

Mejora de la conducta ‘ask’ para perfeccionar la interacción humano-robot en el robot de servicio Golem-III

Emmanuel Maqueda
Facultad de Estudios Superiores
Cuautitlán
Universidad Nacional Autónoma de
México
Estado de México
dmaqueda71@gmail.com

Mario Rosas Otero
Facultad de Estudios Superiores
Cuautitlán
Universidad Nacional Autónoma de
México
Estado de México
maltz15@comunidad.unam.mx

Abstract— A partir del área de oportunidad para realizar la integración de un método dedicado a interacción humano-robot, específicamente para tratar de dar una pauta al flujo conversacional en las interacciones de Golem-III, robot de servicio de IIMAS, UNAM; se ha analizado que dichas interacciones deben ser llevadas a cabo de forma eficiente, para que tanto humano como robot puedan comprender una situación y su contexto de forma clara, debido a ello se optimizó el comportamiento ‘ask’, el cual se encarga de regir la conducta más básica de Golem-III para que se lleve a cabo la comunicación. Sobre todo, cuando se ejecuta en escenarios complejos y no cooperativos (o con nula cooperación por parte del humano), situaciones que son probables de suceder en un ambiente de competencia, como lo puede ser RoboCup.

Keywords— *SitLog, Human-Robot Interaction, Service Robot, Speaking Tasks, Service Robot Programming, The Golem-III robot.*

I. INTRODUCCION

El comportamiento ‘ask’, que después de su optimización es nombrado ‘contextualized_ask’ presente en Golem-III, es a grandes rasgos la forma en que el robot puede interactuar con un usuario de manera hablada, recibir información y mediante una gramática o modelo de lenguaje dar resolución al contenido de una entrada por voz. De esta forma el comportamiento (que cabe destacar es utilizado en cualquier acción o prueba que requiera interacción por medio de voz) ha sido optimizado para que el robot pueda acoplarse a la complejidad del escenario al que se enfrenta, mediante la selección que el programador de una prueba general podrá hacer de acuerdo al contexto en que el robot sea puesto a interactuar, previendo las posibilidades del accionar del usuario, con esto nuestra finalidad es que el robot pueda hacer que el turno conversacional, la atención y respuesta sean beneficiosas para los fines del proyecto.

Con el comportamiento ‘ask’ el robot Golem-III puede interactuar con un usuario de manera hablada mediante diferentes modelos de lenguaje que almacenan y organizan la información que posee el robot en la base de conocimientos (KB) referente a diferentes tópicos, respuestas a preguntas, nombres, lugares etc.

Sin embargo, en escenarios complejos y no cooperativos por parte del usuario el robot no contaba con una estrategia eficaz para llamar la atención y hacer que el turno conversacional

fuera lo suficientemente efectivo. Se planteó una estrategia que diera oportunidad de enfrentarse a los diferentes casos que podría enfrentar el robot en una competencia, o en casos de ambiente cada vez menos controlado, dichos casos, aunque diferentes, variados y quizá impredecibles se condensaron en tres posibilidades; todo esto con la finalidad de suplantar el anterior sistema, que por medio de un sonido ‘beep’, intentaba regir el ritmo en el turno conversacional, lo cual dejaba al muchas variables totalmente incontroladas, tal como la ausencia de un método que asegure y promueva la atención del usuario. Para ello se determinó en primera instancia asegurar que habrá comunicación con una persona que este circundante al robot, otra medida y teniendo como hipótesis el hecho de que el canal de comunicación del usuario está ocupado mientras Golem-III causa expectativa mediante movimientos y por consiguiente se logra la atención necesaria del usuario y que cuando esta se obtenga se liberara dicho canal deteniendo al robot para que la persona hable; también el hecho de que en circunstancias de incertidumbre en donde no se conoce la ubicación del usuario una conducta usual (por los humanos) es ubicar en donde se encuentra la persona y encararla, para posteriormente comenzar la interacción.

II. ESTRUCTURA Y JERARQUIZACIÓN DE MODOS DE INTERACCIÓN

En su estructura general el comportamiento ‘contextualized_ask’ tiene tres modos de posible selección, dichos modos se jerarquizan dependiendo de la proactividad que generará el robot para tratar de asegurarse que el usuario está poniendo atención a su conversación, así de acuerdo al contexto de la situación donde se quiera utilizar dicho comportamiento, el programador podrá elegir el grado de proactividad requerida por el robot para cada caso particular.

