

Selección y ranking de rasgos para caracterizar textos irónicos

Sotolongo-Peña, Anakarla; Arco, Leticia; Bello, Rafael

Published in:
Revista Cubana de Ciencias Informáticas

Publication date:
2020

License:
CC BY

Document Version:
Final published version

[Link to publication](#)

Citation for published version (APA):
Sotolongo-Peña, A., Arco, L., & Bello, R. (2020). Selección y ranking de rasgos para caracterizar textos irónicos. *Revista Cubana de Ciencias Informáticas*, 14(4), 67-84.
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992020000400067

Copyright

No part of this publication may be reproduced or transmitted in any form, without the prior written permission of the author(s) or other rights holders to whom publication rights have been transferred, unless permitted by a license attached to the publication (a Creative Commons license or other), or unless exceptions to copyright law apply.

Take down policy

If you believe that this document infringes your copyright or other rights, please contact openaccess@vub.be, with details of the nature of the infringement. We will investigate the claim and if justified, we will take the appropriate steps.

Tipo de artículo: Artículo original
Temática: Inteligencia Artificial
Recibido: 01/09/2020 | Aceptado: 20/11/2020

Selección y ranking de rasgos para caracterizar textos irónicos

Feature Selection and Ranking to Characterize Ironic Texts

Anakarla Sotolongo-Peña ^{1,3} <https://orcid.org/0000-0003-2642-8745>

Leticia Arco ^{2*} <https://orcid.org/0000-0002-5154-4441>

Rafael Bello ³ <https://orcid.org/0000-0001-5567-2638>

¹ Empresa de Aplicaciones Informática Desoft, Calle Máximo Gómez, No 171 e/ Silvestre Alonso y Frank País, Sancti Spíritus, Cuba, CP 60100. anakarla.sotolongo@desoft.cu

² AI Lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 9, 1050 Brussels, Belgium. larcogar@vub.be

³ Centro de Investigaciones Informáticas, Universidad Central “Marta Abreu” de Las Villas, Carretera a Camajuaní km 5 ½, Santa Clara, Villa Clara, Cuba, CP 54830, Cuba. rbellop@uclv.edu.cu

*Autor para la correspondencia. (larcogar@vub.be)

RESUMEN

Las opiniones textuales imponen grandes retos a las aplicaciones de minería de opinión ya que varios problemas están presentes; entre ellos: la escritura de opiniones de manera irónica o sarcástica. Una de las tendencias que existen para detectar la ironía consiste en la clasificación basada en rasgos. En investigación anterior se propone un conjunto de rasgos que permiten detectar ironía en opiniones textuales; sin embargo,

el cálculo de estos rasgos es costoso computacionalmente. Por lo que en este artículo nos proponemos estudiar dicho conjunto de rasgos con el objetivo de detectar un subconjunto de éste que discrimine entre textos cortos irónicos y no irónicos, sin afectar la eficacia de los clasificadores. El principal resultado de este trabajo consiste en la obtención de un subconjunto de rasgos que logre detectar de manera efectiva la ironía, mediante la aplicación de técnicas de selección y de ranking de rasgos, y la evaluación de varias técnicas de aprendizaje supervisado. El conjunto obtenido de siete rasgos es suficiente para discriminar entre opiniones irónicas y no irónicas, obteniéndose resultados estadísticamente comparables con aquellos obtenidos al utilizar un conjunto mayor y más complejo de rasgos.

Palabras clave: detección de ironía; minería de opinión; selección de rasgos; ranking de rasgos; clasificación.

ABSTRACT

Textual opinions impose great challenges to opinion mining applications since several problems are present; among them: writing opinions ironically or sarcastically. One of the trends that exist to detect irony is the classification based on features. In previous research a set of features that allow detecting irony in textual opinions is proposed; however, the calculation of these features is computationally costly. In this paper, we propose to study this set of features to detect a subset of it that discriminates between ironic and non-ironic short texts, without affecting the effectiveness of the classifiers. The main result of this work consists of obtaining a subset of features that can effectively detect irony, through the application of selection and feature ranking techniques, and the evaluation of several supervised learning techniques. The set obtained from seven features is enough to discriminate between ironic and non-ironic opinions, obtaining statistically comparable results with those obtained by using a larger and more complex set of features.

Keywords: irony detection; opinion mining; feature selection; ranking of features; classification.

Introducción

Trabajar con textos impone grandes retos, por ser datos no estructurados y requerir una etapa de preprocesamiento y representación textual. Las opiniones textuales imponen aún mayores retos a las aplicaciones de minería de opinión ya que varios problemas están presentes como son: el uso de lenguaje informal, la ironía y el sarcasmo, las abreviaturas, los errores ortográficos y tipográficos, la semántica composicional, el nivel cultural y el conocimiento del lenguaje (Amores et al., 2016). Los usuarios al emitir sus criterios en la web utilizan mucho la ironía verbal (Wilson, Sperber, 1992), la cual confunde a las aplicaciones que detectan la polaridad de las opiniones pues conocer su intención de forma automática es muy difícil al no seguir un esquema predeterminado y contener una idea implícita que no está presente en el significado textual. La identificación de opiniones irónicas es sumamente importante en la detección de la polaridad, porque pueden cambiar drásticamente el sentido de lo que se quiere expresar (Hernández, Velásquez, 2015).

