

Detección de sugerencias en tuits de usuarios y seguidores de aerolíneas con procesamiento de lenguaje natural y aprendizaje automático

Rafael Jiménez Castro, Rogelio Florencia Juárez, Vicente García Jiménez y Jorge Rodas-Osollo

Epítome— La interacción que tienen las personas con las redes sociales han creado un cambio cultural en el cual las empresas basan sus campañas de márketing. En este artículo se muestra cómo la minería de sugerencias puede llegar a ser útil para etiquetar comentarios en sugerencias y no sugerencias. Para ello se construirá un repositorio de comentarios reales a partir de las opiniones emitidas por usuarios de redes sociales de aerolíneas. Sobre el repositorio se utilizarán técnicas de procesamiento de lenguaje natural para normalizar los datos contenidos en ellas y facilitar la extracción de características. Este análisis es importante ya que apoyará a la investigación de minería de sugerencias en español al trabajar en un contexto del idioma que se utiliza en México.

Palabras clave— Minería de sugerencias, Clasificación, Aerolíneas, Redes sociales

1 INTRODUCCIÓN

Actualmente el nivel de interacción que tienen las personas en redes sociales ha provocado que las campañas de márketing tengan un cambio en su modelo de negocios [1], pasando de crear anuncios en medios de comunicación impresos, la radio y la televisión con los cuales se podía llegar a un gran número de personas fueran clientes o no, a medios digitales en busca de aumentar el alcance a clientes potenciales. Las nuevas plataformas digitales promueven la interacción con el receptor y gratifican con mayor exposición al contenido que se vuelve viral, lo cual crea valor para la marca y al producto [2]. Las redes sociales han pasado de sitios en línea para interactuar con tus amistades a una plataforma en la cual personas, artistas, marcas y hasta presidentes interactúan con multitudes de personas diariamente.

En particular, las aerolíneas son algunas de las empresas que utilizan las redes sociales como Twitter, para interactuar con sus clientes por medio de mensajes con ofertas, recomendaciones de viaje, videos de colaboraciones con *youtubers* y encuestas.

Los usuarios de las redes sociales cuentan con tres opciones para interactuar con los tuits: 1) compartirlo a los seguidores del usuario por medio de un retuit, 2) darle «me

gusta» cuando el tuit de la aerolínea es de su agrado y 3) responder al tuit emitido por la aerolínea con un comentario en donde el usuario expresa su opinión o sugerencia acerca del tuit al que responde, el servicio o de la compañía.

Dentro de las tantas respuestas que dan los usuarios de las aerolíneas a los tuits, se dejan sugerencias de cómo mejorar sus servicios o procesos. Estas sugerencias pueden llegar a perderse entre los cientos o miles de tuits dirigidos a una empresa. El no atender las sugerencias de los usuarios es un problema ya que se estaría perdiendo información valiosa que estos puedan contener, con lo cual, se desaprovecharía la oportunidad de mejorar el servicio ofrecido por la aerolínea y de retroalimentación para la evaluación de la efectividad de las campañas publicitarias que se están realizando. Además, el usuario al sentir que su opinión no esta siendo tomada en cuenta, puede perder confianza en la aerolínea con lo que es posible ver una afectación negativa de futuras ventas.

En este momento se realizan pocas aplicaciones para la minería de sugerencias en comparación con el número de proyectos de minería de opiniones encaminadas al análisis de sentimientos. En comparación con el análisis de sentimientos, el cual tiene el fin de clasificar los comentarios como positivos, negativos o neutros, la minería de sugerencias trata de identificar los comentarios que contengan consejos de manera implícita ó explícita. La minería de sugerencias se ha aplicado a reseñas con dominio abierto [3] [4], en redes sociales [5] [6] y en reseñas del consumidor [7].

