

URSA - User Review Structure Analysis: Understanding Online Reviewing Trends

Gayatree Ganu
Dept. of Computer Science
Rutgers University
Piscataway, New Jersey 08854

Amlie Marian
Dept. of Computer Science
Rutgers University
Piscataway, New Jersey 08854

Nomie Elhadad
Dept. of Computer Science
Columbia University
New York, NY 10032

April 15, 2010

Abstract

Online reviews are an important asset for users deciding to buy a product, see a movie, or go to a restaurant, as well as for businesses tracking user feedback. However, most reviews are written in a free-text format, and are therefore difficult for computers to understand, analyze, and aggregate. One consequence of this lack of structure is that searching text reviews is often frustrating for users; keyword searches typically do not provide good results as the same keywords routinely appear in good and in bad reviews. The textual body of reviews contains very rich information and user experience in accessing reviews would be greatly improved if the structure and sentiment information conveyed in the content of the reviews were taken into account. Our work focuses on an analysis of free-text reviews by means of classification of reviews at the sentence level, with respect to both the topic and the sentiment expressed in the sentences. Additionally in this article, we report on the insight on user-reviewing behavior and trends that we gained during our analysis. Our work shows that there is large amount of significant data in the hitherto untapped textual part of user reviews. Our large open-source corpus, of peer-authored text with structure and sentiment information, is a valuable resource for researchers to explore several techniques that have so far relied on structured non-textual data.