

Explorando la generación de contenido online por el usuario y su influencia predictiva en la Calidad Relacional. Aplicación al sector hotelero de Andalucía

Exploring the generation of online content by users and its predictive influence on the Relational Quality. Application to the Andalusian hotel sector

M.J. Sánchez-Franco¹, José A. Troyano¹, Fermín L. Cruz¹ and M. Alonso-Dos-Santos²

¹Universidad de Sevilla

²Universidad de Granada

Abstract

Nuestro proyecto se centra en la identificación de factores competitivos de los establecimientos hoteleros en Andalucía. Para ello usaremos técnicas de Procesamiento del Lenguaje Natural aplicadas a las opiniones *online* publicadas por los usuarios, a través de plataformas de infomediación como TripAdvisor. Estamos especialmente interesados en la identificación de métricas que permitan cuantificar el concepto de Calidad Relacional. El equipo de investigadores del proyecto está compuesto tanto por expertos en Marketing Relacional, como por expertos en Tecnologías del Lenguaje. Ambas visiones complementarias serán de gran ayuda, tanto en el diseño experimental como en la transferencia de resultados al sector turístico.

English translation. Our project focuses on the identification of competitive factors of hotel establishments in Andalusia. To do so, we will use Natural Language Processing techniques applied to online reviews published by users through platforms such as TripAdvisor. We are especially interested in the identification of metrics that allow us to quantify the concept of Relational Quality. The research team of the project is composed of both experts in Relationship Marketing and experts in Language Technologies. Both complementary visions will be of great help, both in the experimental design and in the transfer of results to the tourism sector.

Keywords

Contenidos generados por usuarios, calidad relacional, tecnologías del lenguaje, sector turístico.

1. Introducción

El desarrollo de Internet ha supuesto un cambio de paradigma en la forma en la que se inspiran, se contratan, se organizan y se viven los viajes turísticos, y ha cambiado de la manera en la que los turistas toman sus decisiones, debido al intercambio de opiniones y experiencias mediante el uso de aplicaciones, redes sociales y otros espacios en Internet [1]. En particular, los viajeros buscan asesoramiento desinteresado antes de reservar un hotel, y consultan las opiniones emitidas y las valoraciones de los establecimientos hoteleros en plataformas de infomediación (TripAdvisor, Expedia, Yelp, Booking, ...). El entorno global

en que participa el sector, y las transformaciones singulares debidas a la irrupción de las tecnologías de la información y la comunicación junto a los sistemas de información asociados, provocan un tránsito desde modelos tradicionales *offline* de búsqueda de servicios hacia otros modelos de consulta, reserva o compra basados en la publicación de contenidos generados por el propio usuario en sistemas de reservas *online* o de recomendación [2]. Las revisiones *online* son espontáneas, esclarecedoras, e incluso apasionadas, fácilmente accesibles desde cualquier lugar y en cualquier momento [3] son recuerdos o reconstrucciones cognitivas de un viaje o una estancia que reducen significativa y aparentemente el riesgo potencial de la compra [4]. Si bien los contenidos compartidos que recrean las experiencias del huésped pueden estar intencionadamente distorsionados, la información se percibe como creíble y desinteresada, y se convierte en la clave que sustenta sus decisiones [5].

En suma, las opiniones *online* creadas por el usuario desempeñan un papel crucial en la construcción de la reputación de los hoteles, y consecuentemente en la atracción de usuarios y su retención. El

SEPLN-PD 2022. Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations, September 21-23, 2022, A Coruña, Spain

✉ majesus@us.es (M.J. Sánchez-Franco); troyano@us.es (J. A. Troyano); fcruz@us.es (F. L. Cruz); manuelalonso@ugr.es (M. Alonso-Dos-Santos)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



estudio de los contenidos creados es además necesario para intensificar la calidad relacional. Precisamente este concepto de calidad relacional es concebido como una propuesta adecuada para explicar y predecir el éxito de una relación entre el establecimiento hotelero y el usuario medido habitualmente a través de la lealtad, y basado en la teoría del compromiso en las relaciones [6].

Nuestro proyecto de investigación propone una aproximación basada en técnicas de Procesamiento del Lenguaje Natural (PLN) para extraer las cualidades de los productos hoteleros a partir de la experiencia comunicada del turista durante sus estancias hoteleras (por ejemplo, la localización del establecimiento, la calidad del servicio y el valor percibido, la atmósfera del hotel, ...), y por otro lado, predecir a partir de ellas el cumplimiento de las expectativas desde el punto de vista del usuario.

