

CLASIFICACIÓN AUTOMÁTICA DE ACTOS DEL HABLA EN ÁRABE*

AUTOMATED SPEECH ACT CLASSIFICATION IN ARABIC

Lubna A. Shala, Vasile Rus y Arthur C. Graesser**

Resumen

Presentamos en este trabajo un método completamente automatizado para la tarea de clasificación de actos del habla en discurso árabe. La clasificación de actos del habla involucra la asignación de una categoría, obtenida a partir de una serie de categorías de actos del habla, a una oración para indicar la intención del hablante. Nuestra aproximación a la clasificación de actos del habla está basada en la hipótesis de que las palabras iniciales de una oración y/o sus partes-de-discurso son muy útiles para el diagnóstico del acto del habla expresado en la oración. Consideramos además la categorización semántica de estas palabras en términos de entidades nombradas y combinamos este enfoque con algoritmos de aprendizaje mecánicos para derivar automáticamente los parámetros de los modelos que utilizamos para implementar el enfoque. Presentamos experimentos y resultados obtenidos con varios modelos y algoritmos de aprendizaje automático en un corpus de 408 oraciones árabes.

Palabras clave: actos del habla, intención del hablante, categorización semántica.

Summary

We present in this paper a fully-automated method for the task of speech act classification for Arabic discourse. Speech act classification involves assigning a category from a set of predefined speech act categories to a sentence to indicate speaker's intention. Our approach to speech act classification is based on the hypothesis that the initial words in a sentence and/or their parts-of-speech are very diagnostic of the particular speech act expressed in the sentence. We also consider semantic categorization of these words in terms of named entities and combine this approach with machine learning algorithms to automatically derive the parameters of the models we used to implement the approach. Experiments and results obtained with several models and machine learning algorithms on a corpus of 408 Arabic sentences are presented.

* La investigación presentada en este trabajo ha sido financiada por la Fundación Nacional de Ciencia (*The National Science Foundation*) (BCS#0904909).

** Department of Computer Science, The University of Memphis. Dirección: 38152, Memphis, TN, E-mail: lshala@memphis.edu, vrus@memphis.edu, graesser@memphis.edu

Key words: speech acts, speaker intention, semantic categorization.

1. Introducción

La investigación en Procesamiento del Lenguaje Natural en lengua Árabe (PNL-A) ha ganado un interés creciente en los últimos años, por numerosas razones, incluyendo los métodos computacionales básicos para procesarlo. En este trabajo nos focalizamos en la tarea discursiva de clasificación automática de actos del habla de lengua árabe, que según la información que tenemos al respecto no ha sido realizada antes. La tarea de clasificación de actos del habla incluye clasificar una contribución discursiva, p.ej. una oración, dentro de una categoría de un conjunto de clases predeterminadas de actos del habla que cumplen funciones particulares del discurso social (Stolcke, 2000). Nuestra meta global es inferir el estatus del hablante, líder vs. seguidor, basado en un análisis de la distribución de los actos del habla en el discurso del hablante. Por ejemplo, un discurso con muchas órdenes puede ser indicativo de un líder. Otras aplicaciones de la clasificación de actos del habla incluyen tareas mayores de PLN como resumen y traductores automáticos.

Numerosos investigadores han discutido la tarea de clasificación de actos del habla para inglés y otros idiomas (Stolcke, 2000; Panah & Homayounpour, 2008). En este trabajo presentamos un método de clasificación de actos del habla totalmente automático, para el discurso árabe. La tarea de clasificación de actos del habla implica clasificar una contribución discursiva, p. e. una oración, en una categoría entre un conjunto de categorías predeterminadas de actos del habla que cumplen funciones particulares del discurso social. En particular, hemos trabajado con el siguiente conjunto de categorías predefinidas: *aseveración, declaración, negación, evaluación expresiva, saludo, pedido indirecto, pregunta, promesa/denegación, respuesta a una pregunta y respuesta corta*.

