

ATICA 2020

Aplicación de Tecnologías de la
Información y Comunicaciones
Avanzadas y Accesibilidad

OBRAS COLECTIVAS
TECNOLOGÍA 32

Luis Bengochea
Gerardo Contreras Vega
(Editores)

UAH

ANÁLISIS DE SENTIMIENTOS: DETECCIÓN DE SARCASMO EN ESPAÑOL A TRAVÉS DE EMOTICONES EN TWITTER

Caleb Faillace¹, Paola Britos^{1,2}, Ma. Florencia Pollo Cattaneo¹

¹Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)
Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina

² Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Medrano 951
(C1179AAQ) Ciudad Autónoma de Buenos Aires (Argentina)
{faillacec, flo.pollo@gmail.com, paobritos@unrn.edu.ar

Resumen. La información producida por personas, el tipo de contenido expresado en textos, se pueden categorizar en dos grandes grupos: hechos y opiniones. Los hechos son definidos como expresiones objetivas sobre entidades, eventos y sus propiedades, mientras que las opiniones tratan sobre la subjetividad en forma de sentimientos, sensaciones y juicios de valor. Los últimos años han sido testigos de un rápido crecimiento de las plataformas de medios sociales, como Twitter, donde millones de personas comparten sus pensamientos y opiniones sobre varios temas, con el crecimiento explosivo del contenido generado por los usuarios en la Web en los últimos años, el mundo se ha transformado. En este trabajo se presenta una revisión del estado del arte sobre el análisis de sentimientos para la detección de lenguaje figurativo y como los emoticones pueden afectar el análisis de la polaridad de un conjunto de datos extraídos de Twitter.

Palabras clave: Análisis de sentimientos, Emoticones, Ironía, Sarcasmo, Lenguaje figurativo.

1. Introducción

Los últimos años han sido testigos de un rápido crecimiento de las plataformas de medios sociales, como Twitter, donde millones de personas comparten sus pensamientos y opiniones sobre varios temas, con el crecimiento explosivo del contenido generado por los usuarios en la Web en los últimos años, el mundo se ha transformado [1]. Con este crecimiento se generan nuevas posibilidades de explotar la información producida por los consumidores de estos servicios.

Dentro de las opiniones se encuentra el uso extensivo de emojis que han atraído una atención creciente de los investigadores porque los emojis transmiten información semántica y sentimental fructífera para complementar visualmente la información textual que es significativamente útil para comprender las señales emocionales incrustadas en los textos [2].

Se demostró que los usuarios de Twitter utilizan ampliamente los emoticones. En particular, los emoticones que expresan sentimientos positivos, como :) y ;), fueron la mayoría dominante en Twitter [3]. Los emojis pueden resumir y enfatizar la idea original de sus contextos, o expresar semánticas más complejas como la ironía y el sarcasmo al combinarse con contextos de semánticas o sentimientos contradictorios [2].

El objetivo principal del presente trabajo consiste en describir el estado de la cuestión de los diferentes métodos y/o técnicas utilizadas actualmente para Análisis de Sentimientos a través de emoticones o emojis en la red social Twitter.

2. Twitter

Twitter, propiedad y operado por Twitter Inc., es el servicio de microblogging más popular. Los usuarios de Twitter pueden publicar mensajes cortos, llamados tweets, en su perfil de usuario y leer los mensajes de otros en una sola lista agregada en orden cronológico inverso, llamada línea de tiempo. Los tweets son publicaciones basadas en texto limitadas a 280 caracteres [4]. Twitter cuenta con un conjunto de características que lo hacen único entre las redes sociales que se han popularizado con los usuarios de Internet. A continuación, un ejemplo de un tweet en la Fig. 1.



Fig. 1. Ejemplo de un tweet.

Dentro de los tweets se pueden identificar características estilísticas [5]. Algunos ejemplos son emoticones, intensificadores, abreviaturas, términos de argot y signos de puntuación. Una de estas de importancia; la presencia de emoticones. A continuación, en la fig. 2 de los emoticones más comunes dentro de los tweets.

😊 smile	:-) :) :] =)	😬 unsure	:/ :/ :\ :-\
😞 frown	:- (:(:[=(😭 cry	:'(
😛 tongue	:-P :P :-p :p =P	😈 devil	3:) 3:-)
😄 grin	:-D :D =D	👼 angel	O:) O:-)
😮 gasp	:-O :O :-o :o	😘 kiss	:-* :*
😉 wink	;-) ;)	❤️ heart	<3
👓 glasses	8-) 8) B-) B)	😏 kiki	^_^
🕶 sunglasses	8- 8 B- B	😓 squint	^-^
😠 grumpy	>:(>:-(😕 confused	o.O O.o
😡 upset	>:O >:-O >:o >:-o	😬 curly lips	:3

Fig. 2. Ejemplo de emoticones en un tweet.

