

Detección de perfiles de usuarios en la Web 2.0 desde el punto de vista emocional

Lea Canales, Patricio Martínez-Barco
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
{lcanales, patricio}@dlsi.ua.es

Resumen

Actualmente, los estudios sociológicos sobre los estados anímicos se realizan a través de la interpretación de encuestas individuales en las que se formulan preguntas acerca del estado emocional y del bienestar del encuestado. Sin embargo, es bien conocido que el proceso podría dar con interpretaciones erróneas. Es por ello, que el objetivo principal del proyecto de tesis es la creación de técnicas, recursos y herramientas que permitan evaluar y representar el estado emocional de los miembros de una sociedad, usando los canales de comunicación de la Web 2.0, ya que con la aparición de la Web 2.0 las aplicaciones tradicionales han evolucionado a aplicaciones enfocadas en el usuario final, donde es éste el que aporta contenido.

1. Introducción

El objetivo principal del proyecto de tesis doctoral es **la creación de técnicas, recursos y herramientas que permitan evaluar y representar el estado emocional de los miembros de una sociedad**. Esto se llevará a cabo mediante la correcta interpretación de los comentarios que escriben usando los canales de comunicación de la Web 2.0. y teniendo en cuenta la ubicación espacio-temporal de los mismos.

Poder determinar el grado de bienestar de un determinado conjunto social en un determinado lugar y en un rango temporal resulta ser de gran importancia. La utilidad de este sistema está demostrada por la cantidad de encuestas que renombrados organismos a nivel nacional como: el Centro de Investigaciones Sociológicas (CIS), Centro de Investigaciones de la Realidad Social (CIRES) o el Instituto Coca-Cola de la Felicidad¹, desarrollan con el objetivo de analizar y conocer el estado emocional y el bienestar social para distintas finalidades; tanto políticas, como sociales y comerciales. Más concretamente el resultado de estas encuestas se utiliza para analizar las causas que producen ese estado emocional o no (en sus varios grados) o para analizar la relación que existe entre un determinado estado emocional y las diferentes situaciones económicas, sociales o geográficas y así tomar las medidas necesarias.

Actualmente, los estudios sociológicos sobre los estados anímicos se realizan a través de la interpretación de encuestas individuales en las que se formulan preguntas acerca del estado emocional y de bienestar del encuestado. Sin embargo, es bien conocido que el proceso podría dar con interpretaciones erróneas producidas por la falta de sinceridad en el encuestado, la influencia de una situación excepcional ocurrida en el entorno de la persona en el

Copyright © by the paper's authors. Copying permitted only for private and academic purposes.

In: L. Alfonso Ureña López, Jose Antonio Troyano Jiménez, Francisco Javier Ortega Rodríguez, Eugenio Martínez Cámara (eds.): Actas de las V Jornadas TIMM, Cazalla de la Sierra, España, 12-JUN-2014, publicadas en <http://ceur-ws.org>

¹<http://institutodelafelicidad.com>

momento de formular la respuesta, o simplemente, por un error en la interpretación. Sin embargo, un enfoque para el estudio de los estados anímicos a través del análisis de la Web 2.0, presentaría como principales aportaciones las siguientes novedades: a) el estado emocional del sujeto se obtendría directamente de sus intervenciones, no de las respuestas a preguntas cuya fiabilidad depende de la sinceridad del encuestado; b) la ventana temporal del análisis a un sujeto puede estar abierta tanto tiempo como sea conveniente, horas, días o semanas, con el fin de evitar circunstancias extraordinarias que puedan proporcionar una visión errónea en un instante determinado; c) el análisis puede hacerse de manera automática mediante la monitorización de la red social a un individuo, grupo, zona geográfica e incluso a la totalidad de la comunidad participante en la red.

Por contra, el sistema siempre estará limitado a evaluar únicamente a los individuos que interactúan en la Web 2.0, un colectivo que, no obstante, es cada vez mayor y tiende a abarcar a toda la sociedad en próximas generaciones.

