

IQ boost for web intelligence

August 8 2013

Views and opinions can now be filtered out of very large volumes of online text with greater accuracy than ever before. Thanks to an automatic method developed at MODUL University Vienna, ambiguous terms in online content can now be identified and correctly interpreted. The internationally acknowledged technology recognises correlations between the meaning of words and the specific context of the analysed text snippet. The technology is applicable to a wide range of Internet sources and therefore superior to other methods which must first be "trained" for use in a specific domain.

Product assessments, or hotel and movie reviews - opinions are made and intensified with lightning speed on the Internet. The resulting "top" or "flop" can translate into billions in sales or losses. Companies are therefore relying more and more on Web intelligence, the rapid identification of broad sentiment trends through the analysis of Web documents. The economic significance of these trends has generated a desire for accurate methods to identify them automatically. Such methods are now available, thanks to a series of pioneering innovations developed by the team of Professor Arno Scharl, Head of the Department of New Media Technology at MODUL University Vienna.

Resolving ambiguity

The team tackled a well-known problem: the automatic interpretation of terms whose meaning is altered by the context in which they are used. In an online hotel review, for example, using the word "complaint" immediately triggers [negative connotations](#). However, this is not the case

if it arises in a sentence like "my only complaint would be ..." - or, in other words, embedded in a positive review that concludes with constructive criticism. Professor Scharl explains: "Simple systems for the detection of sentiment do not recognise a shift in what is known as polarity from negative to positive. Considered in isolation, the term "complaint" would always be classified as negative. And because the entire text is ultimately assessed according to the frequency of "predominantly negative" or "predominantly positive" terms, the risk of an incorrect analysis increases in such cases."

A key aspect of this method, which has now been published in the renowned expert journal IEEE Intelligent Systems, involves the production of "contextualized sentiment lexicons". The purpose of such a database is to link sentiment terms whose polarity can switch with other terms whose polarity remains constant.

Neighbourhood watch

In the learning phase, the system detects sentiment terms which can, depending on their context, convey positive and negative sentiment. Subsequently, it connects these "ambiguous" terms with context terms, i.e. frequently co-occurring terms. It calculates probabilities for their cooccurrence in texts previously categorised as positive or negative by human readers and stores them in the contextualized sentiment lexicon. In the application phase, the system interprets the context of an ambiguous term in an unread document and infers its polarity from the given co-occurring terms. Professor Scharl further explains how the method works: "All ambiguous terms in a text are assigned a score that expresses the polarity and strength of the expressed sentiment. The scores of ambiguous terms are then added to those of unambiguous terms. The total reflects the sentiment of the entire document."

Another important advantage of the new method is its domain-

independence. Other existing systems that are optimised for film reviews, for example, do not perform well when applied to product reviews. However, the method developed at MODUL University Vienna analyses a wide range of text types to find commonalities among these genres. This particular advantage can be traced back to the comprehensive portfolio of semantic technologies developed at MODUL University Vienna in recent years - particularly through the research project DIVINE (Dynamic Integration and Visualization of Information from Multiple Evidence Sources). The results of this project, which is funded by the Austrian Research Promotion Agency (FFG) and the Federal Ministry for Transport, Innovation and Technology (BMVIT), have been instrumental to advance the webLyzard Web intelligence platform. The latter monitored online opinions in the context of the US presidential election since 2004, and later carried off the first prize in the "Web 2.0" category of the Austrian National Award for Multimedia and e-Business in 2008.

More information: www.weblyzard.com/divine

Provided by MODUL University Vienna

Citation: IQ boost for web intelligence (2013, August 8) retrieved 16 August 2024 from <https://phys.org/news/2013-08-iq-boost-web-intelligence.html>

| |
|--|
| <p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p> |
|--|