3. Emoticones en twitter

Los emoticones son símbolos producidos mediante la reutilización creativa y la combinación de caracteres existentes para representar algo nuevo: significar algo ausente en el lenguaje escrito, o algo más efectivamente "dicho" a través del símbolo. Con mayor frecuencia toman la forma de caras expresivas, como :D, >:(, -_- , o (^o^), pero también pueden incluir otros símbolos, como <3. Las señales no verbales han surgido en la comunicación mediada por computadora. Sin embargo, estas señales son conceptualmente diferentes de las señales no verbales en la comunicación cara a cara: las señales como reír y llorar a menudo se conocen como formas involuntarias de expresarse en una comunicación cara a cara, mientras que el uso de sus respectivas los equivalentes “:-)” y “:-(” en la comunicación mediada por computadora son intencionales [6]. Como tal, los emoticones permiten a las personas indicar cambios sutiles de humor, señalar ironía, sarcasmo y chistes, y expresar, enfatizar, o desambiguar su sentimiento, tal vez incluso más que las señales no verbales en la comunicación cara a cara. Por lo tanto, la recolección de información de emoticones es una estrategia viable para mejorar el Análisis de Sentimientos [7].

Los emoticones se pueden clasificar como, emoticones felices (Positivo), emoticones tristes (Negativo), emoticones neutros. A continuación, se muestra en la tabla 1 ejemplos de algunos emoticones, el sentimiento que estos representan y la clasificación que éstos pueden tener.

Tabla 1. Ejemplos de emoticones

Sentimiento	Emotición	Clasificación
Alegría	  	Positivo
Gusto	  	Positivo
Alivio		Positivo
Dolor		Negativo
Disgusto	  	Negativo
Miedo	  	Negativo
Enfado	  	Negativo
Vergüenza	  	Neutro
Sorpresa	 	Neutro

4. Análisis de Sentimientos

Su objetivo es analizar los sentimientos de las personas, actitudes, opiniones, emociones, etc. hacia elementos tales como productos, individuos, temas, organizaciones y servicios [8].

El lenguaje figurativo plantea un desafío especialmente significativo para los sistemas de Análisis de Sentimientos, ya que los enfoques estándar anclados en el afecto definido por el diccionario de palabras y frases individuales a menudo se muestran inadecuados frente a los significados figurativos indirectos. Sería conveniente si dicho lenguaje fuera raro y se limitara a géneros específicos de texto, como la poesía y la literatura. Sin embargo, la realidad es que el lenguaje figurativo es dominante en casi cualquier género de texto, y es especialmente común en los textos de la Web y en las plataformas de redes sociales como Twitter [9].

La ironía y el sarcasmo son dos conceptos interesantes y muy relacionados. Estos dispositivos figurativos nos dan la oportunidad de explorar la interacción entre cognición y lenguaje. En términos generales, la ironía y el sarcasmo son dispositivos de lenguaje figurativo que sirven para lograr diferentes propósitos de comunicación [10]. La ironía y el sarcasmo son formas de lenguaje no literal que a menudo se usan para comunicar lo contrario de lo que se dice literalmente. El sarcasmo es una forma específica de ironía, que se utiliza cuando el objetivo del comentario es una persona [11]. A continuación, se muestran en la tabla 2 Ejemplos extraídos del corpus de la tarea 11 de SemEval 2015. En la columna “Polaridad” se especifica la polaridad con la que se puntuó de media el tweet mostrado como ejemplo.

Tabla 2. Ejemplos de tweets con lenguaje figurativo

Tweet	Polaridad
There is nothing better than Pitbull singing 'playoffs' as Timber plays in the background. #sarcasm	-2.5
Updated my router and it froze. Now I can't access the internet to google a solution. #irony #thankfulformartphones	-4.14
I've had a lot of wake up calls in my day, but I've always been good at hitting the snooze #metaphor #nailedit	0.22

El sarcasmo escrito puede ser difícil de entender correctamente debido a la ausencia de los marcadores habituales disponibles en las conversaciones cara a cara, como el tono de voz y la expresión facial. Por lo tanto, el uso del sarcasmo en una conversación mediada por computadora puede ser arriesgado, ya que el remitente deja abierta la posibilidad de que el receptor interprete el mensaje literalmente. Sin embargo, el uso de emoticones y otros dispositivos podría ayudar a interpretar un comentario sarcástico en este medio [11]. Ahora existe una amplia gama de emoticones que se usan con frecuencia para expresar emociones, aclarar el significado de un mensaje ambiguo, marcar ironía y sarcasmo, o en general compensar la falta de señales no verbales [11].