Unas de las principales limitaciones de las aplicaciones de minería de sugerencias es el contexto en el que se manejan los comentarios a clasificar, ya que

- F.A. Maestría en Computo Aplicado en Universidad Autonoma de Ciudad Juarez, Ciudad Juárez, Chihuahua, México. E-mail: al182849@alumnos.uacj.mx
- S.B. División Multidisciplinaria en Ciudad Universitaria Universidad Autónoma de Ciudad Juárez Ciudad Juárez, Chihuahua, México. E-mail: vicente.jimenez@uacj.mx.
- T.C. División Multidisciplinaria en Ciudad Universitaria Universidad Autónoma de Ciudad Juárez Ciudad Juárez, Chihuahua, México-E-

tecnológico en forma de un prototipo de clasificación que utilice procesamiento de lenguaje natural, que permita extraer las características de los tuits emitidos por los usuarios de redes sociales de aerolíneas, y aprendizaje automático para etiquetar los comentarios en sugerencias y no sugerencias.

2 OBJETIVO

Clasificación de tuits en español emitidos por usuarios de redes sociales implementando un prototipo que combine técnicas de procesamiento de lenguaje natural y aprendizaje automático.

2.2 Objetivos Específicos

1. Crear un repositorio de tuits de usuarios de aerolíneas.
2. Preprocesar el repositorio mediante la selección, integración y formateo de los datos.
3. Construir y evaluar un modelo de clasificación supervisado.

2.3 Justificación

La implementación de un prototipo para la minería de sugerencias en comentarios realizados mediante la plataforma Twitter tendrá efectos favorables para las aerolíneas que utilicen redes sociales, tales como:

- Clasificar de forma automática los tuits en sugerencias y no sugerencias.
- Detectar de una forma automática las ideas consideradas por los usuarios como apropiadas para mejorar el servicio de las aerolíneas. Al tener las ideas detectadas, se evita tener que implementar recursos en estudios de mercados para identificar cuáles son las necesidades actuales de los clientes.
- Detectar las tendencias en las ideas de los clientes y adaptando las campañas de publicidad para cubrir esas tendencias y aumentar su cuota de mercado.
- Mejorar la atención al cliente y al mismo tiempo reforzar la imagen de la empresa al detectar las sugerencias más populares con lo cual se le puede dar retroalimentación al usuario de que la sugerencia está siendo atendida o esta en proceso de implementación si la empresa cree que sea conveniente.

3 METODOLOGIA

La metodología de investigación que se utilizará será *Action Research* [8], la cual está conformada por cinco fases:

1. Identificación del problema: En esta fase se trata de identificar el área que se quiere explorar, así como determinar el enfoque que se trabajará.
2. Recolección y organización de datos: La recolección de información se hace a través de varias fuentes para entender mejor el enfoque de lo que está sucediendo.
3. Interpretación de los datos: Los datos son analizados

- que ayude a realizar cambios y a estudiar ese cambio.
5. Reflexión: Se evalúan los resultados para determinar si hubo una mejora.

4 IMPACTO, ALCANCES Y LIMITACIONES

4.1 Impacto

El prototipo que se creará será de impacto ya que:

- Podrá identificar las sugerencias de los usuarios de forma automatizada en español.
- Apoyará a la investigación de minería de sugerencias en español al trabajar en un contexto del idioma que se utiliza en México, en contraste con la gran mayoría de la literatura la cual abordan problemas y soluciones utilizando el idioma inglés.

4.2 Alcances

- Se implementará un prototipo para la clasificación que apoye en la detección de sugerencias en tuits de usuarios y seguidores de aerolíneas en Twitter.
- Se integrará un clasificador supervisado para identificar los comentarios que contengan sugerencias.
- Para el entrenamiento se utilizarán los tuits en las cuentas de las aerolíneas mexicanas: Aeromar, Aeroméxico, Interjet, Magnicharters, TAR, Viva Aerobus y Volaris.

