

ADQUISICIÓN Y EVALUACIÓN DE UN CORPUS DE DIÁLOGOS MEDIANTE UNA TÉCNICA DE GENERACIÓN AUTOMÁTICA DE DIÁLOGOS

David Griol, Lluís F. Hurtado, Emilio Sanchis, Encarna Segarra

Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València. E-46022 València, Spain
{dgriol, lhurtado, esanchis, esegarra}@dsic.upv.es

RESUMEN

En este trabajo presentamos una aproximación para adquirir un corpus de diálogos mediante la interacción de un simulador de usuario y un simulador de gestor de diálogo. Inicialmente se define una selección aleatoria de respuestas para el funcionamiento de ambos módulos, evaluándose automáticamente el diálogo adquirido mediante la definición de un conjunto de condiciones de parada. Las probabilidades de las respuestas seleccionadas tras simular con éxito un diálogo se incrementan previamente a una nueva simulación. De este modo, es posible obtener un modelo de diálogo sin la necesidad de disponer de un corpus de diálogos para la tarea. En el artículo se resumen los resultados de la aplicación de esta metodología para la adquisición de un corpus de diálogos para el proyecto EDECÁN.

1. INTRODUCCIÓN

El aprendizaje de modelos estadísticos que permitan desarrollar los diferentes módulos de un sistema de diálogo ha despertado durante la última década el interés de la comunidad científica [1] [2]. Aunque en la literatura pueden encontrarse modelos para el diseño de gestores de diálogo basados en la definición por parte de un experto de un conjunto de reglas, durante los últimos años se han desarrollado aproximaciones basadas en el aprendizaje de un modelo estadístico que define el comportamiento del gestor de diálogo [3] [4] [5].

En este campo, hemos desarrollado una aproximación para gestionar el diálogo utilizando un modelo estadístico aprendido a partir de un corpus de diálogos [6]. Recientemente se ha llevado a cabo la adaptación de este modelo para desarrollar un gestor de diálogo en el marco del proyecto EDECÁN [7].

El éxito de las aproximaciones estadísticas depende del tamaño y la calidad del corpus de diálogos utilizado para realizar el aprendizaje del modelo. La adquisición y etiquetado de un corpus de diálogos con un número suficiente de diálogos para entrenar un buen modelo requiere

un esfuerzo considerable. Una solución para este problema consiste en el desarrollo de un módulo que simule las respuestas del usuario. En este campo se han desarrollado durante los últimos años diferentes técnicas para modelizar el comportamiento del usuario [8] [9] [10] [11].

En este artículo, presentamos una aproximación para adquirir un corpus de diálogos mediante la interacción de un simulador de usuarios y un simulador de gestor de diálogo. La aproximación propuesta se basa en la selección aleatoria de las respuestas del usuario y del sistema. Los únicos parámetros que se requieren para la adquisición son la definición de la semántica de la tarea (es decir, el conjunto de posibles actos de diálogo de usuario y de sistema) y un conjunto de condiciones que permitan descartar automáticamente los diálogos que no alcanzan el objetivo definido. Hemos utilizado esta técnica para adquirir un corpus para una de las tareas definidas en el proyecto EDECÁN. La tarea EDECÁN-UPV consiste en el diseño de un interfaz oral para reservar y proporcionar información sobre las instalaciones deportivas en nuestra universidad.

El corpus obtenido mediante la técnica de simulación automática se ha evaluado utilizándolo para el entrenamiento de nuestro gestor de diálogo estadístico. El gestor de diálogo aprendido se ha evaluado utilizando un conjunto de diálogos persona-persona proporcionados por el personal del Área de Deportes de la Universidad Politécnica de Valencia. Este corpus está compuesto por 150 diálogos (873 turnos de usuario). De este modo, en estos diálogos han participado usuarios que deseaban realmente realizar las diferentes consultas que proporcionará el sistema de forma automática.

2. DEFINICIÓN DE LA SEMÁNTICA DE LA TAREA EDECAN-UPV

De forma similar a la utilizada en muchos otros sistemas de diálogo, la notación seleccionada en la tarea EDECAN-UPV para la representación de los turnos de usuario y de sistema se basa en la utilización de actos de diálogo.

Este trabajo se ha desarrollado en el marco del proyecto EDECÁN subvencionado por el MEC y FEDER número TIN2005-08660-C04-02.

2.1. Actos de diálogo de usuario

Los actos de diálogo de usuario se representan mediante la notación clásica de frames (atributos y conceptos). Para la tarea EDECÁN se han definido un total de siete conceptos:

- Cuatro conceptos dependientes de la tarea, correspondientes a los tipos de consultas que puede solicitar el usuario: conocer la disponibilidad de pistas (*Availability*), realizar una reserva (*Booking*), saber las reservas que tiene vigentes (*Booked*) o cancelar alguna de ellas (*Cancellation*).
- Tres conceptos independientes de la tarea: *Acceptance*, *Rejection* y *Not-Understood*.