5. Recursos

Los recursos en esta sección fueron identificados tras una revisión de los trabajos presentados en la edición 2018 (Tabla 3) del SemEval (International Workshop on Semantic Evaluation), el cual son una serie de talleres internacionales de investigación sobre procesamiento del lenguaje natural cuya misión es avanzar el estado actual del arte en análisis semántico y, ayudar a crear conjuntos de datos anotados de alta calidad.

Tabla 3. Participantes SemEval 2018

Grupo	Trabajo	Grupo	Trabajo
#NonDicevoSulSerio	[12]	LDR	[13]
ALANIS	[14]	NEUROSENT	[15]
Binarizer	[16]	NIHRIO	[17]
CTSys	[18]	NLPRL	[19]
ECNU	[20]	NTUA	[21]
EliRF	[22]	PunFields	[23]
HashCount	[24]	Random Decision Syntax Trees	[25]

La recolección de los trabajos presentados en el taller 2018 del SemEval a través de un mapeo sistemático [26] ayudo para la detección de datos como fue el enfoque utilizado para el análisis de sentimientos sea enfoque supervisado, no supervisado o híbrido, los diferentes modelos de inteligencia artificial los cuales pusieron en entrenamiento y a prueba y los recursos léxicos utilizados de los cuales en su mayoría fueron todos de idioma inglés.

En base a los recursos léxicos utilizados ampliamente en las propuestas de los investigadores, luego de su análisis, se observa que pocos recursos de este tipo toman en cuenta el manejo de emoticones entre los elementos que los conforman. Dentro estos recursos analizados en cuanto el manejo de léxicos en español para la detección de sarcasmos, son muy pocos los que utilizan este lenguaje. Si bien hay un análisis y varias herramientas para detección de sarcasmo, en su gran mayoría son para el idioma inglés. Además de esto se sugiere la aplicación de estos modelos y técnicas en la detección de lenguaje figurativo en el idioma español.

6. Conclusiones

En el presente trabajo se ha llevado a cabo una revisión del estado del arte en el dominio del Análisis de Sentimientos para la detección de lenguaje figurativo con emoticones, específicamente en el ambiente de la red social de Twitter.

Como parte del análisis, se han identificado características que hacen único el desafío del Análisis de Sentimientos en Twitter y sobre todo para la detección de lenguaje figurativo, por ejemplo, la abundancia de mensajes cortos, el uso de un lenguaje informal, con slang, en múltiples idiomas, el uso de características estilísticas como emoticones. De los recursos identificados en los trabajos del SemEval y con los resultados de este análisis surgen futuras líneas de investigación como la focalización y el esfuerzo en el área de los recursos para la detección de sarcasmo en el Análisis de Sentimientos con emoticones. La falta, aplicación de estas técnicas para la detección de lenguaje figurativo en conjunto con emoticones en el lenguaje español es otra de las problemáticas que ha sido identificada.

7. Referencias

- [1] B. Liu, "Sentiment Analysis and Subjectivity," *Handbook of Natural Language Processing*, vol. Second Edition, p. 38, 2010.
- [2] Y. Chen, J. Yuan, Q. You, and J. Luo, "Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM," in *2018 ACM Multimedia Conference on Multimedia Conference - MM '18*, Seoul, Republic of Korea, 2018, pp. 117–125, doi: 10.1145/3240508.3240533.
- [3] H. Wang and J. A. Castanon, "Sentiment expression via emoticons on social media," in *2015 IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, USA, Oct. 2015, pp. 2404–2408, doi: 10.1109/BigData.2015.7364034.