1.1. Experiencia del equipo investigador

En este apartado es de destacar el aspecto multidisciplinar del equipo de investigadores y colaboradores del proyecto. Se ha perseguido la integración de investigadores expertos y conocedores de los campos disciplinares asociados principalmente a las áreas de conocimiento de Comercialización e Investigación de Mercados, y Lenguajes y Sistemas Informáticos, vinculados a cuatro universidades distintas.

Es de esperar que, durante el desarrollo del proyecto, aparezcan sinergias entre investigadores procedentes de muy diversos campos, con experiencias tanto teóricas como experimentales en investigaciones relacionadas con el Marketing y en el Procesamiento del Lenguaje Natural.

1.2. La Calidad Relacional

La satisfacción de los clientes es habitualmente algo difícil de medir, pero es aspecto fundamental para cualquier empresa. En especial para mantener un alto grado de retención de clientes. En este proyecto nos basaremos en el concepto de calidad relacional [7] como un instrumento para cuantificar ese grado de satisfacción. Trabajaremos con tres escalas que dan estructura a este concepto:

- Satisfacción: sentimiento del cliente de que sus expectativas se han cumplido.
- Confianza: certeza del cliente de que la empresa cumplirá sus expectativas.
- Compromiso: deseo del cliente de mantener su relación con la empresa.

La investigación tradicional se basa en cuestionarios para evaluar las distintas escalas de la calidad

relacional. Esto supone un alto coste, además de presentar distintos inconvenientes como son la introducción de sesgos a través de la redacción de las preguntas, el esfuerzo que se le exige al cliente, o la imposibilidad de adaptarse a sectores tan dinámicos como el turismo por la escasa frecuencia con la que se actualizan los propios cuestionarios.

El reto que nos planteamos en este proyecto es precisamente el de construir una alternativa, a este modelo clásico de evaluar la calidad relacional, usando textos publicados por los propios usuarios y automatizando el proceso mediante el uso de tecnologías del lenguaje.

2. Objetivos

El objetivo general del proyecto es validar empíricamente hipótesis del ámbito del Marketing Relacional, mediante la aplicación de técnicas de Procesamiento del Lenguaje Natural. En concreto se pretende:

- Aplicar un marco global de interpretación de los contenidos publicados en las plataformas de infomediación basado en la disciplina de Marketing Relacional. En particular centrándonos en la fase del compromiso verdadero [7].
- Aplicar técnicas de Procesamiento del Lenguaje Natural para identificar métricas que permitan valorar, a través de la calidad relacional, la imagen de los establecimientos hoteleros y por extensión de Andalucía como destino turístico global.

No hemos encontrado trabajos previos que aborden el análisis de la calidad relacional desde la perspectiva del Procesamiento del Lenguaje Natural.

3. Metodología y resultados esperados

En este apartado, resumiremos nuestra aproximación metodológica y los resultados esperados, tanto a nivel experimental como de transferencia de conocimiento.

3.1. Metodología

La figura 1 muestra los elementos más significativos del proceso metodológico que pretendemos seguir en el proyecto. En el diagrama se reflejan los dos tipos de perfiles investigadores mediante un ordenador (experto en tecnologías del lenguaje) y una persona



Figure 1: Proceso metodológico

(experto en el dominio de marketing). Bajo cada una de las tres fases del proceso se indican, con estos iconos, qué perfiles de investigadores estarán involucrados.

En la primera fase de descubrimiento de datos y análisis exploratorio, ambos perfiles son necesarios. El principal resultado de esta fase serán las preguntas de investigación identificadas, que han sido consideradas como interesantes por parte de los expertos en el dominio, y valoradas como viables por los expertos en PLN.

La segunda fase es de corte totalmente experimental, y en ella los expertos en PLN convertirán las preguntas en tareas evaluables, y desarrollarán sistemas para resolver dichas tareas.

En la fase final de análisis, vuelven a participar ambos perfiles. Es el momento de extraer conclusiones e interpretar los resultados obtenidos.

El proceso implica una revisión continua, que se refleja con los dos bucles que permiten identificar nuevas preguntas de investigación interesantes en todo momento.

En cuanto a las técnicas a aplicar, aparte de las herramientas PLN para el desarrollo de *pipelines* clásicos, nuestra intención es basar los experimentos principalmente en modelos pre-entrenados de transformers. Para tareas semánticas como las que pretendemos resolver, son claramente la mejor alternativa.