4.3 Limitaciones

- El prototipo no tendrá conexión a ninguna fuente de información en redes sociales para extraer los comentarios de los usuarios.
- La clasificación sólo podrá utilizar los algoritmos de clasificación incluidos en la librería de herramientas WEKA.
- El usuario tendrá que ingresar de forma manual los comentarios a clasificar.
- Los comentarios para clasificar serán en español mexicano.
- Los comentarios utilizados tendrán un tamaño máximo de 280 caracteres que es el máximo número de caracteres permitidos por Twitter.

4.4 Validación

Para la validación del prototipo se utilizará una validación cruzada con 5 particiones. El método divide el conjunto total de las opiniones, previamente etiquetadas, en \mathcal{K} particiones, donde $\mathcal{K} - 1$ particiones se agrupan para formar el conjunto de entrenamiento y el resto se utiliza como prueba. Este proceso se repite eligiendo una partición diferente como conjunto de prueba.

REFERENCIAS

- [1] M. Saravanakumar y T. SuganthaLakshmi, «Social Media Marketing,» *Life Science Journal*, vol. 9, nº 4, pp. 4444 - 4451, 2012.

brand and value consciousness,» *Asia Pacific Journal of Marketing and Logistics*, vol. 29, nº 1, pp. 129-144, 2017.

- [3] S. Nagi y P. Buitelaar, «Towards the Extraction of Customer-to-Customer Suggestions from Reviews,» de *Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 2015.
- [4] S. Negi, «Suggestion Mining from Opinionated Text,» de *Association for Computational Linguistics*, Berlin, Germany, 2016.
- [5] T. Pitchayawiwat, «A Study on Clustering Customer Suggestion on Online Social Media about Insurance Services by Using Text Mining Techniques,» de *Management and Innovation Technology International Conference*, Bang Saen, Thailand, 2016.
- [6] S. Negi, K. Asooja, S. Mehrotra y P. Buitelaar, «A Study of Suggestions in Opinionated Texts and their Automatic Detection,» de *Conference on Lexical and Computational Semantics*, Berlin, Germany, 2016.
- [7] C. Brun y C. Hagege, «Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments,» *Research in Computing Science*, nº 70, pp. 199-209, 2013.
- [8] E. Ferrance, «Action Research,» 2000. [En línea]. Available: https://www.brown.edu/academics/education-alliance/sites/brown.edu.academics.education-alliance/files/publications/act_research.pdf.

forma en la cual se puede representar. Para fortalecer estos conceptos creamos un mapa conceptual donde se integraba la lógica crips, la inteligencia artificial y las subáreas que la conforman.

5. Modelación Conceptual e Ingeniería de Requisitos: Esta clase nos ayudó a definir el modelo conceptual del dominio de aplicación. Para esto, se cubrieron temas donde se definió que era un modelo mental, los tipos de diagramas, los metamodelos, el dominio de aplicación, y reforzamos la formalidad del lenguaje UML.

Aparte de estas cinco clases, durante el semestre trabajé en mi proyecto de investigación, el cual este contenido en este protocolo. Para lograr este progreso, tuve juntas con mis asesores para determinar el dominio de aplicación e ir revisando lo que iba trabajando durante el semestre.

Reporte Semestral

Durante el semestre cursé cinco clases, las cuales fueron:

1. Matemáticas para Computación: En esta clase se discutieron la teoría y formalidad en temas de funciones, lógica, grafos, probabilidad y estadística y matrices. Así como aplicaciones del mundo real donde se aplican estos conceptos.
2. Innovación y transferencia de tecnología: En esta clase definimos lo que es innovación y que no todos los nuevos productos que salen al mercado son innovadores ya que no tienen un impacto significativo en los esquemas de negocio actualmente utilizados, en lo económico o de impacto social.
3. Seminario de titulación I: En esta clase se nos enseñó a utilizar las fuentes de información que tenemos disponibles, la forma de llevar un registro de los artículos de interés para nuestro proyecto y los tipos de investigación en los cuales nuestro proyecto puede recaer. Además, se trabajó en como ir formando nuestro protocolo de investigación, desde formarmación una introducción hasta como estructurarlo.
4. Seminario de Cómputo Cognitivo: En esta clase