Se han definido un total de seis atributos, relativos a la información que debe aportar el usuario para completar las diferentes consultas contempladas por el sistema. Los atributos definidos son el deporte que se desea practicar (*Sport*), el horario para el que se desea la consulta (*Hour*), la fecha (*Date*), el tipo de pista polideportiva (*Court-Type*), el identificador de pista (*Court-ID*) y el número de orden correspondiente a la pista que se desea reservar (*Order-Number*).

A continuación se muestra un ejemplo de la interpretación semántica de una intervención del usuario:

Turno de usuario: *Quiero reservar una pista de squash para mañana por la tarde.*

Interpretación semántica:

(*Booking*)

Sport: squash

Date: mañana

Hour: tarde

2.2. Actos de diálogo de sistema

El etiquetado de los turnos de sistema se ha realizado de forma similar al de las intervenciones del usuario. Los conceptos definidos pueden clasificarse igualmente en dependientes de la tarea e independientes de la misma. Se han detallado un total de 18 conceptos dependientes de la tarea:

- Conceptos utilizados para informar al usuario del resultado de una determinada consulta: sobre disponibilidad de pistas (*Availability*), sobre la realización de una reserva (*Booking*), sobre las reservas actuales del usuario (*Booked*) o sobre la anulación de una reserva (*Cancellation*).
- Conceptos definidos para requerir al usuario los atributos necesarios para una determinada consulta: deporte (*Sport*), fecha (*Date*), hora (*Hour*) y tipo de pista (*Court-Type*).

- Conceptos utilizados para la confirmación de conceptos (*Confirmation-Availability*, *Confirmation-Booking*, *Confirmation-Booked*, *Confirmation-Cancellation*) y de atributos (*Confirmation-Sport*, *Confirmation-Date*, *Confirmation-Hour*, *Confirmation-Court-Type*).

- Conceptos relativos al gestor de aplicación: infracción de la normativa de reservas (*Rule-Info*) o para indicar la necesidad de seleccionar alguna de las pistas disponibles (*Booking-Choice*).

Se han definido un total de seis atributos, correspondientes a los cinco detallados para el etiquetado de los turnos de usuario (*Sport*, *Court-Type*, *Court-ID*, *Date* y *Hour*) y un atributo relativo al número de pistas que satisfacen los requerimientos del usuario (*Num-Courts*).

Seguidamente se muestra un ejemplo del etiquetado de una respuesta del sistema:

Turno de Sistema: *¿Le reservo la pista de squash 1 del pabellón para el 25 de julio de 20:00 a 20:30?*

Etiquetado:

(*Confirmation-Booking*)

Sport: squash

Date: 25-07-2008

Hour: 20:00-20:30

Court-Type: pabellón

Court-ID: 1

3. TÉCNICA PARA LA ADQUISICIÓN AUTOMÁTICA DE DIÁLOGOS

Como se ha comentado en la introducción, nuestra aproximación para la adquisición de un corpus de diálogo se basa en la interacción de un módulo simulador de usuario y un simulador de gestor de diálogo. Ambos módulos realizan una selección aleatoria de una de las posibles respuestas definidas para la semántica de la tarea (actos de diálogos de usuario y de sistema). Al principio de la adquisición, el conjunto de respuestas de sistema se define como equiprobable. Cada vez que se simula un diálogo con éxito, las probabilidades de las respuestas seleccionadas por el gestor de diálogo durante dicho diálogo se incrementan antes de simular un nuevo diálogo.

El simulador de usuarios proporciona conceptos y atributos que representan la intención del turno de usuario. De este modo, el simulador lleva a cabo las funciones de los módulos de reconocimiento automático del habla y de comprensión del lenguaje.

La semántica seleccionada para el gestor de diálogo se corresponde con las 25 posibles respuestas definidas para el sistema en la tarea. La selección de las posibles respuestas de usuario se llevó a cabo teniendo en cuenta la semántica definida para para el módulo de comprensión.

