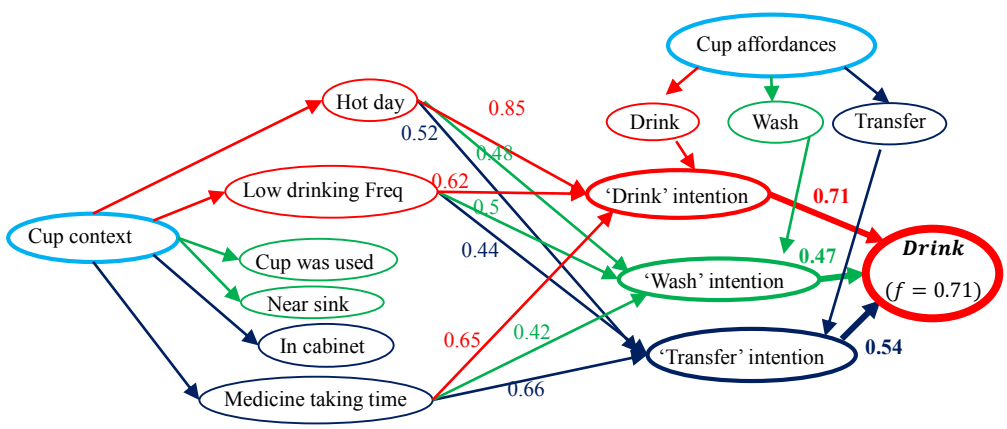


Fig.8 Recognizing the intention 'drink' in specific context by IR model

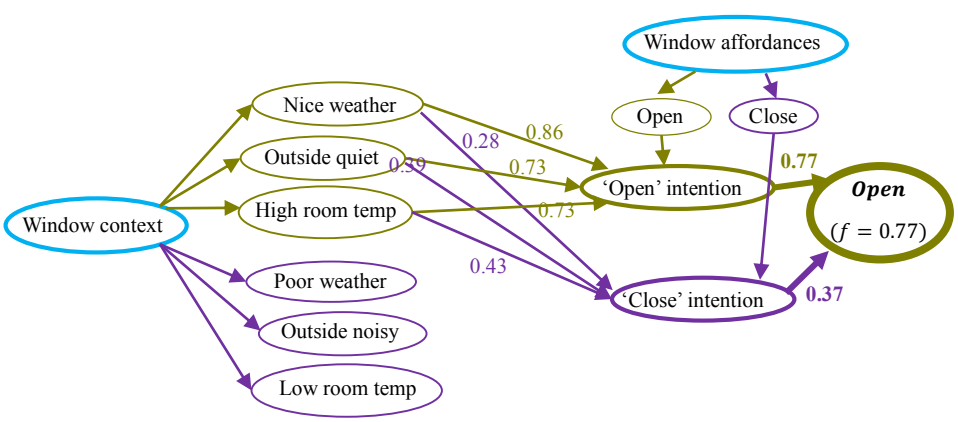
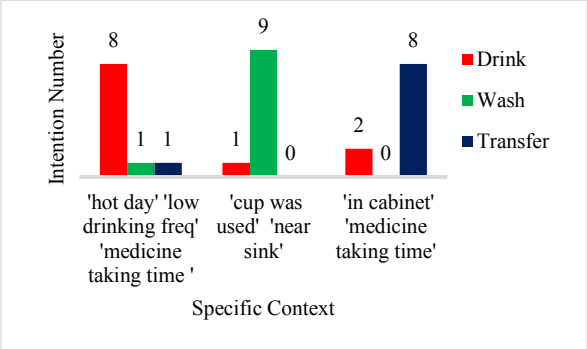
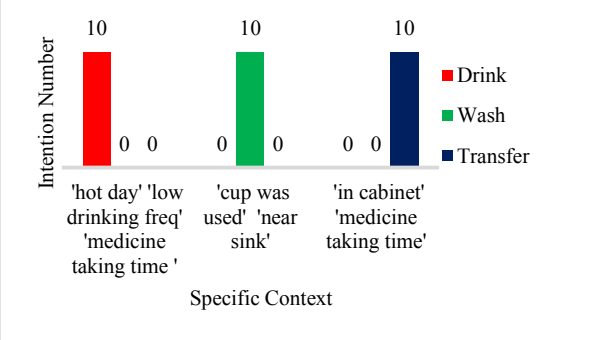


Fig.9 Recognizing the intention 'open' in specific context by IR model

in specific situations. Although the user's intention is not only decided by the context, in most situations the intention could be recognized accurately.