Hoy en día en la Web 2.0 encontramos mucha información sobre la opinión, los sentimientos, emociones y puntos de vista que tiene la sociedad sobre un amplio abanico de temas. Nos encontramos con una gran cantidad de datos ya que la Web 2.0 representa la evolución de las aplicaciones tradicionales hacia las Webs enfocadas en el usuario final². Nos referimos más concretamente a la Web Social, un conjunto de sitios web que se basan en una arquitectura de participación y donde el usuario se convierte en protagonista, aportando los contenidos: comparte fotos, escribe su diario en forma de blog, comenta lecturas, expresa su opinión sobre las noticias o eventos, etc. Ejemplo de este fenómeno pueden ser las diferentes plataformas de blogs (Blogger³ o WordPress⁴), redes sociales (Facebook⁵, Flickr⁶ o MySpace⁷) o servicios de microblogging (Twitter⁸, Khaces⁹ o Plurk¹⁰) disponibles en la Web.

Para la creación de técnicas, recursos y herramienta de evaluación del estado emocional nuestro análisis está enfocado al usuario, es decir, analizamos los comentarios realizados por los usuarios en un determinado intervalo de tiempo para poder determinar cual es su estado emocional y su grado de bienestar. Junto con expertos de Ciencias de la Salud hemos considerado que cambiando el enfoque del análisis podremos realizar un estudio de sentimientos de mayor calidad. De este modo seremos capaces de clasificar a los usuarios en varios rangos, definidos por los expertos en salud, en función de su estado emocional y grado de bienestar. El resultado final por lo tanto será: **ser capaces de plasmar unos perfiles de usuario en función de los sentimientos expresados por los mismos.**

En el apartado 2, presentamos el estado de la cuestión de las áreas que abarca el proyecto de tesis; en el apartado 3, presentamos el trabajo que hemos realizado hasta el momento para cumplir con el objetivo principal de la tesis; y en el apartado 4, indicamos las conclusiones obtenidas hasta el momento y planteamos nuestro trabajo futuro.

2. Estado de la cuestión

Después de haber presentado el contexto de nuestro trabajo y su necesidad, el apartado 2 está dedicado a ilustrar algunos de los trabajos previos más significativos en las áreas que abarca este trabajo.

Nuestra investigación se lleva a cabo desde el área de la *Inteligencia Artificial* (IA) o *Artificial Intelligence* (AI) y más concretamente desde sus subdisciplinas: *Informática Afectiva* o *Affective Computing* (AC), *Personality Computing* (PC) y *Perfiles de Usuario* o *User Profile* (UP).

El concepto *Artificial Intelligence* fue acuñado por John McCarthy en 1956 y lo definió como: "*La ciencia o ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes*"¹¹. Dentro de esta disciplina existen varias subdisciplinas entre las que se encuentran las mencionadas anteriormente.

La *Informática Afectiva* (*Affective Computing*) es aquella que surge de las emociones o de otros fenómenos afectivos¹². Su finalidad es conseguir que los ordenadores sean capaces de distinguir las emociones de los usuarios. Esta subdisciplina está organizada en diferentes modalidades en función de la fuente de información que utilicen

²<http://oreilly.com/web2/archive/what-is-web-20.html>

³<http://www.blogger.com/>

⁴<http://wordpress.com/>

⁵<https://www.facebook.com>

⁶<http://www.flickr.com>

⁷<https://myspace.com>

⁸<https://twitter.com>

⁹<http://www.khaces.com>

¹⁰<http://www.plurk.com/top/>

¹¹<http://www-formal.stanford.edu/jmc/whatisai/node1.html>

¹²<http://affect.media.mit.edu/>

para la detección de las emociones (p.ej., facciones de la cara, voz o texto). Nuestro proyecto está enmarcado dentro de la modalidad de texto, puesto que nuestra investigación se va a centrar en la correcta interpretación de los comentarios que escriben los usuarios en diferentes canales de comunicación de la Web 2.0. Dentro de esta modalidad encontramos diferentes trabajos basados en conocimiento, donde utilizan un recurso léxico para analizar el texto con el objetivo de identificar las palabras que predicen el estado emocional de los autores o lectores [CMP04, KTMA07, HLS07]. También existen aproximaciones basadas en aprendizaje automático [ARS05, SM08]. En [CM10] podemos encontrar un estado de la cuestión sobre *Affective Computing*.