El resto del trabajo se organiza de la siguiente manera. Las primeras secciones proveen una rápida visión general de nuestra taxonomía de los actos del habla y detallan nuestro enfoque para la clasificación de los actos del habla. Luego discutimos el sistema experimental y los resultados obtenidos, así como los problemas mayores que hemos encontrado. Finalmente, las conclusiones e investigaciones futuras son esbozadas.

2. Nuestro enfoque

Nuestro enfoque para la clasificación de actos del habla en árabe se basa en la hipótesis de que las primeras palabras de una oración y/o sus partes del discurso son señaladamente diagnósticas del acto del habla particular que se expresa en la oración. Por ejemplo, las oraciones árabes que empiezan con la partícula de interrogación “*إ*” o “*هل*” son usualmente preguntas, mientras que aquellas que comienzan con verbos que no tienen sujeto pueden ser órdenes, tal como se muestra más abajo en la Tabla 1.

Tabla 1. Initial Words/POS and Speech Acts

هـ /participio وقففت /verbo عم /preposición ل /sustantivo فيضلا /puntuación	Pregunta
أ /verbo بعج /preposición ل /sustantivo ل أو س ل /puntuación	Orden

Una primera decisión que tuvimos que tomar fue elegir el conjunto de actos del habla para nuestra tarea de clasificación, es decir, la taxonomía de preguntas. La taxonomía de los actos del habla elegida refleja un análisis comparativo de los enfoques pasados más importantes para la teoría de los actos del habla. Hemos consultado las taxonomías definidas teóricamente, propuestas por Bach y Harnish (1982), Austin (1962), Searle (1975), Dore (1974), y D'Andrade y Wish (1985). También estudiamos las más recientes taxonomías de Graesser y Person (1994), que fue desarrollada en el contexto de entradas de estudiantes en Sistemas Tutoriales Inteligentes y de Popescu et. al. (2007), que fue desarrollado para la Generación de Lenguaje Natural. La última elección de actos del habla tomó en consideración un análisis de características retóricas del árabe (Hussein, 2006) y del corpus seleccionado. Por ejemplo, hemos decidido colapsar dos categorías (*denegación* y *promesa*) que eran difíciles de diferenciar mediante marcadores en la serie de oraciones de nuestro corpus y también porque tenían baja incidencia, lo que hubiera llevado a estimación inadecuada de parámetros durante el proceso de aprendizaje automático. Los actos del habla seleccionados se asemejan a los de D'Andrade y Wish (1985), con algunas diferencias. Por ejemplo, mientras que nosotros tenemos dos categorías separadas para *Orden* y *Pedido indirecto*, D'Andrade y Wish los combinan en una categoría llamada *Pedidos* y *Directivas*. También unen *Respuestas a preguntas* con *Respuestas cortas* en la categoría *Reacción*. Además nosotros agregamos una clase separada para *Saludos*.

La Tabla 2 muestra las definiciones de la serie final de actos del habla así como ejemplos con su traducción en árabe.

Tabla 2. La taxonomía de nuestros actos del habla

Acto de habla	Definición	Ejemplo
Aseveración	Actos que reportan hechos, reglas de estado, actitud comunicativa	La situación está empeorando cada día. ايحوي عوست ظل احل
Declaración	Actos que cambian la realidad de acuerdo con la propuesta de la declaración	Renuncio. ل يفتس انا Estás despedido. لوصفم تن
Orden	Una dirección autoritaria o instrucción para hacer algo	Dame una prueba de eso اقلد و ل ع دح او ل يلد ين طر ع Limítate a responder a mis preguntas. يبتل يس ا و ل ع بعج ا طوق