La mayoría de los investigadores detectan la ironía basándose en la creación de rasgos (lingüísticos, sintácticos, semánticos y estadísticos) que permitan discriminar entre textos irónicos y no irónicos (Ling, Klinger, 2016 ; Reyes et al., 2013 ; Sotolongo Peña et al., 2018). En (Sotolongo Peña et al., 2018) se propone un conjunto de rasgos que permiten detectar ironía en opiniones textuales con gran eficacia; sin embargo, el cálculo de estos rasgos es costoso computacionalmente. Por lo que en este trabajo nos proponemos estudiar dicho conjunto de rasgos con el objetivo de detectar un subconjunto de éste que discrimine entre textos cortos irónicos y no irónicos, sin afectar la eficacia de los clasificadores. Para ello se aplicarán técnicas de selección y de ranking de rasgos, y la evaluación de varios métodos de aprendizaje supervisado.

Métodos o Metodología Computacional

Descripción de los principales rasgos para caracterizar la ironía

Esta investigación toma como punto de partida los 18 rasgos utilizados en (Sotolongo Peña et al., 2018) para caracterizar la ironía, de los cuales seis fueron seleccionados de la literatura y 12 propuestos en dicha investigación, basándose en ideas planteadas por otros autores o el estudio de ejemplos de ironía que permitió identificar las características distintivas de las oraciones irónicas. La descripción de cada uno de estos rasgos se muestra en la Tabla 1.

Varios autores plantean que las opiniones irónicas deben tener un alto grado de contraste mientras que las no irónicas deben tener un valor casi nulo (Ling, Klinger, 2016 ; Barbieri, Saggion, 2014 ; Rajadesingan et al., 2015 ; Rella et al., 2015), siguiendo este criterio surgió WordPolarityContrast. Por otro lado, los adverbios (Reyes et al., 2013) y signos de puntuación son usados para expresar oposición (Sotolongo Peña et al., 2018), así lo cuantifica AntecedentConsequentPolarityContrast. PolarityStandardDeviation refleja la existencia de palabras con polaridades extremas en la opinión teniendo en cuenta la desviación estándar de la polaridad de las palabras (Sotolongo Peña et al., 2018 ; Rella et al., 2015). Por otro lado, varias personas expresan ironía utilizando frases entre comillas (Ling, Klinger, 2016), MeanPolarityPhrase reconoce frases entre comillas sin tener en cuenta una sintaxis determinada. Los rasgos SynoLowerMean, SynoLowerGap, SynoGreaterMean y SynoGraterGap se basan en el supuesto de que el uso de palabras que se utilizan con poca frecuencia, puede ser consecuencia de una intención de énfasis o de tratar de llamar la atención por parte del hablante, por tanto, puede indicar ironía con alta probabilidad (Barbieri, Saggion, 2014 ; Sotolongo Peña et al., 2018).

Las personas al escribir irónicamente pueden usar caracteres repetidos para hacer énfasis en una determinada idea (Rajadesingan et al., 2015), ese es el supuesto que sigue el rasgo RepetitionRate. El rasgo PresentPastPolarityContrast mide la oposición a través del tiempo, que es otro de los elementos que puede denotar ironía (Reyes et al., 2013 ; Sotolongo Peña et al., 2018). Muchas personas al expresar ironía utilizan emoticonos para hacer énfasis en la idea que se quiere transmitir, EmoticonTextPolarityContrast captura

esta característica (Sotolongo Peña et al., 2018). El desbalance contextual, reflejado en el rasgo DomainRate, es otro de los elementos que caracteriza la ironía, con la intención de buscar lo inesperado teniendo en cuenta el grado de similitud semántica del texto (Sotolongo Peña et al., 2018). La ambigüedad puede ser utilizada en un texto para expresar ironía y varios autores han formulado atributos para capturarla (Rella et al., 2015 ; Barbieri, Saggion, 2014 ; Sotolongo Peña et al., 2018). En el rasgo Ambiguity cuantifica la proporción de palabras con alto y con bajo grado de ambigüedad. Por otro lado, SkipgramPolarityRate detecta si existe contradicción en el texto a través de los Polarity Skipgrams^a. Se han definido varios rasgos dirigidos a identificar interjecciones como indicador de ironía (Ling, Klinger, 2016). FalsePositivePhraseInterjection identifica la presencia de interjecciones teniendo en cuenta la estructura de la frase que lo acompaña, mientras que InterjectionRate calcula la razón entre la cantidad de interjecciones y el total de palabras del texto (Sotolongo Peña et al., 2018).

Tabla 1 – Descripción de los atributos propuestos en (Sotolongo Peña et al., 2018).

