ROBOT CONTROL VIA DIALOGUE

Arkady Yuschenko

Abstract: The most rational mode of communication between intelligent robot and human-operator is bilateral speech dialogue using a problem-oriented language. The dialogue mode of control raises the problem of compliance between the human and robot perception of external world, logic, behavior planning and decision making. We suppose that linguistic variables and fuzzy logic is the most suitable approach for the mentioned problems. The model of the world formed on the base of natural space-time relations allows formulating the basic robot operations in terms of linguistic variables and to solve the problem of planning of robot behavior in previously undetermined environment. Using the fuzzy neural networks it is possible to train robot to fulfill complicated operations by human-operator. Relationship between fuzzy logic approach and the procedures of speech recognition are also under consideration

Keywords: mobile robot, fuzzy logic, human operator, remote control, artificial intelligence, operation planning, robot learning

Introduction

Robot is a technical system for autonomous work in previously undetermined environment. Mobile robots are normally equipped with manipulators, different sensors, including vision systems and on-board computer. Robot may be treated as an intelligent robotic system (IRS) because it is capable to describe an image of the current situation, to analyze the environment and separate objects, and to plan its own behavior necessary to reach the aim stated by human. So robot is a single technical system capable for active cognitive behavior. Such systems are usually controlled by human operator because his experience and intelligence are necessary to fulfill the hazardous and responsible operations. But the mode of control of intelligent robot is sharp different from the tradition modes. The control signals now are the speech commands using a professional problem-oriented language. Feedback is the observation of the robot behavior (if possible) and speech reply from robot to operator in the cases of indeterminacy or lack of information. So the most rational mode of communication between robot and operator is bilateral speech dialogue using a problem-oriented language close to natural speech. Intelligent robotic system controlled via robot-human dialogue is not more a "master-slave" system, as robot now is an equal participant of the process; IRS may be treated as a "master-assistant system". The dialogue mode of control raises the problem of compliance between the human and robot perception of external world, logic, behavior planning and decision making. Some of the problems are under consideration below.

Environment representation and problem of understanding

We suppose that linguistic variables and fuzzy logic may be the most adequate means to solve the problem of external world representation in a human-controlled IRS. The "mental" world model of robot is based on the corresponding representation of external world in human mind. Such description includes the description of physical objects in the environment as well as spatial and temporal relations between them.

To describe the spatial relations of the current scene, extensional and intensional types of relations are applied The former are represented by relations that describes location and orientation of objects. For example *a1 is far, to the right, and a little above a2.* The latter include the relations like R_1 *– to be adjacent to;* R_2 *- to be inside of; etc.* Extensional relations are determined by membership functions, which in its turn may be determined experimentally using the statistically processing estimation of the same spatial relations by human-operators. As for intensional relations – usually we used two unary relations - R_{00} – *to be horizontal* and R_{01} – *to be vertical*, as well as 28 basic spatial binary relations. Other may be determined from the basic ones as a conjunction [Pospelov, 1989].

The space situation is determined as a set of binary frames <*object1-relations-object2*>. Usually the observer, i.e. the robot itself, is one of the objects. For every chain of binary frames may de determined the fussy relation between the initial and the terminal elements of the chain. It make it possible to provide the mobile robot navigation through the fuzzy determined map. Also is possible to change the point of view. For example, to change the point of view of robot (in relative coordinate system) to point of view of operator, working in the inertial coordinate system [Yuschenko,2002]. Since the environment is ever-changing due the motion of the observed objects as well as to the motion of the robot itself, the scene description changes in time respectively. This circumstance requires that we take into account not only spatial but also temporal relations in the external world, such as *to be simultaneous with, to be prior to, to follow* et.c. The temporal relations evidently need the memory of the preceding situations.

The names of specified objects of the current scene and space-temporal relations between them form the thesaurus of the situational problem-oriented language (SPOL) for human-robot communication. The situation now may be represented as a set of binary relations. Scene description allows a formal semiotic representation that uses the spatial-temporal relations logic. So, a complex relation *a_1 is on the surface S far and to the right* can be written as $(a_1\ R_8\ S)\&(a_0\ d_5\ f_7\ a_1)$, where a_0 – is the observer, with respect to whom the distance and orientation relations are formulated, R_8 - *to be on the surface,* d_5 is *to be far,* and f_7 is *to be to the right.*.