Acto de habla	Definición	Ejemplo
Promesa/Denegación	Un compromiso verbal de una persona hacia otra, acordando hacer (o no hacer) algo en el futuro	Responderé todas tus preguntas luego del recreo. لصافلا دعب لفتل يسأ على ع بي ج أس No te interrumpiré más con mis preguntas. نال ا دعب يتل يسأ ابك عطاقأ نل
Pregunta	Una expresión de duda que invita o pide ser respondida	Cómo escapaste de todo eso? فكذ لك نم تبهره فيك
Respuesta a pregunta	Declaración para responder a una pregunta	Sí, he probado todas esas soluciones, pero ninguna funciona para mí ال ب لول حل ا هذه لك تبهر ج دول م عن بدي اف
Respuesta corta	Dar una respuesta corta a una pregunta, usualmente demostrando conformidad (aceptando) o disconformidad (negando) sin ninguna explicación	Sí. م عن No. ال
Saludos	Actos de bienvenida	Hola. اب حرم Gracias. اركش Yo estoy bien, cómo estás tu? فكلا ح فيك بري خب ان ا
Evaluación expresiva	Apreciación del valor de algo	Lo que dijiste es correcto جي حص هتلق ام Este es un punto muy importante ادح فواه قطقن هذه
Pedido indirecto	Una declaración que reclama una acción sin expresar órdenes	Podría repetir las categorías una vez más? عدرخأ ترم عاون ال ا ددعت نأ كن كمب له

3. Estructura experimental

Hemos experimentado con dos algoritmos de aprendizaje automático, Bayes ingenuo (*naïve Bayes*) y Árboles de Decisiones, para inducir los clasificadores de actos del habla para textos árabes. Para modelar la tarea de clasificación de actos del habla, utilizamos como rasgos las primeras 3, 4 o 5 palabras de una oración (el así llamado contexto inicial de la oración), las etiquetas de partes del discurso de estas palabras, y tanto las palabras como las etiquetas, es decir, los pares palabras-etiquetas. Las partes de discurso de las palabras fueron obtenidas automáticamente, utilizando un etiquetador árabe, AMIRA 1.0 desarrollado por Diab y colegas (2007). Este POS etiquetador tiene una exactitud reportada de 96,6%.

También consideramos la categorización semántica de las primeras palabras en una oración en términos de entidades nombradas tales como PERSONA, ORGANIZACIÓN,

LUGAR, TIEMPO, MONEDA, DEIDAD, PORCENTAJE, TÍTULO y VELOCIDAD. Para oraciones cortas (menos de 5 palabras, por ejemplo saludos), utilizamos un valor por defecto NULO para las palabras faltantes (p. e. la palabra #4 o #5 no existe en una oración corta de saludo) y un valor por defecto NINGUNO para entidades nombradas faltantes.

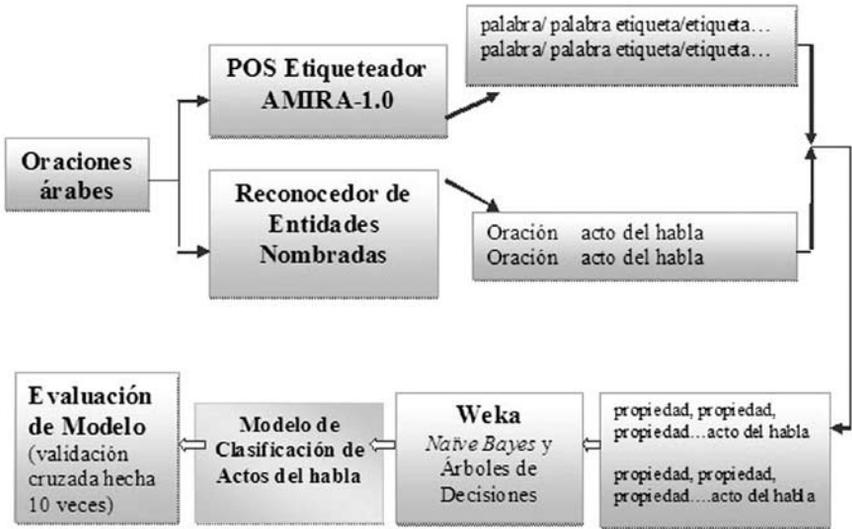
Además, hemos experimentado con varios otros modelos en los que usamos bigramas y trigramas de partes-del-discurso. La idea básica consiste en capturar información posicional/secuencial acerca de las partes del discurso, que puede ser importante a la hora de identificar actos del habla. Para obtener bigramas de partes-del-discurso, hemos simplemente concatenado dos partes consecutivas del discurso en un solo ejemplar. Como antes, solo consideramos partes del discurso para las primeras 3, 4 y 5 palabras de la oración. Armamos pares entre la primera parte-del-discurso y la segunda, entre la segunda y la tercera, y así sucesivamente, generando cinco ejemplares para las primeras cinco palabras. Estos modelos fueron utilizados en conjunto con los mismos algoritmos para inducir clasificadores de actos del habla. Introdujimos una parte de discurso inventada, COMIENZO, antes de la primera palabra para poder generar un bigrama para la primera palabra.