No.	Nombre del atributo	Descripción del atributo
R1	<i>WordPolarityContrast</i>	Mide el contraste teniendo en cuenta la polaridad y la distancia entre los elementos de mayor contraste.
R2	<i>AntecedentConsequentPolarityContrast</i>	Identifica dos partes de la oración que surgen al dividirla por un delimitador (adverbio o signos de puntuación) y pondera la importancia de la diferencia de la cantidad de palabras y de su polaridad en ambas partes de la oración.
R3	<i>MeanPolarityPhrase</i>	Calcula el promedio de la polaridad de las palabras que se encuentran en frases entre comillas.
R4	<i>PolarityStandardDeviation</i>	Devuelve un valor alto cuando existen palabras con polaridades extremas en la opinión, en caso contrario, valores bajos.
R5	<i>SynoLowerMean</i>	Capturan el uso de sinónimos comunes y raros, basándose en un corpus de referencia donde se busca la frecuencia de cada palabra que aparece en el texto y sus sinónimos.
R6	<i>SynoLowerGap</i>	
R7	<i>SynoGreaterMean</i>	
R8	<i>SynoGraterGap</i>	
R9	<i>RepetitionRate</i>	Cuantifica la cantidad de caracteres repetidos y la proporción de éstos con respecto al texto para dar la idea de en qué medida se está exagerando.
R10	<i>PresentPastPolarityContrast</i>	Cuantifica las polaridades de sentimientos entre las dos partes que componen el texto delimitada por los verbos conjugados en pasado y presente.
R11	<i>EmoticonTextPolarityContrast</i>	Calcula el contraste entre los emoticonos y las palabras del texto.
R12	<i>DomainRate</i>	Calcula la razón de la cantidad de dominios presentes en el texto y la cantidad de palabras que pertenecen a estos dominios.
R13	<i>Ambiguity</i>	Evalúa estadígrafos sobre la cantidad de sentidos de cada palabra.

R14	<i>SkipgramPolarityRate</i>	Calcula la razón entre la cantidad de Polarity Skipgrams donde hay oposición y la cantidad de Polarity Skipgrams válidos.
R15	<i>FalsePositivePhraseInterjection</i>	Identifica la interjección acompañada de una frase aparentemente positiva formada por un sustantivo y un adjetivo positivo entre comillas o seguido de dos o más signos de exclamación y/o interrogación.
R16	<i>InterjectionRate</i>	Calcula la razón entre la cantidad de interjecciones y el total de palabras del texto.
R17	<i>PhrasePolarity</i>	Calcula la diferencia absoluta de la mayor polaridad que expresa alguna palabra de la frase y el promedio de las polaridades de las palabras que conforman la frase.
R18	<i>CapitalLetterTextPolarityContrast</i>	Calcula la diferencia entre el promedio de polaridad de las palabras en mayúsculas y la polaridad mínima del texto.

Una forma de expresar ironía es mediante frases positivas o negativas entre comillas o seguidas de signos de exclamación y/o interrogación (Ling, Klinger, 2016). *PhrasePolarity* refleja la polaridad dominante en las frases con esa estructura (Sotolongo Peña et al., 2018). Las palabras en mayúsculas generalmente son usadas en la ironía para decir lo opuesto a lo que se quiere expresar. *CapitalLetterTextPolarityContrast* captura el contraste entre las palabras en mayúsculas y al resto del texto, a diferencia de otros rasgos (Reyes et al., 2013 ; Rajadesingan et al., 2015 ; Ling, Klinger, 2016) que solo tienen en cuenta la cantidad de palabras o letras en mayúsculas.

Experimentos realizados en (Sotolongo Peña et al., 2018) evidencian que estos 18 rasgos en su conjunto logran distinguir entre textos irónicos y no irónicos; sin embargo, se desconoce el poder discriminante de cada uno de ellos y si es necesario calcularlos todos para clasificar los textos en irónicos o no. Debido a la complejidad computacional del cálculo de algunos de estos rasgos, es deseable identificar un subconjunto de ellos que logre igualmente caracterizar los textos irónicos. Además, resulta interesante comprobar si realmente los supuestos indicadores de ironía logran efectivamente caracterizar tales textos. De ahí la necesidad de aplicar técnicas de selección y ranking de atributos, como se muestra en la sección siguiente.

Análisis de los rasgos que caracterizan la ironía

Para estudiar los 18 rasgos partimos de la colección de textos usada en (Ling, Klinger, 2016). Debido a que algunos tweets no estaban disponibles y no se incluyeron aquellos textos que poseían hipervínculos, la colección se redujo a un total de 54623 tweets irónicos o sarcásticos y 5027 regulares. Como esta colección

es desbalanceada, a partir de ésta se formaron 10 colecciones balanceadas donde se incluyeron 5000 textos irónicos y los 5000 regulares. Cada una de estas colecciones ha sido representada utilizando los 18 rasgos en estudio y utilizada en todos los experimentos que se presentan en este artículo. Además, es importante señalar que los algoritmos aplicados se encuentran disponibles en Weka ^b, por lo que en todo el artículo se utilizarán los identificadores de Weka para cada método.

Selección de los rasgos que mejor discriminan entre textos irónicos y no irónicos

En esta investigación es de interés identificar el subconjunto de rasgos que mejor discrimina entre clases, partiendo de los 18 rasgos descritos anteriormente. Primeramente, se aplicaron los algoritmos que construyen árboles de decisión, ya que en ese proceso tienen embebida la selección de atributos. Los experimentos realizados evidenciaron que los mejores valores de las medidas de calidad fueron obtenidos con el algoritmo J48. Este algoritmo se aplicó a todas las colecciones, resultando ganadores los rasgos que se presentan en la Tabla 2. El rasgo de mayor relevancia para todas las colecciones fue CapitalLetterTextPolarityContrast (Sotolongo Peña et al., 2018).