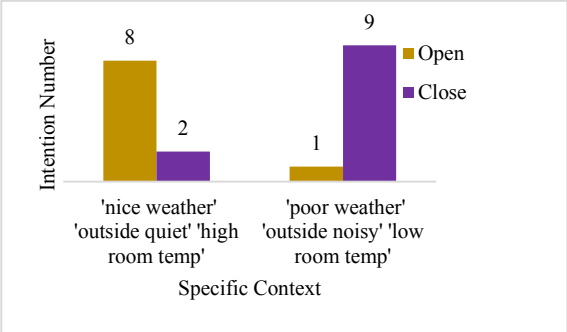


(a) User's actual intentions in specific situations by statistics

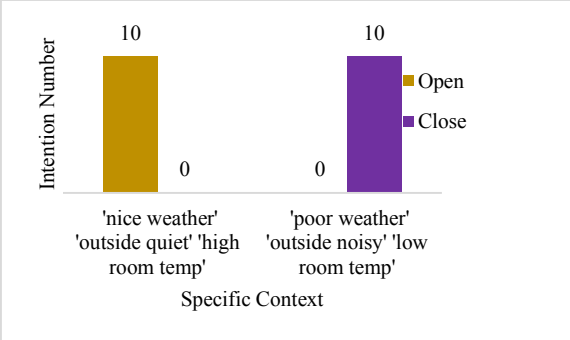


(b) Inferred intentions by the IR model in specific situations

Fig.10 Cup-related IR results in specific situations



(a) User's real intentions in specific situations by statistics



(b) Inferred intentions by the IR model in specific situations

Fig.11 Window-related IR results in specific situations

Because intention various with the context, in a given situation with one intention-related object the IR model solves only one real user real intention. Therefore, in Fig.10(b) and 11(b), there is only one intention inferred by the model for each specific context.

Fig.12 shows the IR accuracy of the model in specific situations. The average IR accuracy is more than 80% showing the effectiveness of this new model.

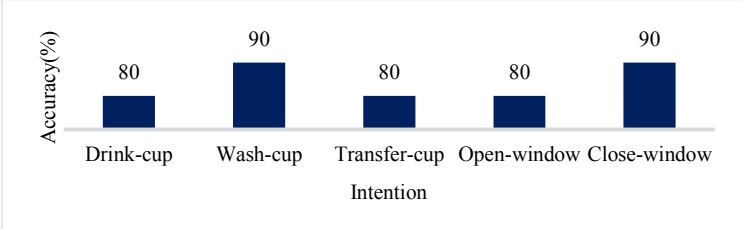


Fig.12 IR accuracy of the model in specific situations

V. Conclusion

In this paper, a new BN based context-specific implicit IR model is established. The strongly intention-related context features are used to calculate the probabilities of the object-related intentions. Features' conditional probability values related to a specific intention are gained from the questionnaires. The intention with the greatest probability value would be considered to be the user's real intention. The experiment results showed that the model could be reliable in assisting the robot to recognize the user's implicit intentions.

This new IR model only requires minimal user performances and context information. It would bring the user especially the elderly great benefits by offering a natural and intuitive interaction means for Human-Robot Interactions in the ADL. In addition, the assistance could be provided accurately and timely based on the good IR performances of the model.

When a completed model has been established towards all objects which would be frequently used in ADL, the completed IR system would be established to help users with every aspects of the daily living. Moreover, this model could also be combined with other action/behavior-recognition-based IR models to improve the IR performance.

References

[1] Pollack, Martha E, "Intelligent Technology for an Aging Population," *AI Magazine*, vol. 26, no. 2, pp. 9-24, September 2005.

[2] Morris, Aaron, et al, "A robotic walker that provides guidance," *Robotics and Automation, Proceedings. ICRA'03. IEEE International Conference*, vol. 1, pp. 25-30, 2003.

[3] Loconsole, Claudio, et al, "A new Kinect-based guidance mode for upper limb robot-aided neurorehabilitation," *Intelligent Robots and Systems*

(IROS), IEEE/RSJ International Conference on, pp. 1037-1042, 2012.

[4] Zhou, Shandan, et al, "A context-aware reminder system for elders based on fuzzy linguistic approach," *Expert Systems with Applications* vol. 39, no. 10, pp. 9411-9419, 2012.

[5] Nef, Tobias, and Robert Riener, "ARMin-design of a novel arm rehabilitation robot," In *Rehabilitation Robotics, ICORR 2005. IEEE 9th International Conference on*, pp. 57-60, 2005.

[6] Roy, Nicholas, et al, "Towards personal service robots for the elderly," *Workshop on Interactive Robots and Entertainment (WIRE 2000)*, vol. 25, 2000.

[7] Heinze, Clint, *Modelling intention recognition for intelligent agent systems*. No. DSTO-RR-0286. DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION SALISBURY (AUSTRALIA) SYSTEMS SCIENCES LAB, 2004.