- [4] A. R. Twitter Ikuhiro Ihara, "Giving you more characters to express yourself," *Giving you more characters to express yourself*, Aug. 26, 2017. https://blog.twitter.com/official/en_us/topics/product/2017/Giving-you-more-characters-to-express-yourself.html.
- [5] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–41, Nov. 2016, doi: 10.1145/2938640.
- [6] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing - SAC '13*, Coimbra, Portugal, 2013, p. 703, doi: 10.1145/2480362.2480498.
- [7] L. Zhang, S. Pei, L. Deng, Y. Han, J. Zhao, and F. Hong, "Microblog sentiment analysis based on emoticon networks model," in *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service - ICIMCS '13*, Huangshan, China, 2013, p. 134, doi: 10.1145/2499788.2499832.
- [8] H. Naderi, B. Haji Soleimani, S. Mohammad, S. Kiritchenko, and S. Matwin, "DeepMiner at SemEval-2018 Task 1: Emotion Intensity Recognition Using Deep Representation Learning," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 305–312, doi: 10.18653/v1/S18-1045.
- [9] A. Ghosh *et al.*, "SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, 2015, pp. 470–478, doi: 10.18653/v1/S15-2080.
- [10] D. I. H. Farias and P. Rosso, "Irony, Sarcasm, and Sentiment Analysis," in *Sentiment Analysis in Social Networks*, Elsevier, 2017, pp. 113–128.
- [11] R. Filik, A. Ṫurcan, D. Thompson, N. Harvey, H. Davies, and A. Turner, "Sarcasm and emoticons: Comprehension and emotional impact," *Quarterly Journal of Experimental Psychology*, vol. 69, no. 11, pp. 2130–2146, Nov. 2016, doi: 10.1080/17470218.2015.1106566.
- [12] E. W. Pamungkas and V. Patti, "#NonDicevoSulSerio at SemEval-2018 Task 3: Exploiting Emojis and Affective Content for Irony Detection in English Tweets," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 649–654, doi: 10.18653/v1/S18-1106.
- [13] B. Ghanem, F. Rangel, and P. Rosso, "LDR at SemEval-2018 Task 3: A Low Dimensional Text Representation for Irony Detection," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 531–536, doi: 10.18653/v1/S18-1086.
- [14] K. Swanberg, M. Mirza, T. Pedersen, and Z. Wang, "ALANIS at SemEval-2018 Task 3: A Feature Engineering Approach to Irony Detection in English Tweets," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 507–511, doi: 10.18653/v1/S18-1082.
- [15] M. Dragoni, "NEUROSENT-PDI at SemEval-2018 Task 3: Understanding Irony in Social Networks Through a Multi-Domain Sentiment Model," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 512–519, doi: 10.18653/v1/S18-1083.

- [16] N. Nikhil and M. M. Srivastava, “Binarizer at SemEval-2018 Task 3: Parsing dependency and deep learning for irony detection,” *arXiv:1805.01112 [cs]*, May 2018, Accessed: May 16, 2020. [Online]. Available: <http://arxiv.org/abs/1805.01112>.
- [17] T. Vu, D. Q. Nguyen, X.-S. Vu, D. Q. Nguyen, M. Catt, and M. Trenell, “NIHRIO at SemEval-2018 Task 3: A Simple and Accurate Neural Network Model for Irony Detection in Twitter,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 525–530, doi: 10.18653/v1/S18-1085.
- [18] M. Sherif, S. Mamdouh, and W. Ghazi, “CTSys at SemEval-2018 Task 3: Irony in Tweets,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 576–580, doi: 10.18653/v1/S18-1094.
- [19] H. Rangwani, D. Kulshreshtha, and A. Kumar Singh, “NLPRL-IITBHU at SemEval-2018 Task 3: Combining Linguistic Features and Emoji pre-trained CNN for Irony Detection in Tweets,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 638–642, doi: 10.18653/v1/S18-1104.
- [20] Z. Yin, F. Wang, M. Lan, and W. Wang, “ECNU at SemEval-2018 Task 3: Exploration on Irony Detection from Tweets via Machine Learning and Deep Learning Methods,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 600–606, doi: 10.18653/v1/S18-1098.
- [21] C. Baziotis *et al.*, “NTUA-SLP at SemEval-2018 Task 3: Tracking Ironic Tweets using Ensembles of Word and Character Level Attentive RNNs,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 613–621, doi: 10.18653/v1/S18-1100.
- [22] J.-Á. González, L.-F. Hurtado, and F. Pla, “ELiRF-UPV at SemEval-2018 Tasks 1 and 3: Affect and Irony Detection in Tweets,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 565–569, doi: 10.18653/v1/S18-1092.
- [23] E. Mikhalkova, Y. Karyakin, A. Voronov, D. Grigoriev, and A. Leoznov, “PunFields at SemEval-2018 Task 3: Detecting Irony by Tools of Humor Analysis,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 541–545, doi: 10.18653/v1/S18-1088.
- [24] W. I. Cho, W. H. Kang, and N. S. Kim, “HashCount at SemEval-2018 Task 3: Concatenative Featurization of Tweet and Hashtags for Irony Detection,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 546–552, doi: 10.18653/v1/S18-1089.
- [25] A. San, “Random Decision Syntax Trees at SemEval-2018 Task 3: LSTMs and Sentiment Scores for Irony Detection,” in *Proceedings of The 12th International Workshop on Semantic Evaluation*, New Orleans, Louisiana, 2018, pp. 560–564, doi: 10.18653/v1/S18-1091.
- [26] Kitchenham, B., Charters, S.: Guidelines for performing Systematic Literature Reviews in Software Engineering. School of Computer Science and Mathematics, Keele University (2007)