En la subdisciplina *Personality Computing* se enmarcan aquellos trabajos que utilizan cualquier tecnología que implica la comprensión, la predicción y la síntesis de la conducta humana. En esta subdisciplina existen tres modalidades o subáreas: *Automatic Personality Recognition* (predecir la verdadera personalidad de un individuo a partir de su comportamiento), *Automatic Personality Perception* (predecir la personalidad de una persona dada según la percepción de los observadores) y *Automatic Personality Synthesis* (generación de personalidades artificiales a través de agentes personalizados) [MM14]. Consideramos que nuestro trabajo está enmarcado en la subárea *Automatic Personality Recognition*, puesto que nuestro objetivo es definir perfiles de usuarios que nos permitan saber cómo es la persona que se encuentra detrás de esos comentarios realizados en los diferentes canales de comunicación de la Web 2.0. Más concretamente en la modalidad cuya fuente de información es el texto, ya que realizaremos un análisis basado en texto escrito. Dentro de esta modalidad encontramos diferentes aproximaciones basadas en aprendizaje automático y/o recursos léxicos como: [MWMM07, LD08] donde analizan un conjunto de relatos escritos por estudiantes que han realizado el cuestionario *NEO-Five Factor Inventory* (NEO-FFI). Este cuestionario es uno de los utilizados para evaluar la personalidad de un individuo [MC04]. Otras aproximaciones analizan los textos escritos en diferentes blogs [GNO09, ON]. En cuanto a *Social Media* encontramos trabajos como: [GRET11, GRT12] donde además de analizar los comentarios de los usuarios, utilizan características como el número de seguidores, número de hashtags, número de amigos, etc. Y aproximaciones donde sólo utilizan estas características y no analizan el texto [QKSC11, BZC12].

En cuanto a los trabajos sobre la detección de perfiles de usuarios (*User Profile*) hemos encontrado principalmente trabajos cuyo objetivo es conocer los gustos de los usuarios para realizar recomendaciones personalizadas sobre productos y/o noticias [TM09, AGHT]. También hemos encontrado trabajos sobre perfiles de usuarios pero entendiendo como tal el conjunto de acciones o opciones que puede el usuario configurar para determinar su perfil dentro de una red social [GT, MVGD].

En este primer análisis de las tres subdisciplinas relacionadas con el proyecto de tesis, podemos extraer como primera conclusión, que no existen trabajos que fusionen la detección de emociones con los perfiles de usuario, es decir, que proporcionen perfiles emocionales de usuario.

Una vez analizado el estado de la cuestión y detectado nuestro nicho de investigación, en la siguiente sección detallamos el trabajo que hemos realizado para conseguir nuestro objetivo: la creación de técnicas, recursos y herramientas que permitan evaluar y representar el estado emocional de los miembros de una sociedad.

3. Trabajo realizado

Después de presentar los trabajos previos más significativos en las áreas relacionadas con el proyecto de tesis, en este apartado 3 vamos a detallar los trabajos realizados.

Para poder cumplir el objetivo principal del proyecto de tesis doctoral, hemos empezado a trabajar junto con expertos de la Facultad de Ciencias de la Salud de la Universidad de Alicante. Con ellos, hemos desarrollado un corpus emocional y hemos realizado una experimentación emocional.