The world model is sufficient for robot navigation but does not sufficient for manipulation with the objects of the world. The model of world is incomplete without sensation. Robotic system is equipped with tactile sensors, force-torques sensors and other necessary types of sensors. In make it possible to introduce to the world model such terms as heavy or light, warm or cold, smooth or rough [Roy,2003]. All the terms are fussy and may be represented in the knowledge base of robot by corresponding membership function. As show our results the linguistic representation of feeling is possible but doesn't usually effective in robotics. The same as for human the reflective robot motion is a synergetic complex perception-action. That is why for robot is more effective to form the same complexes by previous training using teachable fuzzy (or hybrid) neural networks [Vechkanov 2002]. So the robot model of world seems to be of hybrid type: linguistic relations are complemented by sensory stereotypes.

As far as robotic system can present the image of the scene constructed from the sensors data in linguistic terms, it may describe the situation for the human by text in SPOL or by speech. And vice versa if operator represents the situation by speech, robot can present the graphic image of situation with before known objects. The cognitive activity of robot makes it possible to transform one form of presentation into another which may be determined as "understanding" the situation by robot.

Operations description and commands

External world description allows formalizing the robot's operations within it [Yuschenko,2002]. We assume that complex operations performed by robot can be represented as a sequence of a relatively few types of elementary operations. These are define in advance and are stored in the IRS knowledge base as frames of typical operations. A frame of this kind contains linguistic variables based description of the aims of an operation, the initial stage scene, and the preconditions for the feasibility of the operation. The latter may depend on the specific situation, the capabilities of the robot in question, and the properties of the object of the operation. Thus, the structure of a typical operation frame is as follows: <*operation name*> <*operation object*> <*initial situation* *(modifier of place)*> <*target situation*> <*operation feasibility conditions (preconditions)*>.<*additional details*>. For example: <*move*> <*object A*> <*object A on B*> <*object A on C*> <*object A is free*> <*install object A shock-free*>. While performing technological operations this frame should sometimes have an extra slot <operation performance method (modifier of manner)>.

Preconditions are one peculiarity of the discussed operations description approach. Generally, all preconditions can belong to one of at least three types: a) situational, e.g. the condition *object A is free* means that *there are no other objects on object A*; b) preconditions stipulated by the robot's capabilities: *the robot is equipped with the gripper suitable for type and size of the object*; and c) preconditions connected with the peculiarities of the object: *the object is a rigid body and can withstand the force developed by the gripper without any damage*. The operator can control a robotic system directly by giving the names and aims of the typical operations in the problem-oriented language, e.g., forming the commands as ‹*move object A to plane C*› ‹*insert shaft A into orifice O*›. Such commands call the full frame as before. Some slots of the operation frame may be empty. To fulfill them robot may address queries to the operator with corresponding questions. Another way of the command complication is the cognitive activity of robot, which can change the point of observation or investigate the manipulation object using the sensors. Preconditions description may not always be complete in the sense that some of them may not be defined. For example, it may not be known whether there is free space on plane C, on which object A is to be put. Then a query to the cognitive operations base is formed, and an operation is selected for examining plane C that is supposed to provide for filling in the empty slot. The system can also formulate address queries to the operator, if cognitive actions yield no results or uncertainty persists.

The description of typical operations expands the situation description thesaurus introduced above. Another problem for complicated sentences is the syntax of the SPOL. As may be seen above the problem is solved by the rigid structure of the operation frame. It is not the best way for operator, but it allows to solve the problem of separation of words in continuous speech. The commands perceived by robot may be corrected via dialogue. Possibility of IRS to transform the information from one form into another mentioned above allow to present the goal situation in graphic form and to use means of computer aim-pointing. Last time other ways to command by robots are under consideration such as arm movements and gesture [Chernakova, 2009].

Complex operations planning

A distinctive feature of planning procedure in robotics, as compared to numerous methods of AI- planning, is the possibility of continuous comparison of the real situation observations and the conditions defined during the planning stage. The contradiction between the real and prescribed situation generate the plan to resolve contradiction between the reality and it desirable image [Magazov, 2007]. In the case when preconditions of prescribed operation are not satisfied the operation can not be classified as elementary ones. For example the object is to be transported from initial place to another but the precondition of situation type is not satisfied (object is not free). The contradiction emerged from the prescribed and real conditions result to find another operation to solve the contradiction (to make the object free). But preconditions of new operation may be also not satisfied. At last a chain of consented elementary operations will be constructed in such way that result of n-th operation meets the preconditions for the (n+1)-th operation and the last operation in the chain achieve the aim. This procedure can be represented as directed graph, with its root being the target situation [Yuschenko, 2005].

