Procesamiento del Lenguaje Natural, Revista nº 58, marzo de 2017, pp. 159-162 recibido 01-12-2016 revisado 11-01-2017 aceptado 06-02-2017

Detección de Opinion Spam usando PU-Learning

Opinion Spam detection using PU-Learning

Donato Hernández Fusilier
Departamento de Sistemas Informáticos y Computación
Universitat Politècnica de València
Camino de Vera s/n 46022, Valencia, España
doherfu@doctor.upv.es

Resumen: Tesis doctoral realizada por Donato Hernández Fusilier en la Universitat Politècnica de València, dirigida por los Doctores Paolo Rosso (Universitat Politècnica de València, España, Manuel Montes-y-Gómez (Instituto Nacional de Astrofísica, Óptica y Electrónica, México) y Rafael Guzmán (Universidad de Guanajuato, México). La defensa se efectuó el 20 de enero de 2016 en Valencia. El tribunal estuvo conformado por la Dra. Raquel Martinez Unanue de la Universidad Nacional de Educación a Distancia, Madrid como vocal, por el Dr. Carlos David Martinez Hinajeros de la Universitat Politècnica de València como secretario y por el Dr. Rafael Berlanga LLavori de la Universitat Jaume I, Castelló como presidente. La tesis obtuvo una calificación de Sobresaliente.
Palabras clave: Análisis de Opiniones, aprendizaje supervisado, aprendizaje no supervisado, clasificadores, PU-Learning.

Abstract: Doctoral thesis written by Donato Hernández Fusilier at the Universitat Politècnica of València, directed by Ph.D. Paolo Rosso (Universitat Politècnica of València, Spain), Ph.D. Manuel Montes-y-Gómez (National Institute of Astrophisics, Optics and Electronics, México) and Ph.D. Rafael Guzmán (University of Guanajuato, México). The defense took place on January 20, 2016 in Valencia. The doctoral committee was integrated by the following doctors: Ph.D. Raquel Martinez Unanue of National Distance Learning University, Madrid as panel member, Ph.D. Carlos David Martinez Hinajeros de la Universitat Politècnica of València as secretary and by Ph.D. Rafael Berlanga LLavori of the Universitat Jaume I, Castelló as president. The thesis was graded as Outstanding.
Keywords: Analysis of Opinions, supervised learning, unsupervised learning, classifiers, PU-Learning.

1 Introducción

Con el uso general de las tecnologías de información es cada vez más común que los usuarios de servicios y los consumidores de productos escriban sus opiniones a favor o en contra de los productos o servicios que adquirieron. Estas referencias escritas comúnmente en foros, blogs y en general en las redes sociales, sirven de ayuda a otros consumidores que desean adquirir algunos productos o servicios similares. También sirven a los fabricantes o prestadores de servicios para identificar nuevas áreas de oportunidad por parte de los consumidores y les permite saber no solo la opinión sobre los mismos, sino además ver sus usos, costumbres, satisfacción, etcétera. Los consumidores utilizan las opiniones para recibir información sobre los productos, tales como calidad y utilidad, también son utilizadas para proporcionar datos sobre su propia experiencia con el producto a otros consumidores. Por otro lado, los fabricantes utilizan estos comentarios positivos o negativos, para identificar características que son importantes para los consumidores. Estas características son entonces incluidas en la comercialización y desarrollo de nuevos productos o servicios.

ISSN 1135-5948

Con este tipo de opiniones se presenta el problema del Opinión Spam que, en otras palabras, son aquellas opiniones falsas, escritas deliberadamente para promover o desacreditar un producto o servicio. Son opiniones escritas por personas que no han adquirido producto o servicio alguno, pero que fueron contratados para escribir opiniones engañosas. Estas opiniones falsas hacen creer a los posibles usuarios, que el producto o servicio es bueno o, según sea la causa para la cual fueron inducidos.

2 Objetivos

Desarrollar un método semi-supervisado, basado en el método de PU-Learning, para la detección de opiniones falsas, que considere tanto las características temáticas como estilísticas y que además sea robusto, adecuado y efectivo en escenarios reales que presentan una escasez de datos etiquetados.

Encontrar los atributos que consideren información tanto del contenido como del estilo de escritura de las opiniones.

Establecer el efecto de la polaridad de las opiniones, mediante el entrenamiento con una polaridad y la prueba con otra polaridad.

Probar el método propuesto en ambientes de dominios cruzados, donde se entrena con opiniones de un dominio y se prueba con opiniones de otro dominio diferente.

3 Organización

Este trabajo de tesis está organizado en seis capítulos: en el Capítulo 1 se describe la introducción al tema de la detección de opiniones falsas, los siguientes tres capítulos (Capitulo 2, Capítulo 3 y Capítulo 4), corresponden a las publicaciones realizadas sobre el nuevo método de detección de opiniones falsas basado en PU-Learning que se propone, a este nuevo método le llamamos PU-Learning*. En el Capítulo 5 se realiza la discusión de los resultados obtenidos y por último en el Capítulo 6 se presentan las conclusiones y el trabajo futuro.

Una breve síntesis del contenido de los capítulos presentados se muestra a continuación:

Capítulo 2: Se presenta el trabajo:"Using PU-Learning to detect deceptive Opinion Spam"publicado en el workshop WASSA 2013 (Hernández et al., 2013). En este artículo se utiliza el método de PU-Learning* por primera vez, evaluándolo en un conjunto de opiniones positivas (falsas y verdaderas) sobre hoteles situados en el centro de la ciudad de Chicago. En esta publicación se compara el nuevo método con otros que requieren un conjunto completamente etiquetado de opiniones para su funcionamiento y se obtienen valores de f-measure promedio de hasta 0.840.