Para considerar estos tres casos se tomó como base el actuar en el turno conversacional humano, comportamiento que si se analiza a fondo no es tan fácil de sintetizar y caracterizar debido a que realmente la comunicación humana ha sido un proceso largo que se ha logrado desarrollar a partir de la evolución humana y se ha perfeccionado a lo largo de siglos. Sin embargo, aunque complicado como humanos generalmente tenemos una gran habilidad para conseguir una sincronización casi perfecta y moderar un turno conversacional. En este sentido para tener una comunicación efectiva, tener la atención del interlocutor y poder transmitir

un mensaje de manera efectiva utilizamos el lenguaje corporal, la visión, percepción inconsciente del movimiento de labios y músculos en el rostro entre otras; ¿Como un robot podría mantener una comunicación efectiva sin contar con todos estos mecanismos?

A. Modo Regular

En el primer caso, se intuye y se da por hecho que tenemos la atención del usuario, además que está frente a nosotros, la acción principal es asegurarse que si esté el humano frente al robot, utilizando una cámara para reconocer un rostro, similar a lo que sucede entre humanos cuando existe atención para recibir un mensaje y se está frente a alguien; dado esto se comienza el diálogo.

B. Modo Cooperativo

En este modo se toma en cuenta el lenguaje corporal como parte esencial de una comunicación efectiva, el robot hace uso de movimientos corporales y busca tener la atención del usuario para que el mensaje que se quiere dar sea recibido eficazmente, así mismo se verifica mediante el reconocimiento de un rostro con una cámara que el humano este frente al robot antes de empezar el dialogo.

C. Modo Contextualizado

Por último, en el modo contextualizado se busca al interlocutor, y se decide si hay que realizar un traslado y una orientación hacia el humano, se realiza dicha acción en caso de ser necesaria y posterior a esto se comienza la comunicación, este es el caso con más ambigüedad ya que se desconoce en donde se encuentra el usuario por lo cual hay que encararlo y posterior a esto comenzar el turno conversacional.

De esta manera se considera que se tiene una estrategia más robusta previendo las posibilidades que pueden existir en competencia y provocando una conducta del usuario para lograr llegar a una interacción humano-robot eficaz, en donde se logre una comunicación efectiva de emisor-receptor.

III. MÉTODOS

Para la realización de estos tres modos, se utilizó SitLog, un lenguaje y entorno de programación para la especificación e interpretación de ingeniería de comportamiento para tareas de robot de servicio de una manera simple y flexible.

Al momento de realizar el diseño y dar estructura al algoritmo, se tomó en cuenta las características propias y jerarquización de cada ‘modo de la conducta’ descritos anteriormente. De la misma manera, se realizó un análisis a nivel de hardware, las habilidades y limitaciones propias del robot.

A. Modo Regular

Como se expuso antes en este modo de conducta se requiere que el robot verifique que un usuario esta frente a él, además se asume que ya se tiene la atención del usuario, con esto se puede dar comienzo a la interacción. Debido a la estructura de SitLog, donde se trabaja por situaciones donde cada una de estas da un argumento de salida el cual determina el éxito o fallo de la situación, como uno de los puntos principales se incluyo en una situación un modelo de dialogo llamado *detect_body*, dedicado a la detección de personas mediante la estructura corporal; donde se pueden modificar parámetros como la distancia a la que se detecta a una persona,

numero de intentos que se realiza dicha acción y un argumento de salida que se menciona anteriormente.

En este sentido las expectativas de la situación en caso de que esta se lleva a cabo satisfactoriamente son que el robot realiza un acercamiento de 15 centímetros al usuario para favorecer la posibilidad que tiene el robot de escuchar al usuario de forma correcta; además el semblante del robot, pasa a representar felicidad (*mood(feliz)*), y continua hacia la siguiente situación.

En caso de que la situación no se lleve a cabo satisfactoriamente, el robot cambiara a un semblante triste (*mood(triste)*), e indicará verbalmente al usuario que se coloque frente a él iniciando nuevamente la situación de forma recursiva.

Cabe destacar que esta situación se utiliza en los demás modos de conducta para evitar que el robot realice una tarea sin un usuario frente a él.

```
[
  id ==> ask(regular),
  type ==> recursive,
  embedded_dm==>detect_body(nearest,3,Body_Position,
  Detect_Status),
  arcs ==> [
    success : [advance(0.15,_),mood(feliz)] => ask_r,
    error : [mood(triste), say('I dont see you please stand
    in front of me')] => ask(regular)
  ]
]
```

B. Modo Cooperativo

Para este nivel de interacción, en el sentido de incorporar el lenguaje corporal, se incluyeron dos movimientos, uno que le permite navegar llamada ‘nav’, el segundo movimiento llamado ‘grasp’, que es la posición en que se coloca el robot para tomar objetos; aunado a esto se agrega el avance de 15 centímetros, se eleva el robot a 1.25 centímetros respecto a su base y pone un semblante feliz (*mood(feliz)*), todo esto con la finalidad de llamar la atención del usuario.