Tabla 2 - Listado de atributos con alto valor de ganancia al aplicar J48.

Atributos
<i>CapitalLetterTextPolarityContrast</i>
<i>WordPolarityContrast</i>
<i>PolarityStandardDeviation</i>
<i>AntecedentConsequentPolarityContrast</i>
<i>SynoLowerMean</i>
<i>SynoGreaterMean</i>
<i>PresentPastPolarityContrast</i>
<i>SynoLowerGap</i>
<i>EmoticonTextPolarityContrast</i>

<i>Ambiguity</i>
<i>DomainRate</i>

También se aplicaron métodos de evaluación de subconjuntos de atributos y los de un solo atributo para obtener los rasgos con mayor poder discriminativo entre opiniones irónicas y no irónicas. Los evaluadores de subconjuntos toman un subconjunto de atributos y devuelven una medida numérica que guía la búsqueda. CfsSubsetEval (Cfs) (Hall, 1998) evalúa la capacidad predictiva de cada atributo individualmente y el grado de redundancia entre ellos, prefiriendo conjuntos de atributos altamente correlacionados con la clase, pero con baja correlación entre ellos. ConsistencySubsetEval (Consist) (Liu, Setiono, 1996) evalúa conjuntos de atributos por el grado de consistencia en los valores de la clase cuando las instancias de entrenamiento se proyectan en el conjunto. Este evaluador generalmente se usa junto con una búsqueda aleatoria o exhaustiva que busca el subconjunto más pequeño con una consistencia que es la misma que la del conjunto completo de atributos. Mientras que los evaluadores de subconjuntos previamente mencionados son filtros de selección de atributos, ClassifierSubsetEval (Classifier) y WrapperSubsetEval (Wrapper) (Kohavi, John, 1997), son métodos de envoltura y ambos usan un clasificador para evaluar conjuntos de atributos, distinguiéndose WrapperSubsetEval por el empleo de la validación cruzada para estimar la precisión del esquema de aprendizaje para cada conjunto. Al aplicar estos métodos en nuestra experimentación, se seleccionó para cada caso el método de búsqueda sugerido en la literatura (Witten et al., 2011).

Tabla 3 – Atributos ganadores al aplicar evaluadores de subconjuntos de atributos.

	Cfs	Consist	Classifier	Wrapper		Cfs	Consist	Classifier	Wrapper
R1		X	X	X	R10	X	X		X
R2	X	X	X	X	R11	X	X	X	X
R3	X	X		X	R12		X		X
R4		X	X	X	R13		X		X
R5		X		X	R14		X	X	X
R6		X		X	R15				X

R7		X			R16	X	X		X
R8	X	X		X	R17				X
R9				X	R18	X	X	X	X

Al analizar la Tabla 3, resultante de aplicar los evaluadores de subconjuntos de atributos, los rasgos que tuvieron mayor ganancia; i.e., aquellos que resultaron ganadores en al menos tres métodos, fueron *AntecedentConsequentPolarityContrast* (R2), *EmoticonTextPolarityContrast* (R11), *CapitalLetterTextPolarityContrast* (R18), *WordPolarityContrast* (R1), *MeanPolarityPhrase* (R3), *PolarityStandardDeviation* (R4), *SynoGraterGap* (R8), *PresentPastPolarityContrast* (R10), *SkipgramPolarityRate* (R14) y *InterjectionRate* (R16).

Ranking de los atributos

Para identificar el subconjunto de rasgos que mejor discrimina entre opiniones irónicas y no irónicas no solo se tuvieron en cuenta los resultados de los métodos de selección de rasgos. Además, se aplicaron los evaluadores de atributos que realizan ranking de los rasgos, para identificar aquellos que alcanzan mejores posiciones. Estos métodos, descritos en la Tabla 4, ofrecieron un ranking de los rasgos por cada colección analizada, de ahí que se hizo necesario aplicar métodos de agregación de ranking que permitan obtener un ranking consenso para definitivamente identificar los rasgos con mejores puntuaciones.

Tabla 4 - Métodos aplicados en este artículo para realizar el ranking de los atributos (Witten et al., 2011).

Nombre	Descripción
ReliefFAttributeEval	Evalúa los atributos basándose en las instancias (Kira, Rendell, 1992)
InfoGainAttributeEval	Evalúa los atributos midiendo su ganancia de información con respecto a la clase (Lorenzo Navarro, 2002)
ChiSquaredAttributeEval	Evalúa los atributos mediante el cálculo de la estadística chi-cuadrado con respecto a la clase
Gain-RatioAttributeEval	Evalúa los atributos midiendo su relación de ganancia con respecto a la clase

SymmetricalUncertAttributeEval	Evalúa los atributos midiendo su incertidumbre simétrica con respecto a la clase
OneRAttributeEval	Usa la medida de precisión simple adoptada por el clasificador OneR (Holte, 1993)

Se utilizó el método *RankAggreg* del paquete *RankAggreg* de R^2 para agregar los rankings obtenidos de los 18 rasgos en estudio. Se obtuvo un ranking consenso por cada colección, considerando los resultados de cada método de ranking aplicado, y también, buscamos un ranking consenso por cada método de ranking aplicado a las 10 colecciones textuales utilizadas en la experimentación. En la Tabla 5 se muestra el ranking consenso obtenido por colecciones y por métodos de ranking. Estos resultados fueron obtenidos aplicando un algoritmo genético, con la distancia de Spearman y con *score* 26.75 y 25.09 al buscar el consenso a partir de los rankings por colecciones y a partir de los métodos de ranking, respectivamente.

