A Survey of Multimodal Sentiment Analysis

Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih-Fu Chang, Maja Pantic

PII: S0262-8856(17)30119-1
DOI: doi: 10.1016/j.imavis.2017.08.003
Reference: IMAVIS 3638

To appear in: *Image and Vision Computing*

Received date: 31 July 2016
Revised date: 25 May 2017
Accepted date: 22 August 2017

Please cite this article as: Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih-Fu Chang, Maja Pantic, A Survey of Multimodal Sentiment Analysis, *Image and Vision Computing* (2017), doi: 10.1016/j.imavis.2017.08.003

A Survey of Multimodal Sentiment Analysis

Mohammad Soleymani[a,*], David Garcia[b], Brendan Jou[c,**], Björn Schuller[d,e,a], Shih-Fu Chang[c], Maja Pantic[e,f]

[a]*Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland*
[b]*Chair of Systems Design, ETH Zürich, Zurich, Switzerland*
[c]*Electrical Engineering Department and Computer Science Department, Columbia University, New York, NY, USA*
[d]*Chair of Complex and Intelligent Systems, University of Passau, Passau, Germany*
[e]*Department of Computing, Imperial College London, London, United Kingdom*
[f]*Human Media Interaction, EEMCS, University of Twente, Enschede, The Netherlands*

Abstract

Sentiment analysis aims to automatically uncover the underlying attitude that we hold towards an entity. The aggregation of these sentiment over a population represents opinion polling and has numerous applications. Current text-based sentiment analysis rely on the construction of dictionaries and machine learning models that learn sentiment from large text corpora. Sentiment analysis from text is currently widely used for customer satisfaction assessment and brand perception analysis, among others. With the proliferation of social media, multimodal sentiment analysis is set to bring new opportunities with the arrival of complementary data streams for improving and going beyond text-based sentiment analysis. Since sentiment can be detected through affective traces it leaves, such as facial and vocal displays, multimodal sentiment analysis offers promising avenues for analyzing facial and vocal expressions in addition to the transcript or textual content. These approaches leverage emotion recognition and context inference to determine the underlying polarity and scope of an individual's sentiment. In this survey, we define sentiment and the problem of multimodal sentiment analysis and review recent developments in multimodal sentiment analysis in different domains, including spoken reviews, images, video blogs, human-machine and human-human interaction. Challenges and opportunities of this emerging field are also discussed leading to our thesis that multimodal sentiment analysis holds a significant untapped potential.

Keywords: sentiment, affect, sentiment analysis, human behavior analysis, computer vision, affective computing

1. Introduction

Sentiment is a long-term disposition evoked when a person encounters a specific topic, person, or entity [1]. Understanding people's position, attitude or opinion towards a certain entity has many applications. For example, companies are interested in understanding how their products or their brand is perceived among their customers [2]. Political parties are interested in opinion polling to gauge voting intentions [3]. Automatic sentiment analysis is the computational understanding of one's position, attitude or opinion towards an entity, person or topic [4]. With the advent of the World Wide Web and shortly after, the social web, individuals are enabled to broadly express their opinions through these media. This has provided a very rich resource for opinion mining and sentiment analysis and promoted the development of automatic sentiment analysis [5, 4, 6]. Text-based sentiment analysis has long been the standard bearer in this area and only recently has sentiment analysis from other modalities, such as speech and vision, begun to be considered.

Liu and Zhang [4] defined sentiment analysis as a problem of automatic identification of four components of a sentiment, including, entity, aspect, opinion holder, aspect's sentiment. For example, in the sentence "Sally likes the screen resolution in Nexus 6P," "Nexus 6P" is the entity, "screen resolution" is the aspect, Sally is the opinion holder and the associated sentiment is positive. A successful automatic sentiment analysis system should be able to extract all these four components correctly. The available user-generated data on the Internet, containing people's opinion or sentiment, is unstructured and noisy [6]. Challenges such as negation, irony and ambiguous phrases with implicit hints add to this challenge. Therefore, correct extraction of all these components is very challenging.

Sentiment analysis from text is a well-researched topic that now enjoys a number of industry solutions. Text-based sentiment analysis has been applied to a broad set of applications including movie box-office performance prediction [7], stock market performance prediction [8] and

*Corresponding author
**Now at Google Inc
Email addresses: mohammad.soleymani@unige.ch (Mohammad Soleymani), dgarcia@ethz.ch (David Garcia), bjou@caa.columbia.edu (Brendan Jou), bjoern.schuller@imperial.ac.uk (Björn Schuller), shih.fu.chang@columbia.edu (Shih-Fu Chang), m.pantic@imperial.ac.uk (Maja Pantic)