3.1. Corpus emocional

Para la creación del recurso partimos del recurso léxico SEL (Spanish Emotion Lexicon) [SMjVj⁺12], el cual contiene 2.036 términos. Cada uno de los términos tiene asociada una de las emociones básicas: alegría (joy), miedo (fear), tristeza (sadness), rabia (anger), asco (disgust), sorpresa (surprise) y un factor de probabilidad de uso afectivo (Probability Factor of Affective, PFA). Este factor representa la frecuencia con la que se utiliza dicho término para expresar la emoción que tiene asociada.

Es uno de los pocos recursos léxicos escritos en español. Contiene un conjunto de términos en español que no son utilizados en España porque es un recurso desarrollado en latino-América y aunque sea el mismo idioma, en cada país las emociones se expresan con diferente terminología. Por ello, el recurso fue analizado y adaptado por un conjunto de psicólogos de la Facultad de Ciencias de la Salud de la Universidad de Alicante.

En primero lugar, realizaron una selección de los términos que se utilizan en España, quedando finalmente 1.827 términos. A continuación, evaluaron cada uno de los términos y se les asignó un grado de relevancia o peso por cada una de las emociones básicas.

3.2. Experimentación emocional

Realizamos una experimentación en colaboración con la Facultad de Salud de la Universidad de Alicante que hemos denominado experimentación emocional. Es una experimentación que realizamos con 15 estudiantes de la Facultad de Salud de la Universidad de Alicante. Se dividió en 4 sesiones en las que a través de diferentes imágenes se les provocaba a los estudiantes diferentes emociones básicas, concretamente: alegría, tristeza, asco y rabia. Después de mostrar las imágenes ellos debían escribir comentarios en Twitter¹³ en la relación a las imágenes que habían visto y que sentimientos/emociones habían producido esas imágenes en ellos. Todos estos mensajes fueron monitorizados y almacenados para su posterior análisis y procesamiento.

Las imágenes que utilizamos en la experimentación fueron extraídas del Sistema Internacional de Imágenes Afectivas (*International Affective Picture System, IAPS*) [LBC08], un conjunto de imágenes emocionales ampliamente utilizado.

4. Conclusiones y Trabajos futuros

Una vez detallados los trabajos realizados, en este apartado exponemos nuestras conclusiones y los trabajos futuros que nos planteamos.

Volviendo al comienzo de este artículo recordamos que el objetivo principal de nuestra investigación es la creación de técnicas, recursos y herramientas que permitan evaluar y representar el estado emocional de los miembros de una sociedad, a través de la Web 2.0.

En el estado de la cuestión hemos analizado las áreas que abarcan este proyecto: *Informática Afectiva* o *Affective Computing* (AC), *Personality Computing* (PC) y *Perfiles de Usuario* o *User Profile* (UP). Detectando un nuevo nicho de investigación en la detección de perfiles de usuario desde el punto de vista emocional.

El primer paso para alcanzar este objetivo final ha sido la creación de un corpus emocional y la realización de una experimentación emocional. Ambas tareas han sido desarrolladas conjuntamente con la Facultad de Ciencias de la Salud de la Universidad de Alicante.

Pero consideramos que éste es sólo el primer paso de nuestra investigación y por tanto nos planteamos los siguientes trabajos futuros:

- Incorporar el corpus emocional a nuestro sistema y proporcionar una primera aproximación basada en conocimiento.
- Analizar el uso de técnicas de aprendizaje automático en nuestro sistema, o incluso una aproximación híbrida donde también utilicemos el corpus emocional.
- Evaluar nuestro sistema con la información generada en la experimentación emocional.