Tabla 5 – Ranking consenso general obtenido a partir de los rankings por colecciones y por métodos de ranking.

Colecciones	1	2	3	4	5	6	7	8	9
	R18	R1	R4	R14	R2	R13	R11	R6	R8
	10	11	12	13	14	15	16	17	18
	R16	R10	R5	R7	R3	R15	R17	R9	R12
Métodos de ranking	1	2	3	4	5	6	7	8	9
	R1	R18	R14	R4	R2	R11	R8	R6	R13
	10	11	12	13	14	15	16	17	18
	R7	R5	R16	R10	R17	R15	R3	R9	R12

Al observar la Tabla 3 y la Tabla 5, los siete mejores rasgos tanto en selección como en rankings consenso fueron R1, R18, R14, R4, R2, R11, R8. De ellos, cinco fueron creados en (Sotolongo Peña et al., 2018) (R1, R18, R14, R2 y R11) y tratan de identificar el contraste en el texto al usar elementos como palabras con polaridades extremas, delimitadores, emoticonos y palabras en mayúscula; y dos fueron seleccionados de la literatura (R4 y R8) y miden el uso de sinónimos comunes y la desviación estándar de las polaridades de las palabras. A estos se les puede añadir R16 que ganó en selección y en uno de los rankings; R13 y R6 que

ganaron en los dos rankings y no en selección; y R3, R10 y R7 que ganaron en ranking o en selección. De estos últimos tres rasgos (R3, R10 y R7), dos fueron creados en (Sotolongo Peña et al., 2018) y capturan la polaridad de las frases entre comillas y el contraste de dos partes de la oración que están en tiempos distintos, y el otro fue tomado de la literatura y detecta el uso de sinónimos comunes y raros; y de los otros tres (R16, R13 y R6), el primero fue creado en (Sotolongo Peña et al., 2018) y mide el uso de interjecciones, y los otros dos fueron seleccionados de la literatura y miden el uso de la ambigüedad y de sinónimos comunes y raros.

Resultados y discusión

En esta sección se analizará el comportamiento de los cuatro clasificadores más utilizados en la clasificación de ironía describiendo las colecciones de tweets por el conjunto de los siete mejores rasgos, los 13 mejores y el total de 18 rasgos. El objetivo de la experimentación consiste en explorar si solo con los siete mejores rasgos es suficiente para discriminar entre tweets irónicos y no irónicos. Por tanto, se estudia el desempeño de los clasificadores para los tres conjuntos de rasgos definidos.

Los clasificadores aplicados son Naïve Bayes (NB) (John, Langley, 1995), la máquina de vectores de soporte entrenada con el algoritmo de optimización secuencial mínima (Sequential Minimal Optimization; SMO) (Platt, 1998), J48 correspondiente al clásico algoritmo C4.5 (Quinlan, 1993) y Random Forest (RF) (Ho, 1995). En SMO se utilizó un kernel polinomial y una tolerancia de 0.001. En J48 se definió un factor de confianza de 0.25 para la poda pues valores menores incurren en más podas y se establecieron dos instancias como mínimo por hoja. En Random Forest no se limitó la profundidad de los árboles.

Estos clasificadores fueron aplicados directamente a los subconjuntos de rasgos para cada colección. A cada resultado de clasificación se aplicaron las medidas Exactitud (E), Precisión (P), Exhaustividad (R) y la medida-F (F) (Hossin, Sulaiman, 2015), realizando una validación cruzada considerando 10 particiones para indicar conjuntos de entrenamiento y prueba.

El diseño experimental de la validación realizada para determinar si existen diferencias significativas entre los subconjuntos de rasgos al aplicar los clasificadores a las diferentes colecciones para detectar la ironía consta de los siguientes pasos:

1. Creación de una tabla por cada medida de calidad, que incluye los valores de ésta al aplicar los clasificadores a las colecciones representadas con los subconjuntos de rasgos propuestos.
2. Aplicar el test de Friedman para identificar si existen diferencias significativas entre los subconjuntos de rasgos considerando una medida de calidad dada y un valor de significación $\alpha=0.05$. Este test genera un ranking de los algoritmos.
3. Aplicar un test Post-Hoc para identificar diferencias significativas entre pares de subconjuntos de rasgos. Para ello se utilizó el test de Holm con umbral de significación $\alpha=0.05$.
4. Graficar los resultados del ranking de los subconjuntos de rasgos y la presencia o no de diferencias significativas por cada medida de calidad considerada en el estudio, donde los resultados conectados no son significativamente diferentes.

Al analizar la Fig. 1 se observa que no existen diferencias significativas en los valores de exactitud obtenidos al aplicar los algoritmos NB, RF, J48 y SMO a los conjuntos utilizando siete, 13 y 18 rasgos. Se evidencia que el conjunto de siete rasgos tiene igual poder discriminante entre opiniones irónicas y no irónicas. Al analizar la Fig. 2 se observa que los mejores valores de precisión al aplicar SMO se obtuvieron sobre los conjuntos de 18 y 13 rasgos, existiendo diferencias significativas entre estos resultados y los valores de precisión obtenidos al caracterizar los textos irónicos utilizando siete rasgos. No obstante, no existen diferencias significativas en cuanto a la precisión obtenida por los algoritmos NB, RF y J48 al caracterizar los datos con los tres conjuntos de rasgos en análisis.