En este caso esta situación es exclusiva ya que posteriormente se verifica si el usuario esta frente al robot, antes de comenzar el dialogo.

```
[
  id ==> ask(cooperativo),
  type ==> neutral,
  arcs ==> [
    empty:[advance(0.15,_),roboheight(1.25),
    switchpose('nav'),sleep,switchpose('grasp'),
    mood(feliz) ] => see
  ]
]
```

C. Modo Contextualizado

Por último, en este modo de conducta se asume que no existe la atención del usuario es nula, y que incluso se encuentra relativamente alejado de él, aunque si al alcance de su visión, debido a ello, primero se lleva a cabo el protocolo realizado en el modo de conducta cooperativo, para después verificar la posición y orientación de la persona, para que el robot se traslade, y encare al usuario (*approach_person*). Cuando llega al punto vuelve a verificar que el usuario siga en dicha posición, para proseguir con la comunicación. En caso de que la persona no se encuentre su cambio de visión realiza una rotación sobre su propio eje de 70° para buscar a la persona y de manera recursiva iniciar el protocolo descrito anteriormente.

```
[
  id ==> approaching,
  type ==> recursive,
  embedded_dm==>approach_person(body,Final_Position,
                                Status),
  arcs ==> [
    success : empty => ask_question,
    error : turn(70.0,_) => approaching
  ]
]
```

IV. ETAPA DE PRUEBAS

Se realizaron pruebas posteriores a la implementación de los modos de conducta en el comportamiento 'ask', dichas pruebas fueron realizadas en el Centro de Investigación en Diseño Industrial (CIDI), en UNAM. En dichas pruebas se registraron datos de éxito y fracaso en la interacción humano robot, para su posterior análisis (etapa en desarrollo). Para esto se creó un modelo de diálogo en el cual el robot promueva e induzca al desarrollo de las situaciones idóneas para que cada uno de los modos sean comprobados.

V. CONCLUSIONES

A falta de análisis estadístico de las pruebas realizadas las conclusiones serán meramente cualitativas en donde se considera el cumplimiento de la hipótesis planteada. Se logró

plasmar el comportamiento deseado tomando en cuenta las limitaciones del hardware del robot Golem-III, además de esto se explotó al máximo las habilidades con las que cuenta el robot, tales como incluir los módulos del rostro, brazos, navegación y visión para lograr captar la atención del usuario. Además de esto se incluyó una condicional en la cual no se pueda avanzar de situación hasta que haya una persona frente al robot, todo esto con la final de evitar ambigüedad y que no se avance en una prueba y/o tarea sin que un usuario este al menos en el campo de visión del robot.

De esta manera se espera que este desarrollo sea de utilidad para el Grupo Golem en desarrollos posteriores para la programación de tareas y pruebas, así como apoyo en la mejora de la interacción humano-robot en la que era necesaria una reestructuración completa, todo esto con vista en la próxima competencia internacional RoboCup 2020 en donde el éxito es el objetivo principal.

AGRADECIMIENTOS

Agradecemos el apoyo de los miembros de Grupo Golem por el soporte durante la implementación de la conducta. A el Dr. David Tinoco Varela, por la motivación y comentarios brindados y en especial a la Universidad Nacional Autónoma de México por permitirnos llegar a cumplir nuestros sueños.

REFERENCES

- [1] Mauricio E Reyes , Ivan V Meza and Luis A Pineda, 'Robotics facial expression of anger in collaborative human-robot interaction', In 'International Journal of Advanced Robotic Systems' January-February 2019: 1-13. pp 1-11, 2019.
- [2] Luis A. Pineda, Lisset Salinas, Ivan V. Meza, Caleb Rascon and Gibran Fuentes, 'SitLog: A Programming Language for Service Robot Tasks.', In 'International Journal of Advanced Robotic Systems', Vol. 10, pp. 1-12, 2013.
- [3] Jorge Garcia Flores, Meza Iván, Émilie Colin, Gardent, Claire, 'Robot experience stories: First person generation of robotic task narratives in SitLog', In Journal of Intelligent and Fuzzy Systems', Tomo 34, N.º 5, 2018.
- [4] Knapp ML, Hall JA, and Horgan TG. Nonverbal communication in human interaction. Boston: Cengage Learning, 2013.
- [5] Fong, T, Nourbakhsh, I, Dautenhahn, K. A survey of socially interactive robots. Robotics and autonomous systems 2003; 42(3-4): pp 143-166.
- [6] Seib V, Giesen J, Gruntjens D, et al. Enhancing human-robot interaction by a robot face with facial expressions and synchronized lip movements. In: Skala V (ed) Proceedings of the 21st International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG). 2013, pp. 70-77