Para evaluar nuestro enfoque hemos recolectado 408 oraciones árabes de dos fuentes de noticias árabes: el periódico de Al-hayat y el canal de televisión de Aljazeera. Las oraciones fueron manualmente anotadas por un árabe parlante con un único acto del habla correspondiente a nuestra serie predefinida, que mostramos en la Tabla 2. Debido a la baja incidencia de la categoría “orden”, fue eliminada con la expectativa de poder incluirla en experimentos futuros, cuando una serie mayor de datos sería anotada. Las oraciones restantes conformaron nuestro conjunto de datos experimentales para construir clasificadores automáticamente y evaluar su performance.

Un enfoque de excelencia fue utilizado para la evaluación. Es decir, las oraciones recolectadas fueron anotadas manualmente por un árabe parlante cuyos actos del habla eran correctos. Esta serie de datos anotados es nuestro estándar de excelencia que sirve para evaluar la performance de nuestro método automático. El académico de lengua árabe también asignó entidades nombradas manualmente a las palabras, ya que no pudimos encontrar un reconocedor de entidades nombradas disponible sin costo para árabe, a pesar del hecho de que numerosos trabajos han discutido y propuesto tales herramientas (Benajiba, Rosso, & Benedíruiz, 2007; Benajiba, Diab & Rosso, 2008). Dadas las etapas incipientes del PLN árabe, esto no nos sorprende. La evaluación fue conducida basada en un método de 10 validaciones cruzadas en el cual el que la serie de datos disponible se divide en 10 sesiones y por cada sesión un clasificador es inducido. El clasificador es derivado de 9 sesiones y testeado en la siguiente sesión. La performance global consiste en el promedio de las 10 sesiones.

El proceso de clasificación de actos del habla se resume en el Cuadro 1, que figura abajo.

Cuadro 1. Proceso de clasificación de actos del habla



4. Resultados

Las Tablas 3-7 resumen los resultados de la clasificación de actos del habla para cada modelo en nuestro experimento, dando valores de exactitud (porcentaje de oraciones correctamente clasificadas) así como también valores kappa. El índice Kappa mide el grado de acuerdo entre actos del habla predichos por nuestro enfoque y actos del habla correctos de acuerdo con el estándar de excelencia, teniendo en cuenta los acuerdos al azar. Los mejores resultados globales fueron obtenidos con un modelo de 4-palabras-solamente y ninguna entidad nombrada (-EN), en combinación con el algoritmo de aprendizaje naïve Bayes (ver Tabla 4; exactitud=41.73, kappa=0.22).