During the realization of the plan we propose that comparison of the real and planned situation is fulfilled after every stage of the procedure. If after the actual completion of n-th operation the prescribed conditions are not achieved, the planning process is repeated, with the current situation being assumed as the initial.

The conflict resolution approach is rather similar to human cognitive activity while planning actions, which is also based on comparing the operative image of the situation and the target image. The conflict resolution principle application requires a further extension of IRS control language. Besides the typical operations we now need a thesaurus for situational conflicts resolution by means of performing typical operations. If spatial relations are intentional then each type of conflict induces its own typical operation to resolve it. For example: if the aim is: *(a_1 R_8 S),* i.e. object a_1 is on the surface S, while in fact *(a_1 ⌐R_8 S),* then the conflict induces a typical operation *move a_1 to S* . If the aim is defined as *(a_1 R_2 C),* i.e. *shaft a_1 is inside orifice C*, while observation results show *(a_1 ⌐R_2 C),* then a typical operation is induced: *insert a_1 into C.* If the condition *a1 is free* is necessary for further operations, while in fact we have: *(a_2 R_8 a_1),* i.e. *a2 is on a1*, then a typical operation *remove a_2 from a_1* is induced. One can easily proceed with this list of action that resolve intentional type conflicts.

If the relations are extensional there is no need for a special vocabulary for matching the situation with the required typical operation. Conflict can be resolved by performing a typical operation aimed at the relation specified as its precondition. If a mobile robot R is expected to in position *(R d_1 f_1 N)* with respect to observer *N*, while in fact a different conditions holds true: *(R d_2, f_2 N),* then the required operation will be defined in the form of: *move robot R from position (R d_2 f_2 N) to position (R d_1 f_1 N).*

Robot learning by neural fuzzy network

A disadvantage of the existing approach is that the operator has to define the rules for different situations beforehand, hence the situations should also be known in advance. The rules of behavior cannot be formalized in advance when the human control robots using his personally sensory skill. For example a skillful human operator can successfully control the mobile robot movements in environment with complicated obstacles but can not describe the rules he used. Another example is the assembly of the “shaft-hole” type.

For such cases operator guides the robot through typical situations after which information is processed in, e.g. teachable fuzzy (hybrid) neural networks. The network ANFIS (adaptive neuro-fuzzy intelligent system) type was proposed for mobile robot control in the environment with obstacles. The system able to teach itself using the telemetry data received for skillful human operator remote robot control. The robot under consideration had three pairs of caterpillars and the problem to control all of them in real time scale was complicated for human. The task was to teach the mobile robot to get over the typical obstacles by itself. In the experiment the robot was controlled by experienced operator. The input variables were the current angles of the robot platform orientation and the torques of drivers. As the output variables the controlled angles of four robot caterpillars were considered. The training of the net was realized by error back propagation method. The data obtained during experiment by telemetry had been processed by cluster analysis method. It proved to determine both the typical situations and the corresponding control signals formed by operator. After the training procedure the fuzzy network showed results of control close to the same of human control [Vechkanov 2002]. As a matter of fact the hybrid network constructed could by itself formalize the fuzzy rules which human operator could not determine. The fuzzy mode of control allows the robot to get over the obstacles of the prescribed type for wide range of their parameters. But as new situation critically differs from the previous ones the training process has to be repeated again. All the previous tunings may be remembered in the system memory, so the robot “experience” is expanding after all new training cycles.

The same learning control systems may be applied in other types of robotic system to form the “skill” of the artificial system. A good example of such system is the industrial manipulation robot for assembly operations of “shaft-hole” type using a 6 d.o.f. forces-torques sensor. The number of possible situations during such operations is not large and all of them may be described beforehand. So the task of the fuzzy system is to recognize the situation and form the fuzzy rules similar to those for experienced operator. The experiments showed that for situation recognition in this case may be reliably applied an artificial neural network with one hidden layer.

As a matter of fact the reflective behavior of robots use mainly the sensor information so it effectiveness may be improved using other sources of information. The vision system may recognize the type of the obstacle and to adjust the fuzzy controller previously. To recognize the type of the obstacle the robot control system may be taught using the simulation of the environment. The type of the obstacle also may be determined in dialogue with operator.

Another problem under consideration nowadays is to teach robot by itself (i.e. without the teacher), because there are many situations when operator has not necessary information of the robot position and state. For such situation the training without teacher may be realized on the base of trainable neural networks [Zhdanov, 2008]. Author treated the basic feature of such systems as robot emotion, which help the system to recognize the positive or negative result of its behavior in memorized situation. Most perspective direction seems to combine the previously trained system with possibility of self – teaching.

