

## ANÁLISIS DEL GÉNERO DISCURSIVO APLICADO A LA CLASIFICACIÓN AUTOMÁTICA DE LA POLARIDAD EN COMENTARIOS SOBRE PRODUCTOS

JOHN A. ROBERTO

[johnroberto2001@yahoo.es](mailto:johnroberto2001@yahoo.es)

Esta tesis trata sobre el análisis de la polaridad en comentarios sobre productos, más exactamente, sobre la clasificación de comentarios como positivos o negativos a partir del uso de información lingüística. En la tesis presento un enfoque al análisis de la polaridad basado en el género discursivo de los comentarios. Según este enfoque, primero se identifican los segmentos que caracterizan el género discursivo de los comentarios y, posteriormente, se evalúa la utilidad que cada tipo de segmento tiene para determinar la polaridad de los comentarios.

La tesis se divide en dos partes. En la primera parte, caracterizo los comentarios como un género mediante el análisis de su estructura discursiva y su registro lingüístico. Sobre la base de ambos análisis postulo que los comentarios se componen de tres tipos principales de segmentos: valorativo, narrativo y descriptivo. En la segunda parte de la tesis, utilizo estos segmentos para calcular la polaridad de los comentarios. La hipótesis de partida es que no todos los segmentos que forman parte del género discursivo de los comentarios contribuyen de la misma manera a expresar la polaridad.

Para validar esta hipótesis evalué tres métodos complementarios que tienen como objetivo detectar y determinar de forma automática la utilidad que tienen los tipos de segmentos para predecir la polaridad de los comentarios. El primer método utiliza información léxica y morfosintáctica para identificar el tipo de segmento que expresa mejor la polaridad del comentario. El segundo método analiza la función que desempeñan las secuencias narrativas en el cálculo de la polaridad. El tercer método se basa en el cálculo de la complejidad sintáctica para identificar y eliminar las oraciones que tienen una polaridad opuesta a la del comentario (oraciones asimétricas) como paso previo a la identificación de los comentarios positivos y negativos.

La conclusión principal que se desprende de estos análisis es que existe una relación directa entre el tipo de segmento y la polaridad expresada en el comentario: los usuarios suelen emplear de manera diferente los segmentos según se trate de un comentario positivo o uno negativo. Estas diferencias en el uso de los segmentos me ha llevado a plantear la existencia de dos (sub)géneros discursivos asociados a la expresión de opiniones sobre productos en la Web: el (sub)género de los comentarios positivos y el (sub)género de los comentarios negativos.

## GENRE-DISCOURSE ANALYSIS APPLIED TO AUTOMATIC CLASSIFICATION OF POLARITY IN CUSTOMER REVIEWS

This thesis is about polarity analysis of reviews, that is, classifying reviews as either positive or negative based on linguistic evidence. I describe a genre-based approach for the polarity analysis of customer reviews. Genre is characterized by a schematic structure of the discourse composed of different types of stages, each one with a goal-oriented function. This approach to polarity analysis, first, distinguishes stages in the genre of reviews and, subsequently, evaluates the usefulness of each type of stage in the determination of the polarity of the entire review.

The thesis is broadly divided into two parts. In the first part, I characterize customer reviews as a discursive genre by analyzing both their structure and their linguistic register. Based on this analysis, I postulate that customer reviews are composed of three main types of stages: valorative, narrative and descriptive. In the second part of the thesis, I determine the usefulness of the different type of stages for the classification of positive and negative reviews. The rationale behind our approach is the assumption that within the specific genre of customer reviews, not all parts of a text contributes equally to expressing the underlying sentiment.

In order to validate this hypothesis, I evaluate three alternative methods used to automatically detect and determine the usefulness of each type of stage in the detection of the polarity of the entire review. The first method applies lexical and morphosyntactic information to identify the type of stage that best expresses the polarity of the whole review. The second method analyzes the role played by narrative chains in determining the polarity of reviews. The third method is based on the measurement of syntactic complexity to detect and remove descriptive sentences with the opposite polarity to that of the entire document (asymmetric sentences) as a previous step to identify positive and negative reviews.

The main conclusion that has been drawn is that there is a relationship between the types of stages and the polarity expressed in the review: users often employ stages differently according to whether its polarity is positive or negative. These differences in use of stages leads me to the conclusion that there are two (sub)genres, rather than one, for the expression of opinions on the Web: the (sub)genre of positive reviews and the (sub)genre of negative reviews.