3.2. Resultados experimentales

La evaluación de las distintas tareas PLN identificadas, nos permitirá medir la eficacia de los sistemas desarrollados a la hora de responder a las preguntas de investigación.

Hay algunas tareas definidas de antemano, en especial las que están relacionadas con la definición de una métrica para dar respaldo al concepto de calidad relacional. Pero más allá de eso, no nos cerramos a usar ninguna de las técnicas y herramientas disponibles en el área del PLN. Desde soluciones léxicas para extraer términos relevantes [8], pasando por modelos basados en bolsas de palabra para tareas de clasificación [9], *pipelines* clásicos para tareas de extracción de información [10], hasta las más recientes técnicas basadas en redes profundas [11] o *transformers* [12].

Los corpus que utilizaremos serán principalmente de desarrollo propio. De hecho, invertiremos un porcentaje importante de las horas de dedicación del proyecto a la recopilación y etiquetado de corpus específicos, que nos permitan entrenar y evaluar nuestros sistemas.

3.3. Resultados de transferencia

El hecho de que el equipo de investigación sea multidisciplinar, y que desde el principio participen en la definición de las tareas expertos en el dominio de marketing, hace que la transferencia de conocimiento sea un paso natural en nuestro proyecto.

Estamos convencidos de que muchas de las tareas que identifiquemos pueden dar lugar a productos software que tengan un claro interés gerencial, tanto desde la perspectiva general del marketing relacional, como en el dominio específico de la gestión hotelera.

Nuestra intención es difundir nuestros resultados, tanto entre agentes públicos como privados, con idea de encontrar oportunidades de transferencia. La revisión de los resultados experimentales desde

la perspectiva de los gestores, nos dará un punto de vista mucho más aplicado y nos permitirá refinar nuestros experimentos para orientarlos al desarrollo de productos de utilidad para el sector hotelero.

Acknowledgments

Financiado por el proyecto US-1380960 de la convocatoria de proyectos de I+D+i en el marco del Programa Operativo FEDER. Consejería de Transformación Económica, Industria, Conocimiento y Universidades. Junta de Andalucía.

References

- [1] J. y. A. L. Consejería de Turismo, Regeneración, Plan General de Turismo Sostenible de Andalucía Horizonte 2020, Technical Report, Junta de Andalucía, 2016.
- [2] E. Raguseo, P. Neirotti, E. Paolucci, How small hotels can drive value their way in information. the case of 'italian hotels vs. otas and tripadvisor', *Information & Management* 54 (2017) 745–756.
- [3] Y. Guo, S. J. Barnes, Q. Jia, Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation, *Tourism management* 59 (2017) 467–483.
- [4] B. A. Sparks, K. K. F. So, G. L. Bradley, Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern, *Tourism Management* 53 (2016) 74–85.
- [5] M. J. Sánchez-Franco, A. Navarro-García, F. J. Rondán-Cataluña, Online customer service reviews in urban hotels: A data mining approach, *Psychology & Marketing* 33 (2016) 1174–1186.
- [6] A. Parasuraman, D. Grewal, The impact of technology on the quality-value-loyalty chain: a research agenda, *Journal of the academy of marketing science* 28 (2000) 168–174.
- [7] B. Fogg, D. Eckles, The behavior chain for online participation: how successful web services structure persuasion, in: *International Conference on Persuasive Technology*, Springer, 2007, pp. 199–209.
- [8] A. Peñas, F. Verdejo, J. Gonzalo, et al., Corpus-based terminology extraction applied to information access, in: *Proceedings of corpus linguistics*, volume 2001, 2001, p. 458.
- [9] T. M. Alam, M. J. Awan, Domain analysis of information extraction techniques, *Int. J. Multidiscip. Sci. Eng* 9 (2018) 1–9.
- [10] Y. HaCohen-Kerner, D. Miller, Y. Yigal, The influence of preprocessing on text classification using a bag-of-words representation, *PLoS one* 15 (2020) e0232525.
- [11] W. Li, L. Zhu, Y. Shi, K. Guo, E. Cambria, User reviews: Sentiment analysis using lexicon integrated two-channel cnn-lstm family models, *Applied Soft Computing* 94 (2020) 106435.
- [12] M. Munikar, S. Shakya, A. Shrestha, Fine-grained sentiment classification using bert, in: *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, volume 1, IEEE, 2019, pp. 1–5.