Speech interface

Speech interface consists of recognition and linguistic blocks. The recognition block is a device for transforming speech signals as well as interpreting them as separate words or phrases. The linguistic block performs the interpretation of statements into SPOL, as well as the representation of these statements in a semiotic form.

At present there are two most widely used methods of speech recognition: Dynamic Time Warping (DTW), or template matching, and hypothesis probability estimation using Hidden Markov Models (HMM). For continuous speech recognition one can employ template phrases construction, using the information on the grammar of the SPOL. The HMM method using hypothesis probability estimation with Viterbi algorithm (beam-search) allows to recognize continuous speech almost independent of the speaker. However this method requires a high quality and expensive teaching speech database. N-grams method may be also successfully applied in robotics

Operator/s statements for IRS control can be formulated in the SPOL mentioned above. In the case of successful recognition we obtain, that (1): command is split into semantic units (*action, object, modifiers*), each of the part of the statement-command is assigned permissible and relevant values from the program data-base, feasibility of the operation is assesses, based on the current working scene status. The linguistic analyzer performs the syntactic and semantic decomposition of the statement which is supposed to result in filling the slots of the frame that describes operations. The slots of the frame are filled in with relevant subordinate parts of the command-sentence. The linguistic recognition stage output is a set of encapsulated frames that can be uniquely interpreted over the further stages of command-sentence completion. It was shown above that the sentences represented by linguistic frames can be expressed in the inner semiotic language as a sequence of symbols. Now the formal logic operation and verification of the commands are available.

A part of speech interface is the dialogue control system. There are some possible forms of dialogue scenarios. For example, dialogue for correction of the command syntax, for correction of the slots content, call for operator consent to fulfill the operation and so on. The dialogue control system form the call from robot to human to point the necessary information.

The most complicated problem with speech interface is to provide the continuous operator speech processing. As for speaker dependence the problem is not crucial as there are few operators for a robot and the speech interface may de adjusted individually. The problem will be much more complicate in the case when robot has to support the dialogue with everybody.

Conclusion

The preliminary research has shown that the implementation of speech dialogue type control mode for a robotic system by way of formulating separate commands is inefficient. It is necessary to develop a speech interface meant focused on the use of the situational problem-oriented language similar in its structure to natural language. This allows a substantial simplification of the task of robot control, as it no longer requires any special skills from the operator. There are however a number of tasks in this field that are yet to be solved. In particular, the speech interface application for teaching the robot, rather than merely controlling it. We also attribute crucial importance to the psychological aspects of interaction between a human and an intelligent robotic system, connected with “mutual” ideas about the situations and reasonable behavior. The final aim of the robot control via dialogue investigation seems the including robots to humanitarian socium where everybody can communicate with robots using practically natural language.

Bibliography

[Chernakova, 2009] S.E.Chernakova. Simulation of the intelligent mechatronic systems and robots training with movement demonstration. In: Mechatronika, Automation , Control, № 2, 2009, p.p.53-56

[Magazov, 2007]. S.S.Magazov. Cognitive processes and models. URSS, Moscow, 2007.

[Pospelov, 1989] E.Ju.Kandrashina, L.V.Litvinceva, D.A.Pospelov. Representation of knowledge of time and space in intelligent systems. – Nauka - Phys.matt.lith., Moscow, 1989 .

[Roy,2003] K.-Y.Xsiao, N.Mavridis, D.Roy. Coupling perception and simulation. Steps towards conversational robotics. In: Proc.IEEE/RSJ Int. Conf on Intelligent robotics and systems, 2003.

[Yuschenko,2002] A.S.Yuschenko. Robot distance control using fuzzy concepts, In: Iskusstvennyi intellect, Vol. 4, NAS Ukraine, 2002, pp. 388-396

[[Yuschenko,2005] A.S.Yuschenko. Intelligent planning in robot operation. In: Mekhatronika, Vol.3, 2005, pp. 5-18

[Vechkanov 2002] V.V. Vechkanov, D.V. Kiselev, A.S.Yuschenko. Adaptive system for mobile robot fuzzy control. In: Mekhatronika, Vol. 1, 2002, pp. 20-26

[Zhdanov, 2008] A. A..Zhdanov . Autonomous artificial intelligence. BINOM, Moscow, 2008.

Authors’ Information

Arkady Yuschenko – Professor, head of Department “Robotic Systems”, Bauman Moscow State Technical University, 105005, Moscow, 2-nd Baumanskaya str., 5, Russia: e-mail: robot@bmstu.ru