Tabla 3. Resumen de los resultados con etiquetas POS (exactitud/kappas). Las dos primeras filas incluyen resultados con Entidades Nombradas (+EN)

Algoritmo	3-palabras	4-palabras	5-palabras
Naïve Bayes (+EN)	29,63/0,03	31,11/0,05	29,88/0,04
Árboles de Decisiones (+EN)	33,33/0,08	31,85/0,06	33,83/0,07
Naïve Bayes (-EN)	30,54/0,05	32,02/0,08	32,02/0,09
Árboles de Decisiones (-EN)	34,73/0,12	33,00/0,07	32,02/0,07

Tabla 4. Resumen de los resultados con palabras (exactitud/kappas). Las dos primeras filas incluyen resultados con Entidades Nombradas (+EN)

Algoritmo	3-palabras	4-palabras	5-palabras
Naïve Bayes (+EN)	38,52/0,16	39,75/0,19	39,51/0,18
Árboles de Decisiones (+EN)	32,35/0,08	34,32/0,06	34,32/0,03
Naïve Bayes (-EN)	41,48/0,20	41,73/0,22	38,27/0,18
Árboles de Decisiones (-EN)	32,35/0,08	34,32/0,06	34,32/0,03

Tabla 5. Resumen de los resultados con pares de palabras-etiquetas (exactitud/kappas). Las dos primeras filas incluyen resultados con Entidades Nombradas (+EN)

Algoritmo	3-palabras	4-palabras	5-palabras
Naïve Bayes (+EN)	39,01/0,16	38,02/0,16	38,52/0,17
Árboles de Decisiones (+EN)	32,35/0,05	31,85/0,02	32,10/0,03
Naïve Bayes (-EN)	40,74/0,18	41,23/0,21	38,02/0,18
Árboles de Decisiones (-EN)	32,35/0,05	31,85/0,02	32,10/0,03

Tabla 6. Resumen de los resultados con bigramas de etiquetas POS (exactitud/kappas). Las dos primeras filas incluyen resultados con Entidades Nombradas (+EN)

Algoritmo	3-palabras	4-palabras	5-palabras
Naïve Bayes (+EN)	33,58/0,05	33,83/0,04	34,32/0,05
Árboles de Decisiones (+EN)	33,58/0,12	35,56/0,12	35,31/0,11
Naïve Bayes (-EN)	37,28/0,12	35,80/0,10	35,80/0,10
Árboles de Decisiones (-EN)	34,32/0,13	35,56/0,11	35,56/0,11

Tabla 7. Resumen de los resultados con trigramas de etiquetas POS (exactitud/kappas). Las dos primeras filas incluyen resultados con Entidades Nombradas (+EN)

Algoritmo	3-palabras	4-palabras	5-palabras
Naïve Bayes (+EN)	38,77/0,19	38,02/0,19	38,02/0,19
Árboles de Decisiones (+EN)	29,63/0,06	27,90/0,04	28,89/0,01
Naïve Bayes (-EN)	39,26/0,20	38,27/0,19	38,27/0,20
Árboles de Decisiones (-EN)	28,89/0,05	28,64/0,02	28,89/0,01

5. Conclusiones y futuras direcciones

Los resultados precedentes revelan que los valores de precisión de la clasificación de todos los modelos fueron modestos, el mayor fue 47,73, con un índice kappa de 0,22. En general, los modelos que utilizan las palabras en sí mismas como rasgos presentaban mayor precisión y valores kappa más altos que aquellos que utilizan etiquetas de partes del discurso o pares palabra-etiqueta. El uso de entidades nombradas no tuvo efectos o, en algunos casos, disminuyó la performance. Los clasificadores basados en Árboles de decisiones funcionaron un poco mejor en contextos de 3-palabras, mientras que Naïve Bayes fue mejores para contextos de 4- o 5-palabras. Los resultados obtenidos del uso de secuencias (bigramas y trigramas) de etiquetas POS mostraron mayor precisión cuando utilizaron el clasificador Naïve Bayes pero menores o similares valores de precisión cuando usaron árboles de decisiones.

