

# Análisis de sentimientos en Twitter mediante modelos ocultos de Markov

Debashis Naskar<sup>1</sup>, Miguel Rebollo<sup>1</sup> and Eva Onaindía<sup>1</sup>

<sup>1</sup>Universitat Politècnica de Valencia, Camino de Vera s/n 4602 Valencia

Las técnicas actuales de análisis de sentimientos de la actividad de los usuarios en redes sociales asumen que el sentimiento del usuario corresponde exactamente con la emoción demostrada en el mensaje. Pero el sentimiento real del usuario es algo desconocido, que solo podemos determinar con precisión mediante la evaluación de la persona (mediante cuestionarios, reconocimiento facial o incluso mediante el estudio de la actividad cerebral).

Los modelos ocultos de Markov (HMM) [1] son modelos estadísticos que permiten determinar el estado de un sistema cuando existen parámetros desconocidos. En nuestro caso, los estados ocultos del modelo representan los sentimientos reales de los usuarios de Twitter y los estados observables se corresponden con las emociones que se demuestran en los mensajes escritos.

Los sentimientos se caracterizan a través del modelo de Russell [2], que identifica 16 sentimientos diferentes situados en un espacio circular en función de dos parámetros: la valencia y la activación (arousal). Estos dos valores se calculan para cada tuit en función de las palabras que incluye y de la carga emocional de las mismas.

Para comprobar la validez de la propuesta, se han extraído los tuits referentes a diferentes conversaciones identificadas mediante una etiqueta (hashtag). Los tuits recopilados se dividen en dos partes: con una de ellas se entrena el modelo y luego se comprueba su validez con el conjunto de tuits reservado para el test. Aunque hay una gran correlación entre los estados ocultos del HMM y las emociones mostradas en los tuits, los resultados obtenidos muestran que los modelos ocultos de Markov son más precisos que los modelos clásicos a la hora de predecir los sentimientos.

La principal complicación de los HMM es determinar el estado inicial. Se han examinado algunas de las inicializaciones habituales, pero las que mejor resultado han obtenido han sido aquellas que tenían en cuenta además de los sentimientos de los tuits enviados, los sentimientos de los mensajes leídos por los usuarios.

**Agradecimientos** Este trabajo ha sido financiado por el Ministerio de Economía, Industria y Competitividad bajo el proyecto con N° de expediente TIN2015-65515-C4-1-R.

[1] Baum, L.E. . *An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process*. Inequalities. 3: 1–8 (1972)

[2] James A. Russell. *Core affect and the psychological construction of emotion*. Psychological review, 374 110(1):145–172. (2003)

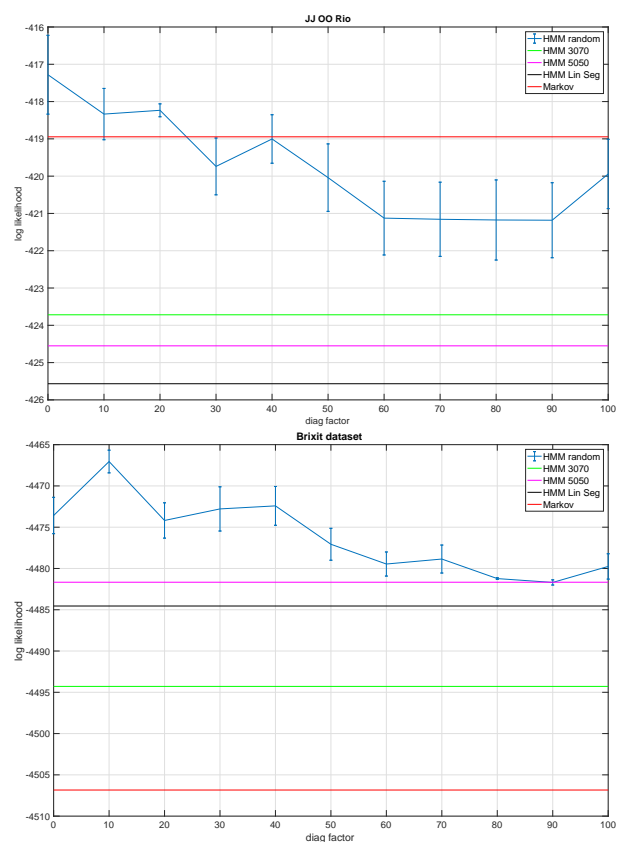


Figure 1: Comparativa de distintas configuraciones de HMM y el modelo de Markov clásico, según la log-likelihood, para el análisis de sentimientos. Los ejemplos corresponden a la final de Badminton en los JJ.OO. de Río y las consulta en UK sobre el Brexit