5. Agradecimientos

Queremos agradecer a: a) la Facultad de Ciencias de la Salud de la Universidad de Alicante su colaboración en este trabajo. En concreto a Miguel Richart, a Juan Diego Ramos y Maria Jose Cabañeros; b) al programa de Formación de Personal Investigador (FPI) del Ministerio de Economía y Competitividad del Gobierno de España por su apoyo a través de una de sus becas pre-doctorales de investigación (BES-2013-065950); c) Ministerio de Economía y Competitividad del Gobierno de España por su apoyo a través de los proyectos TEXT-MESS 2.0 (TIN2009-13391-C04-01), LEGOLANG (TIN2012-31224), y ATTOS (TIN2012-38536-C03-03), al Gobierno de la Generalitat Valenciana por su apoyo a través del proyecto PROMETEO/2009/119 y a la Unión Europea, por la financiación del proyecto FIRST(FP7-ICT-2011-7).

Referencias

- [AGHT] Fabian Abel, Qi Gao, Geert-jan Houben, and Ke Tao. Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web.

¹³<https://twitter.com>

- [ARS05] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. Emotions from text: machine learning for text-based emotion prediction. *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354, 2005.
- [BZC12] Shuotian Bai, Tingshao Zhu, and Li Cheng. Big-Five Personality Prediction Based on User Behaviors at Social Network Sites. Technical report, 2012.
- [CM10] Rafael A Calvo and Senior Member. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. 1(1):18–37, 2010.
- [CMP04] Michael a Cohn, Matthias R Mehl, and James W Pennebaker. Linguistic markers of psychological change surrounding September 11, 2001. *Psychological science*, 15(10):687–693, October 2004.
- [GNO09] Alastair J Gill, Scott Nowson, and Jon Oberlander. What Are They Blogging About? Personality, Topic and Motivation in Blogs. *Proceedings of the international AAAI Conference on Weblogs and Social Media*, pages 18–25, 2009.
- [GRET11] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting Personality from Twitter. *2011 IEEE Third Int’l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int’l Conference on Social Computing*, pages 149–156, October 2011.
- [GRT12] J. Golbeck, C. Robles, and K. Turner. Predicting Personality with Social Behavior. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 302–309, August 2012.
- [GT] G Gee and H Teh. Twitter Spammer Profile Detection.
- [HLS07] Jeffrey T. Hancock, Christopher Landrigan, and Courtney Silver. Expressing emotion in text-based communication. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, page 929, 2007.
- [KTMA07] J Kahn, R Tobin, A Massey, and J Anderson. Measuring Emotional Expression with the Linguistic Inquiry and Word Count. *Am. J. Psychology*, 120:263–286, 2007.
- [LBC08] P.J. Lang, M.M. Bradley, and B.N Cuthbert. International affective picture system (IAPS): Affective ratings of pictures and instruction manual, 2008.
- [LD08] Kim Luyckx and Walter Daelemans. Using syntactic features to predict author personality from text. *Proceedings of Digital Humanities*, pages 146–149, 2008.
- [MC04] Robert R. McCrae and Paul T. Costa. A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36(3):587–596, February 2004.
- [MM14] Alessandro Vinciarelli Member and Gelareh Mohammadi. A Survey of Personality Computing. *IEEE Transactions on Affective Computing*, 2014.
- [MVG D] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You Are Who You Know : Inferring User Profiles in Online Social Networks.
- [MWMM07] F Mairesse, M A Walker, M R Mehl, and R K Moore. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research*, 30:457–500, 2007.
- [ON] Jon Oberlander and Scott Nowson. Whose thumb is it anyway? Classifying author personality from weblog text.
- [QKSC11] Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. *Proceedings of the IEEE International Conference on Social Computing*, pages 180–185, October 2011.
- [SM08] Carlo Strapparava and Rada Mihalcea. Learning to identify emotions in text. *Proceedings of the 2008 ACM symposium on Applied computing - SAC '08*, pages 1556–1560, 2008.

- [SMjVj⁺12] Grigori Sidorov, Sabino Miranda-jiménez, Francisco Viveros-jiménez, Ismael Díaz-rangel, and Sergio Suárez-guerra. Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. 2012.
- [TM09] Yasufumi Takama and Yuki Muto. Profile Generation for TV Program Recommendation Based on Utterance Analysis. 13(2):86–90, 2009.