Un análisis más profundo del enfoque y de la estructura experimental reveló varios factores que podrían explicar los modestos resultados que obtuvimos y que podrían constituir futuros caminos hacia mejoras. *Primero*, la serie de datos utilizados en nuestros experimentos es relativamente pequeño, 408 oraciones, lo cual no es suficiente para estimar correctamente los parámetros de los modelos durante la fase de aprendizaje mecánico. *Segundo*, la serie de datos no resulta equilibrada cuando el número de ocurrencias varía ampliamente entre diferentes categorías de actos del habla. Por ejemplo, la serie de datos contiene 132 oraciones *declarativas*, 92 oraciones de *evaluación-expresivas*, pero solo 11 *pedidos-indirectos*. Hemos tenido que eliminar también la categoría *orden* debido a su baja incidencia, como ya hemos mencionado. *Finalmente*, luego de examinar detenidamente las oraciones clasificadas manualmente en la serie de datos, hemos llegado a la conclusión de que algunas oraciones no fueron correctamente clasificadas por nuestro árabe parlante. Estamos en proceso de recolectar una serie de datos más larga y equilibrada de oraciones árabes. Adicionalmente, nos encontramos en proceso de identificar varios árabes académicos que anotarían las oraciones independientemente y luego realizarían el proceso de acuerdo inter-anotador. No hemos utilizado varios árabes académicos hasta ahora ya que son escasos y que el proceso de anotación es costoso en términos de tiempo y de otros recursos, lo que convierte en un desafío la retención de dichos académicos por periodos largos de tiempo.

Referencias

- Austin, J.L. (1962). *How to Do Things with Words*. Cambridge, Mass.: Harvard University Press.
- Bach, K. and Harnish, R.M. (1982). *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.
- Benajiba, Y.; Diab, M.; Rosso, P. (2008). "Arabic Named Entity Recognition: An SVM-based approach". In *Proceedings of the International Arab Conference on Information Technology*.

Benajiba, Y.; Diab, M. & Rosso, P. (2008). "Arabic Named Entity Recognition using Optimized Feature Sets". In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*.

Benajiba, Y.; Rosso, P. & Benedíruiz, H.M. (2007). "ANERSys: An Arabic Named Entity Recognition System Based on Maximum Entropy". In *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing*, pp. 18-24.

D'Andrade, R.G. & Wish, M. (1985). "Speech Act Theory in Quantitative Research on Interpersonal Behavior". *Discourse Processes*, 8(2), pp. 229-59.

Diab, M.; Hacıoglu, K. and Jurafsky, D. (2007). "Arabic Computational Morphology: Knowledge-based and Empirical Methods" (Chapter 9). En Soudi, Abdelhadi; van den Bosch, Antal and Neumann, Gunter (eds.). Springer, pp. 159-179.

Dore, J. (1974). "A pragmatic description of early language development". *Journal of Psycholinguistic Research*, 3(4), pp. 343-350.

Graesser, A. & Person, N. (1994). "Question asking during tutoring". *American Educational Research Journal*, 31, 104-137.

Hussein, A. (2006). *Arabic Rhetoric: A Pragmatic Analysis (Culture and Civilization in the Middle East)*. London: Routledge.

Popescu, V.; Caelen, J. & Burileanu, C. (2007). "Logic-Based Rhetorical Structuring Component in Natural Language Generation for Human-Computer Dialogue". In *Proceedings of TSD 2007*. Pilsen, Czech Republic. LNCS/LNAI, Springer Verlag.

Panah, A.S. & Homayounpour, M.M. (2008). "Speech acts classification of Farsi texts, IST 2008". *International Symposium on Telecommunications*, pp. 539-542.

Searle, J. (1975). "A Taxonomy of Illocutionary Acts". In Gunderson, K. (ed.). *Minnesota Studies in the Philosophy of Language*. Minnesota: Univ. of Minnesota Press, pp. 334-369.

Stolcke, A.; Ries, K.; Coccaro, N.; Shriberg, E. & Bates, R. et al. (2000). "Dialogue act modeling for automatic tagging and recognition of conversational speech". *Computational Linguistics*, 26: pp. 339-373.

Fecha de recepción: 15/12/09

Fecha de aceptación: 10/05/10