Capítulo 3: Se presenta el artículo "Detecting positive and negative opinions using PU-Learning"publicado en la revista Information Processing & Management (Hernández et al., 2015a), donde se evalúa el método de PU-Learning* en un conjunto de opiniones que contiene tanto opiniones positivas como negativas de 20 hoteles del área del centro de Chicago. En este artículo se compara la precisión obtenida usando el método de PU-Learning* contra el método de PU-Learning tradicional, empleando diferentes tipos de atributos. También se realiza un análisis de significancia estadística entre los atributos empleados, para determinar la mejor combinación que nos lleva a obtener resultados con la precisión más alta.

Capítulo 4: Se presenta la publicación de la conferencia CICLing 2015 "Detection of opinion spam with character n-grams"(Hernández et al., 2015b), donde se evalúa y compara el método de PU-Learning* para detectar las opiniones falsas utilizando diferentes tipos de atributos, particularmente n-gramas de palabras y n-gramas de caracteres. En este trabajo se presenta también un análisis sobre la robustez de los n-gramas de caracteres en la clasificación con pocos datos de entrenamiento.

Capítulo 5: Se presenta una descripción completa de los corpus utilizados en los experimentos realizados para la evaluación del método de PU-Learning*, así como los resultados obtenidos en los experimentos diseñados para probar su eficacia. También se realizan pruebas de dominios cruzados, donde se entrena con un dominio diferente al dominio con el que se realiza la prueba.

Capítulo 6: Se presentan las conclusiones del desarrollo del trabajo de tesis, algunas ideas para desarrollar en el futuro y la lista de las publicaciones generadas.

4 Contribuciones

La principal contribución de este trabajo es haber diseñado el método llamado PU-Learning*, que permite detectar opiniones falsas partiendo de un conjunto pequeño de

instancias etiquetadas como opiniones falsas y otro conjunto más grande de instancias no-etiquetadas. Los resultados experimentales nos permiten concluir que sí es posible detectar de una manera más efectiva las opiniones falsas usando el método propuesto.

El método de PU-Learning* se puede describir de una manera breve por medio del algoritmo 1. En este algoritmo se puede apreciar como se van extrayendo del conjunto no etiquetado, aquellas instancias a las cuales el clasificador les asigna una etiqueta positiva. Al final de un número finito de iteraciones, solo permanecen en el conjunto de instancias no-etiquetadas, aquellas que pasaran a formar parte del conjunto de instancias negativas.

Algoritmo 1 PU-Learning* para la detección de Opinion Spam

$i \leftarrow 1$
$|W_0| \leftarrow |U_1|$
$|W_1| \leftarrow |U_1|$
while $|W_i| <= |W_{i-1}|$ **do**
  $C_i \leftarrow Generate\_Classifier(P, U_i)$
  $U_i^L \leftarrow C_i(U_i)$
  $W_i \leftarrow Extract\_Positives(U_i^L)$
  $U_{i+1} \leftarrow U_i - W_i$
  $i \leftarrow i + 1$
Return Classifier C_i

Mediante la aplicación del algoritmo 1 y usando los atributos apropiados, se obtiene como resultado la clase faltante, la de las opiniones verdaderas.

Adicionalmente los experimentos realizados nos permitieron formular las siguientes conclusiones respecto al método PU-Learning*:

- El método de PU-Learning* es más efectivo para la detección de opiniones falsas que el método PU-Learning tradicional; esto se debe en gran manera al uso de criterios más conservadores y exigentes para la selección de las opiniones (presumiblemente) falsas del conjunto no-etiquetado.
- El método de PU-Learning* es efectivo para la detección de opiniones falsas de ambas polaridades, sin embargo, los resultados obtenidos fueron mejores en la detección de opiniones falsas positivas. Este comportamiento del método propuesto coincide con trabajos previos, y se origina a partir de la mayor diversidad y mayor nivel de especificidad, de las opiniones falsas negativas.
- El método de PU-Learning* es una solución adecuada para la detección de opiniones falsas en escenarios de dominios cruzados, particularmente cuando los dominios fuente y objetivo son afines, es decir, cuando estos muestran una alta intersección en el vocabulario empleado.

Adicional al diseño del método PU-Learning*, una segunda contribución de este trabajo es la propuesta de una representación de los documentos apropiada para esta tarea, que incorpora tanto aspectos de su contenido como de su estilo. Esta representación fue realizada a través de los n-gramas de caracteres. Los resultados obtenidos con esta representación son superiores a los obtenidos con la representación tradicional de bolsa de palabras, permitiendo concluir que la información estilística es también importante para la detección de opiniones falsas.

Los experimentos realizados con ambas representaciones, también indican que para modelar adecuadamente el estilo de escritura de las opiniones falsas y verdaderas es necesario disponer de grandes conjuntos de entrenamiento; usando conjuntos pequeños las diferencias en los resultados de ambas representaciones fueron estadística-mente no significativas.

Bibliografía

Hernández, D., R. Guzmán, M. M. y Gómez, y P. Rosso. 2013. Using PU-learning to detect deceptive opinion spam. En *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, volumen 1606 de *Association for Computational Linguistics*, páginas 38–45, Atlanta, Georgia. Association for Computational Linguistics.

Hernández, D., M. M. y Gómez, P. Rosso, y R. Guzmán. 2015a. Detecting positive and negative deceptive opinions using PU-learning. *Information Processing & Management*, 51(4):433–443.

Hernández, D., M. M. y Gómez, P. Rosso, y R. Guzmán. 2015b. Detection of opinion spam with character n-grams. En A. Gelbukh, editor, *Proc. of 16th International*

conference on Intelligent Text Processing and Computational Linguistics (CICLing-2015), Lecture Notes in Artificial Intelligence, páginas 285–294, Cairo, Egypt, April. Springer-Verlag. Vol. 9042 part II.