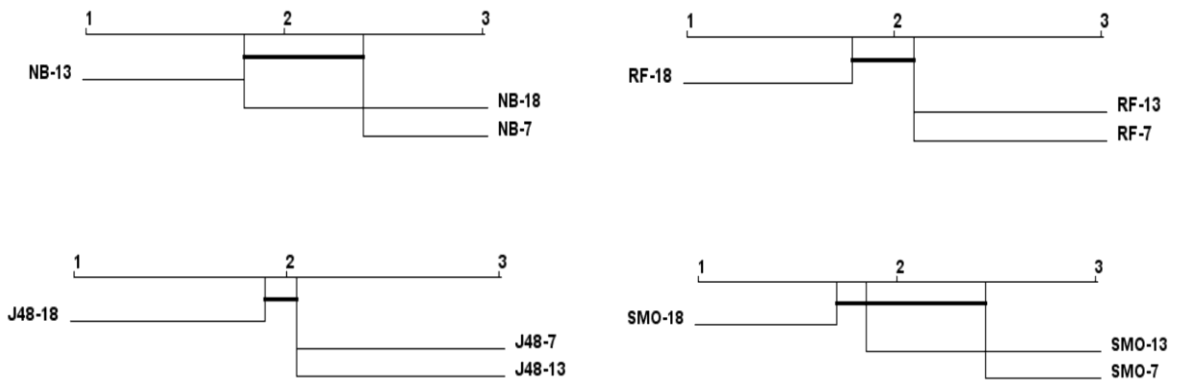


Fig. 1 – Ranking de los algoritmos según los valores de exactitud.

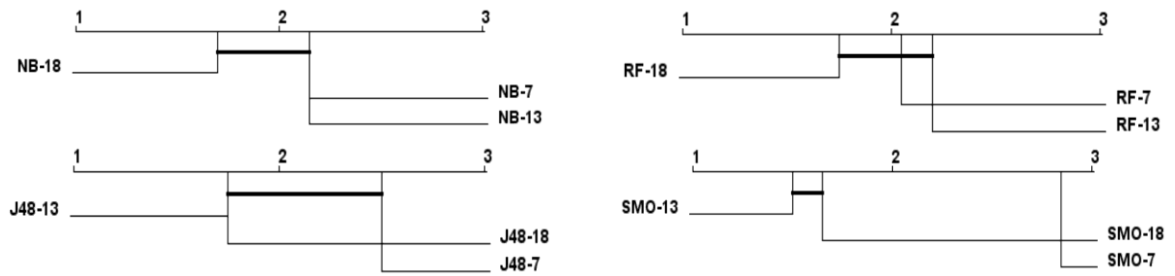


Fig. 2 – Ranking de los algoritmos según los valores de precisión.

Los mejores valores de exhaustividad se obtuvieron al aplicar los algoritmos RF y J48 al conjunto de siete rasgos, como se muestra en la Fig. 3. No así, al aplicar los algoritmos NB y SMO, donde los valores de exhaustividad para conjuntos de siete rasgos fueron más bajos. No obstante, no existieron diferencias significativas en los resultados obtenidos al aplicar estos clasificadores sobre los conjuntos de siete, 13 y 18 rasgos, evidenciándose el poder discriminatorio del conjunto de siete rasgos. Finalmente, la Fig. 4 muestra que no existen diferencias significativas en los valores obtenidos de la medida-F al aplicar los algoritmos a los conjuntos de rasgos en estudio.

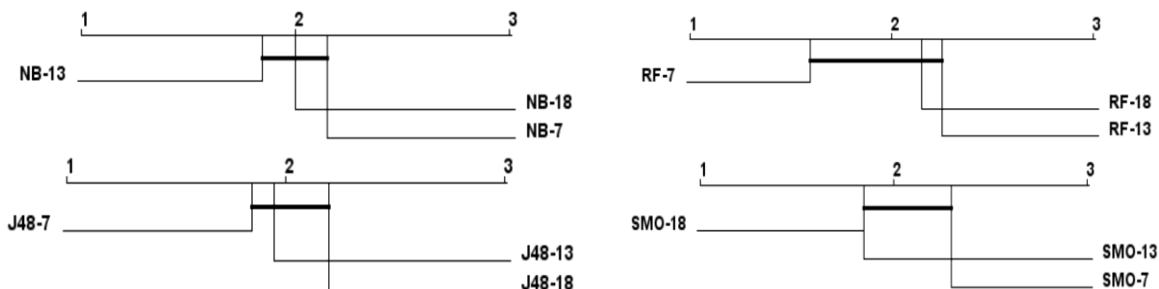


Fig. 3 – Ranking de los algoritmos según los valores de exhaustividad.