Adicionalmente, se ha desarrollado un módulo que realiza la generación de errores y la incorporación de medidas de confianza. Esta información modifica los frames

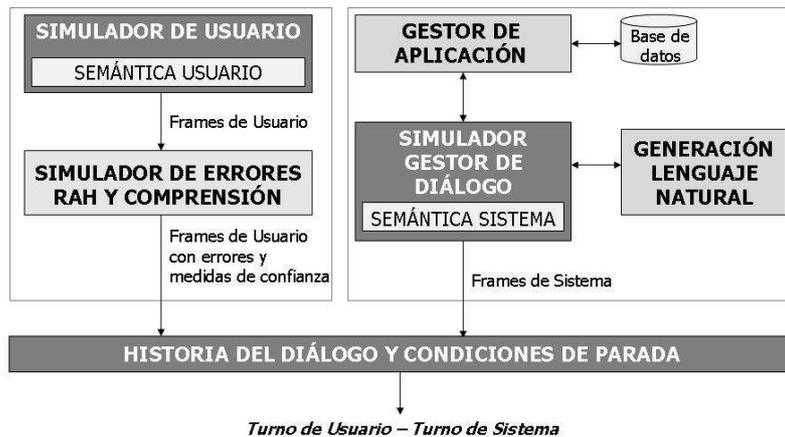


Figura 1. Esquema de módulos que componen la técnica de simulación de diálogos propuesta

generados por el simulador de usuario. De forma experimental, hemos detectado 2,7 errores por diálogo de un análisis de un corpus adquirido para la tarea DIHANA mediante la utilización de un reconocedor del habla y un módulo de comprensión reales. Este valor se puede modificar para adaptar el módulo simulador de errores con respecto al funcionamiento de cualquier módulo de RAH y de comprensión. La Figura 1 muestra la arquitectura de la metodología de adquisición automática de diálogos desarrollada.

Se selecciona una solicitud por parte del usuario para finalizar el diálogo una vez el sistema ha proporcionado la información definida en el objetivo del mismo. Los diálogos que cumplen esta condición antes de un número máximo de turnos se consideran exitosos. El gestor considera que el diálogo no ha alcanzado el objetivo prefijado cuando alguna de las siguientes condiciones tiene lugar:

- El diálogo excede del máximo de turnos.
- La respuesta seleccionada por el gestor de diálogo se corresponde con una consulta no requerida por el simulador de usuarios.
- El módulo gestor de la aplicación genera un mensaje de error debido a que el simulador del usuario no ha proporcionado la información obligatoria necesaria para llevar a cabo una consulta a la base de datos.
- El generador de respuestas proporciona un mensaje de error cuando la respuesta seleccionada por el gestor implica el uso de datos no proporcionados por el simulador del usuario.

La Tabla 1 resume las estadísticas de la adquisición de un corpus de diálogos para la tarea EDECÁN-UPV. Se definió un conjunto de 15 escenarios que especifican los objetivos de los diálogos, simulándose un total de 100.000 diálogos.

Diálogos simulados	100.000
Diálogos con éxito	2.521
Diálogos diferentes	1.973
Número de turnos de usuario por diálogo	4,2

Tabla 1. Estadísticas de la adquisición del corpus mediante la técnica de simulación de diálogos

4. EVALUACIÓN

El corpus descrito en la sección previa se ha utilizado para aprender un gestor de diálogo estadístico para la tarea EDECÁN-UPV de acuerdo con la metodología presentada en [7]. El corpus proporcionado por el Área de Deportes de nuestra universidad se ha utilizado como conjunto de evaluación para evaluar así el comportamiento del gestor de diálogo con un corpus adquirido con usuarios reales.

Hemos definido tres medidas para evaluar el funcionamiento del gestor. Estas medidas se calculan comparando turno a turno la respuesta automáticamente generada por el gestor de diálogo con respecto a la respuesta de referencia anotada para dicho turno en el corpus proporcionado por el Área de Deportes. La primera medida es el porcentaje de respuestas que coinciden con la respuesta de referencia en el corpus (*%exacta*). La segunda medida es el porcentaje de respuestas que son coherentes con el estado actual del diálogo (*%correcta*). Finalmente, la tercera medida es el porcentaje de respuestas que se consideran erróneas con respecto al estado actual del diálogo y ocasionarían el fallo del mismo (*%error*). Estas dos últimas medidas se han obtenido tras la revisión manual de las respuestas proporcionadas por el gestor de diálogo.

La Tabla 2 muestra los resultados de la evaluación del gestor de diálogo. Los resultados obtenidos tras la experimentación muestran que el gestor de diálogo aprendido se adapta correctamente a los requisitos de la tarea

EDECÁN-UPV, proporcionando un 89,8 % de respuestas que son coherentes con el estado actual del diálogo. Además, el 75,3 % de las respuestas coinciden con la de referencia en el corpus. No obstante, el porcentaje de respuestas proporcionadas por el gestor que pueden ocasionar el fallo del diálogo es considerable (3,9 %). Además, hay un 6,3 % adicional de respuestas que no suponen el fallo del diálogo, pero no son coherentes con el estado actual del mismo (por ejemplo, respuestas que requieren al usuario información que ya ha proporcionado previamente). Queremos reducir ambos porcentajes mediante la incorporación de nuevos diálogos al corpus inicial.