[8] Krauthausen, Peter, *Learning Dynamic Systems for Intention Recognition in Human-Robot-Cooperation*. KIT Scientific Publishing, 2012.

[9] Tahboub, Karim A, "Natural and manmade shared-control systems: an overview," In *ICRA*, pp. 2655-2660, 2001.

[10] Simpson, Richard C, "Smart wheelchairs: A literature review," *Journal of rehabilitation research & development*, vol. 42, no. 4, 2005.

[11] Tahboub, Karim A, "Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition," *Journal of Intelligent and Robotic Systems*, vol. 45, no. 1, pp. 31-52, 2006.

[12] Repeat

[13] Bui, H. Hai, "A general model for online probabilistic plan recognition," In *IJCAI*, vol. 3, pp. 1309-1315, 2003.

[14] Pynadath, David V. "Probabilistic grammars for plan recognition." PhD diss., The University of Michigan, 1999.

[15] Mao, Wenji and Jonathan G, "A utility-based approach to intention recognition." In *AAMAS 2004 Workshop on Agent Tracking: Modeling Other Agents from Observations*, 2004.

[16] Jacob, Mithun G, and Juan P. "Context-based hand gesture recognition for the operating room," *Pattern Recognition Letters*, vol. 36, pp. 196-203, 2014.

[17] Blakemore, Sarah, and Jean D, "From the perception of action to the understanding of intention," *Nature Reviews Neuroscience*, vol. 2, no. 8, pp. 561-567, 2001.

[18] Abowd, Gregory D, et al, "Towards a better understanding of context and context-awareness," In *Handheld and ubiquitous computing*, pp. 304-307, Springer Berlin Heidelberg, 1999.

[19] Franklin, David, and Joshua F, "All gadget and no representation makes jack a dull environment," In *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*, pp. 155-160, 1998.

[20] Zhang D, Daniel G, et al, "Modeling individual and group actions in meetings: a two-layer hmm framework," In *Computer Vision and Pattern Recognition Workshop, CVPRW'04. IEEE Conference on*, pp. 117-117, 2004.

[21] Melinger, Alissa, and Willem J, "Gesture and the communicative intention of the speaker," *Gesture* 4, no. 2, pp. 119-141, 2004.

[22] Aarno, Daniel, and Danica K, "Motion intention recognition in robot assisted applications," *Robotics and Autonomous Systems* vol. 56, no. 8 pp. 692-705, 2008.

[23] Koppula, Hema S, and Ashutosh S, "Anticipating human activities using object affordances for reactive robotic response," *Robotics: Science and Systems, Berlin*, 2013.

[24] Sun, Y, Shaogang R, and Yun L, "Object–object interaction affordance learning," *Robotics and Autonomous Systems*, 2013.

[25] Kjellström, Hedvig, et al, "Visual object-action recognition: Inferring object affordances from human demonstration," *Computer Vision and Image Understanding*, vol. 115, no. 1, pp. 81-90, 2011.

[26] Kim, Sangwook, et al, "Intention Estimation and Recommendation System Based on Attention Sharing," In *Neural Information Processing*, pp. 395-402, Springer Berlin Heidelberg, 2013.

[27] Baldauf, Matthias, et al, "A survey on context-aware systems," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 2, no. 4, pp. 263-277, 2007.

[28] Favela, Jesus, et al, "Activity recognition for context-aware hospital applications: issues and opportunities for the deployment of pervasive networks," *Mobile Networks and Applications*, vol. 12, no. 2-3, pp. 155-171, 2007.

[29] Zhou, Shandan, et al, "A context-aware reminder system for elders based on fuzzy linguistic approach," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9411-9419, 2012.

[30] Kiefer, Peter, et al, "Exploring Context-Sensitivity in Spatial Intention Recognition," *BMI*, vol. 296, pp. 102-116, 2007.

[31] Krauthausen, Peter, et al, "Situation-specific intention recognition for human-robot cooperation," In *KI 2010 Advances in Artificial Intelligence*, pp. 418-425, Springer Berlin Heidelberg, 2010.

[32] Pereira, Luís M, "Intention recognition via causal Bayes networks plus plan generation," In *Progress in Artificial Intelligence*, pp. 138-149. Springer Berlin Heidelberg, 2009.

[33] Röning J and Jukka R, "Context-Aware Mobile Robots: Part of Smart Environment,"

[34] Schrempf, Oliver C, et al, "Tractable probabilistic models for intention r±ecognition based on expert knowledge," In *Intelligent Robots and Systems, IROS 2007, IEEE/RSJ International Conference on*, pp. 1429-1434, 2007.