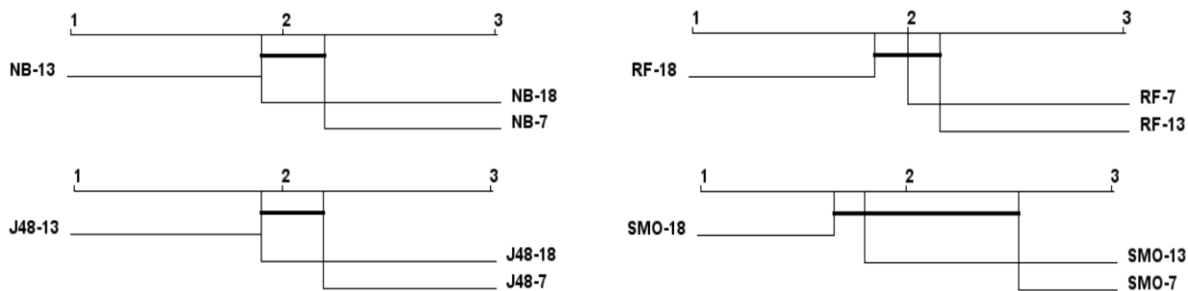


Fig. 4 – Ranking de los algoritmos según los valores de medida-F.

La Fig. 5; **Error! No se encuentra el origen de la referencia.** evidencia que prácticamente no hay diferencias entre los valores promedios de las medidas de calidad exactitud, precisión, exhaustividad y medida-F obtenidos al aplicar los clasificadores NB, RF, J48 y SMO a los conjuntos de siete y 18 rasgos.

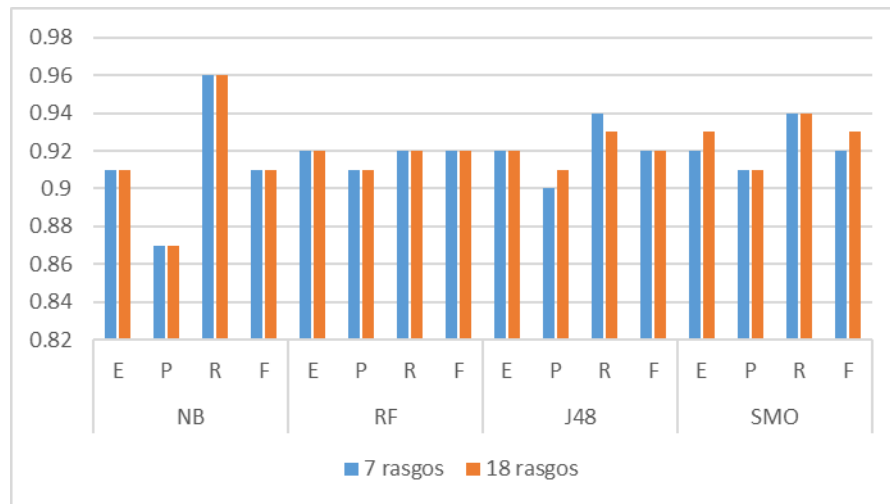


Fig. 5 – Valores promedios de las medidas de calidad por los clasificadores aplicados al conjunto de siete y de 18 rasgos.

Los resultados experimentales nos permiten afirmar que en la mayoría de los casos son suficientes los siete rasgos seleccionados para discriminar entre opiniones irónicas y no irónicas. Los resultados obtenidos evidencian el poder discriminante de los rasgos propuestos en (Sotolongo Peña et al., 2018) (R1, R18, R14, R2 y R11), así como de aquellos resultantes del análisis de la literatura (R4 y R8). La mayoría de estos rasgos tienen en cuenta elementos semánticos (R1, R2, R4, R11, R14 y R18) y calculan proporciones, por tanto, se basan en algunas técnicas estadísticas (R4, R8 y R18), y en menor medida tienen en cuenta elementos lingüísticos (R8 y R14) y sintácticos (R2 y R14).

Conclusiones

Al aplicar técnicas de selección y ranking el conjunto de rasgos originalmente diseñado en (Sotolongo Peña et al., 2018) se redujo a siete rasgos con los cuales los valores promedios de exactitud, precisión, exhaustividad y medida-F fueron de 0.92, 0.90, 0.94 y 0.92 respectivamente, siendo éstos similares a los obtenidos con el conjunto de 18 rasgos.

El subconjunto de rasgos ganador se basa en la identificación del contraste en el texto al usar elementos como palabras con polaridades extremas, delimitadores, emoticonos y palabras en mayúscula, el uso de sinónimos comunes y la desviación estándar de las polaridades de las palabras; reafirmando de esta forma los principales supuestos acerca de los elementos que distinguen a los textos irónicos.

A pesar de que ha sido interesante la reducción de rasgos a calcular, aún quedan por explorar otras tareas de preprocesamiento como la desambiguación del sentido de las palabras y el tratamiento de las jergas; así como la creación de rasgos que consideren el escenario emocional, lo cual podría conllevar describir mejor las colecciones y consecuentemente obtener mejor desempeño de los clasificadores.