	%exacta	%correcta	%error
Respuesta del gestor	75,3 %	89,8 %	3,9 %

Tabla 2. Resultados de la evaluación del gestor de diálogo estadístico desarrollado para el proyecto EDECÁN

5. CONCLUSIONES

En este artículo, hemos descrito una aproximación para adquirir automáticamente un corpus de diálogos mediante la interacción de un simulador de usuarios y un simulador de gestor de diálogo. Para el desarrollo de ambos módulos definimos la semántica de las posibles respuestas de sistema y un conjunto de condiciones de parada que nos permiten evaluar automáticamente si el diálogo es exitoso o no. De este modo, el esfuerzo necesario para adquirir y etiquetar un corpus de diálogos se reduce notablemente.

El corpus que se ha obtenido mediante la aplicación de esta técnica a la tarea EDECÁN-UPV se ha utilizado para aprender un gestor de diálogo para dicha tarea, utilizando un modelo de diálogo estadístico. Hemos utilizado un corpus adquirido con usuarios reales para evaluar este gestor de diálogo. Los resultados de la evaluación muestran que es posible utilizar el modelo de diálogo aprendido para el desarrollo de un gestor de diálogo, generado sin mucho esfuerzo y con un alto rendimiento.

Actualmente, nuestro objetivo es llevar a cabo la evaluación de los diferentes módulos que componen el sistema de diálogo EDECÁN con usuarios reales. Esta evaluación se está llevando a cabo de forma supervisada, utilizando el gestor de diálogo presentado en este trabajo. Los diálogos que se adquieran se utilizarán además para mejorar el modelo de diálogo inicial.

6. BIBLIOGRAFÍA

- [1] S. Young, “The Statistical Approach to the Design of Spoken Dialogue Systems,” Tech. Rep., CUED/F-INFENG/TR.433, Cambridge University Engineering Department, Cambridge (Reino Unido), 2002.
- [2] E. Levin, R. Pieraccini, y W. Eckert, “A stochastic model of human-machine interaction for learning dialog strategies,” in *IEEE Transactions on Speech and Audio Processing*, 2000, vol. 8(1), pp. 11–23.
- [3] J. Williams y S. Young, “Partially Observable Markov Decision Processes for Spoken Dialog Systems,” in *Computer Speech and Language*, 2007, vol. 21(2), pp. 393–422.
- [4] H. Cuayáhuitl, S. Renals, O. Lemon, y H. Shimodaira, “Learning Multi-Goal Dialogue Strategies Using Reinforcement Learning with Reduced State-Action Spaces,” in *Proc. of the 9th International Conference on Spoken Language Processing (Interspeech/ICSLP)*, Pittsburgh (Estados Unidos), 2006, pp. 469–472.
- [5] F. Torres, E. Sanchis, y E. Segarra, “Development of a stochastic dialog manager driven by semantics,” in *Proc. of European Conference on Speech Communications and Technology (Eurospeech’03)*, Ginebra (Suiza), 2003, pp. 605–608.
- [6] David Griol, Lluís F. Hurtado, Encarna Segarra, y Emilio Sanchis, “A statistical approach to spoken dialog systems design and evaluation,” in *Speech Communication*, 2008, vol. 50, pp. 666–682.
- [7] D. Griol, L.F. Hurtado, E. Sanchis, y E. Segarra, “Adaptación de un Gestor de Diálogo Estadístico a una Nueva Tarea,” in *Revista de la Sociedad Española de Procesamiento del Lenguaje Natural*, 2007, vol. 39, pp. 231–238.
- [8] K. Scheffler y S. Young, “Corpus-based Dialogue Simulation for Automatic Strategy Learning and Evaluation,” in *Proc. of NAACL-2001. Workshop on Adaptation in Dialogue Systems*, Pittsburgh (Estados Unidos), 2001.
- [9] O. Pietquin y T. Dutoit, “A probabilistic framework for dialog simulation and optimal strategy learning,” in *IEEE Transactions on Speech and Audio Processing, Special Issue on Data Mining of Speech, Audio and Dialog*, 2005, vol. 14, pp. 589–599.
- [10] K. Georgila, J. Henderson, y O. Lemon, “Learning user simulations for information state update dialogue systems,” in *Proc. of the 9th European Conference on Speech Communication and Technology (Eurospeech’05)*, Lisboa (Portugal), 2005, pp. 893–896.
- [11] H. Cuayáhuitl, S. Renals, O. Lemon, y H. Shimodaira, “Reinforcement learning of dialogue strategies with hierarchical abstract machines,” in *Proc. of IEEE/ACL Workshop on Spoken Language Technology (SLT)*, Palm Beach (Aruba), 2006, pp. 182–186.