Referencias

- Amores, m., arco, l. Y barrera, A., 2016. Efectos de la negación, modificadores, jergas, abreviaturas y emoticonos en el análisis de sentimiento. In : *Proceedings of the 2nd International Workshop on Semantic Web (IWSW)*. La Habana: CEUR. 2016.
- Barbieri, f. Y saggioN, H., 2014. Automatic detection of irony and humor in Twitter. In: *Proceedings of the Fifth International Conference on Computational Creativity* [en línea]. Ljubljana, Slovenia: Josef Stefan Institute, Ljubljana, Slovenia. 2014. Disponible en : <http://computationalcreativity.net/iccc2014/wp-content/uploads/2014/06//9.2{ }Barbieri.pdf>.
- Hall, M A, 1998. *Correlation-based feature subset selection for machine learning*. Hamilton, New Zealand: University of Waikato.
- Hernández, V. A. y Velásquez, J. D., 2015. *Identificación de la presencia de ironía en el texto generado por usuarios de Twitter utilizando técnicas de opinion mining y machine learning*. S.l.: Universidad de Chile.
- Ho, T. K., 1995. Random Decision Forests. In: *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)* [en línea]. Washington, DC, USA: IEEE Computer Society. 1995. p. 278-. Disponible en : <http://dl.acm.org/citation.cfm?id=844379.844681>.
- Holte, R. C., 1993. Very simple classification rules perform well on most commonly used datasets. In: *Machine Learning*. 1993. Vol. 11, p. 63-91.

- Hossin, M. y Sulaiman, M. N., 2015. A review on evaluation metrics for data classification evaluations. In: *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. 2015. Vol. 5, n° 2. DOI 10.5121/ijdkp.2015.5201.
- John, G. H. y Langley, P., 1995. Estimating continuous distributions in bayesian classifiers. In: *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann. 1995. p. 338-345.
- Kira, K. y Rendell, L. A., 1992. A practical approach to feature selection. In: SLEEMAN, D. H. y Edwards, P. (eds.), *Ninth International Workshop on Machine Learning*. S.l.: Morgan Kaufmann. 1992. p. 249-256.
- Kohavi, R. y John, G. H., 1997. Wrappers for feature subset selection. In: *Artificial Intelligence*. 1997. Vol. 97, n° 1-2, p. 273-324.
- Ling, J. y Klinger, R., 2016. An empirical, quantitative analysis of the differences between sarcasm and irony. In : Sack, H., Rizzo, G., Steinmetz, N., Mladeníć, D., Auer, S. y Lange, C. (eds.), *The Semantic Web*. S.l.: Springer International Publishing. p. 203-216. ISBN 978-3-319-47602-5.
- Liu, H. y Setiono, R., 1996. A probabilistic approach to feature selection - A filter solution. In: *13th International Conference on Machine Learning*. S.l.: s.n. 1996. p. 319-327.
- Lorenzo Navarro, J. J., 2002. *Selección de atributos en aprendizaje automático basado en la teoría de la información*. S.l.: Universidad de Las Palmas de Gran Canaria.
- Platt, J., 1998. Fast training of Support Vector Machines using Sequential Minimal Optimization. In: Schoelkopf, B., Burges, C. y Smola, A. (eds.), *Advances in Kernel Methods - Support Vector Learning* [en línea]. S.l.: MIT Press. Disponible en: <http://research.microsoft.com/%5C~jplatt/smo.html>.
- Quinlan, R., 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers.
- Rajadesingan, A., Zafarani, R. y Liu, H., 2015. Sarcasm detection on Twitter: A behavioral modeling approach. In: *8th ACM International Conference on Web Search and Data Mining, WSDM 2015*. Shanghai, China: Association for Computing Machinery, Inc. 2015. p. 97-106.
- Rella, J., Saggion, H. y Barbieri, F., 2015. *TwIrony: Identificación de la ironía en tweets en Catalán*. S.l.: Universitat Pompeu Fabra.
- Reyes, A., Rosso, P. y Veale, T., 2013. A multidimensional approach for detecting irony in Twitter. In: *Language Resources and Evaluation*. 2013. Vol. 47, n° 1, p. 239-268. DOI 10.1007/s10579-012-9196-x.
- Sotolongo Peña, A., ARCO, L. y Rodríguez Dosina, A., 2018. Detección de ironía en textos cortos enfocada

a la minería de opinión. In: *IV Conferencia Internacional en Ciencias Computacionales e Informáticas (CICCI' 2018)*. La Habana, Cuba: s.n. 2018.

Wilson, D. y Sperber, D., 1992. On verbal irony. In: *Lingua*. 1992. Vol. 87, n° 1, p. 53-76. DOI 10.1016/0024-3841(92)90025-E.

Witten, I. H., Frank, E. y HALL, M. A., 2011. *Data mining: practical machine learning tools and techniques*. Third Edit. S.l.: Morgan Kaufmann. ISBN 9780123748560.

Conflicto de interés

Los autores declaran no tener ningún conflicto de intereses y autorizan la distribución de este artículo.

Contribuciones de los autores

Leticia Arco: Contribuyó en Conceptualización, Metodología, Supervisión, Redacción – revisión y edición.

Rafael Bello: Contribuyó en Conceptualización, Supervisión.

Anakarla Sotolongo-Peña: Contribuyó en Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación, Visualización, Redacción – borrador original.

Financiación

Este artículo forma parte de la investigación realizada por la Lic. Anakarla Sotolongo Peña para alcanzar el título de Máster en Ciencia de la Computación de la Universidad Central “Marta Abreu” de Las Villas, bajo la asesoría de la Dra. Leticia Arco y el Dr. Rafael Bello.

^a <http://www.romanklinger.de/ironysarcasm/>

^b <https://cran.r-project.org/>