SPRINGER BRIEFS IN COGNITIVE COMPUTATION

Erik Cambria · Amir Hussain

Sentic Computing
Techniques, Tools, and Applications

Springer

Erik Cambria
Media Laboratory
Massachusetts Institute of Technology
Cambridge
USA

Amir Hussain
Department of Computing Science
University of Stirling
Stirling
UK

ISSN 2212-6023 ISSN 2212-6031 (electronic)
ISBN 978-94-007-5069-2 ISBN 978-94-007-5070-8 (eBook)
DOI 10.1007/978-94-007-5070-8
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012942029

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1 Introduction 1
1.1 Sentic Computing 2
1.1.1 Motivations 3
1.1.2 Aims 5
1.1.3 Methodology 7
References 9

2 Background 11
2.1 Opinion Mining and Sentiment Analysis 11
2.1.1 The Buzz Mechanism 12
2.1.2 Origins and Peculiarities 13
2.1.3 Sub-Tasks 14
2.2 Main Approaches to Opinion Mining 15
2.2.1 From Heuristics to Discourse Structure 16
2.2.2 From Coarse to Fine Grained 17
2.2.3 From Keywords to Concepts 18
2.3 Towards Machines with Common Sense 19
2.3.1 The Importance of Common Sense 19
2.3.2 Knowledge Representation 21
2.3.3 From Logical Inference to Digital Intuition 25
2.4 Conclusions 30
References 31

3 Techniques 35
3.1 Affective Blending: Enabling Emotion-Sensitive Inference 36
3.1.1 AffectNet 36
3.1.2 AffectiveSpace 40
3.2 Affective Categorisation: Modelling Human Emotions 47
3.2.1 Categorical Versus Dimensional Approaches 48
3.2.2 The Hourglass of Emotions 51

3.3 Sentic Medoids: Clustering Affective Common Sense 56
3.3.1 Partitioning Around Medoids 56
3.3.2 Centroid Selection. 57
3.4 Sentic Activation: Two-Level Affective Reasoning 58
3.4.1 Unconscious Reasoning 59
3.4.2 Conscious Reasoning. 60
3.5 Sentic Panalogy: Switching Between Different Ways to Think 61
3.5.1 Changing Reasoning Strategies. 62
3.5.2 Changing Reasoning Foci 63
3.6 Conclusions 64
References 64

4 Tools 69
4.1 SenticNet: A Semantic Resource for Opinion Mining 70
4.1.1 Building SenticNet 70
4.1.2 Working with SenticNet. 72
4.2 Sentic Neural Networks: Brain-Inspired Affective Reasoning . . . 73
4.2.1 Discrete Versus Continuous Approach. 73
4.2.2 Affective Learning 75
4.3 Open Mind Common Sentics: An Emotion-Sensitive IUI 77
4.3.1 Games for Knowledge Acquisition 77
4.3.2 Collecting Affective Common Sense Knowledge 81
4.4 Isanette: A Common and Common Sense Knowledge Base 85
4.4.1 Probase 86
4.4.2 Building the Instance-Concept Matrix 87
4.5 Opinion Mining Engine: Structuring the Unstructured 90
4.5.1 Constitutive Modules. 91
4.5.2 Evaluation 93
4.6 Conclusions 98
References 99

5 Applications. 103
5.1 Development of Social Web Systems. 103
5.1.1 Troll Filtering. 104
5.1.2 Social Media Marketing. 107
5.1.3 Sentic Album 113
5.2 Development of HCI Systems 123
5.2.1 Sentic Avatar 124
5.2.2 Sentic Chat 130
5.2.3 Sentic Corner 131
5.3 Development of E-Health Systems. 136
5.3.1 Crowd Validation 136
5.3.2 Sentic PROMs 138

5.4 Conclusions 142
References 142

6 Concluding Remarks 147
6.1 Summary of Contributions 147
6.1.1 Techniques 148
6.1.2 Tools 148
6.1.3 Applications 149
6.2 Limitations and Future Work 149
6.2.1 Limitations 149
6.2.2 Future Work 151
6.3 Conclusions 151
References 153

Acronyms

3NF	Third Normal Form
AI	Artificial Intelligence
AKAPU	Average Knowledge Acquired Per User
ANN	Artificial Neural Network
API	Application Programming Interface
BACK	Benchmark for Affective Common sense Knowledge
BCNF	Boyce-Codd Normal Form
CF–IOF	Concept Frequency–Inverse Opinion Frequency
CMC	Computer Mediated Communication
CRF	Conditional Random Field
CRM	Customer Relationship Management
DAG	Directed Acyclic Graph
DAU	Daily Active User
DL	Description Logic
ECA	Embodied Conversational Agent
ELM	Extreme Learning Machine
EQ	Emotional Quotient
FMRI	Functional Magnetic Resonance Imaging
FOAF	Friend Of A Friend
FOL	First Order Logic
GUI	Graphic User Interface
GWAP	Game With A Purpose
HCI	Human Computer Interaction
HEO	Human Emotion Ontology
HRQoL	Health Related Quality of Life
HTML	Hyper Text Markup Language
HVS	Human Visual System
ICA	Independent Component Analysis
IM	Instant Messaging
IT	Information Technology
IUI	Intelligent User Interface

KNN	K-Nearest Neighbor
KR	Knowledge Representation
JSON	JavaScript Object Notation
LSA	Latent Semantic Analysis
LSR	Label Sequential Rules
MAU	Monthly Active User
MDS	Multi-Dimensional Scaling
MLP	Multi-Layer Perceptron
MMO	Massively Multiplayer Online
MVE	Minimum Volume Ellipsoid
NB	Naïve Bayes
NELL	Never-Ending Language Learning
NMF	Non-negative Matrix Factorisation
NLP	Natural Language Processing
NP	Nondeterministic Polynomial
OCR	Optical Character Recognition
OMCS	Open Mind Common Sense
OMR	Ontology for Media Resources
OWL	Ontology Web Language
PAM	Partitioning Around Medoids
PCA	Principal Component Analysis
POS	Part Of Speech
PROM	Patient Reported Outcome Measure
RDBMS	Relational Database Management Systems
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SBoC	Small Bag of Concepts
SKOS	Simple Knowledge Organisation System
SQL	Structured Query Language
SNN	Sentic Neural Network
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TF–IDF	Term Frequency–Inverse Document Frequency
TMS	Truth Maintenance System
TSVD	Truncated Singular Value Decomposition
UGC	User Generated Content
UML	Unified Modeling Language
VSM	Vector Space Model
XML	Extensible Markup Language
W3C	World Wide Web Consortium
WNA	WordNet-Affect

Chapter 1
Introduction

We can understand almost anything,
but we can't understand how we understand
Albert Einstein

In a world in which millions of people express their opinions about commercial products in blogs, wikis, forums, chats, and social networks, the distillation of knowledge from this huge amount of unstructured information can be a key factor for marketers who want to create an image or identity in the minds of their customers for their product, brand, or organisation [1]. The automatic analysis of online opinions, however, involves a deep understanding of natural language text by machines, from which we are still very far [2]. Online information retrieval, in fact, is still mainly based on algorithms relying on the textual representation of web pages [3]. Such algorithms are very good at retrieving texts, splitting them into parts, checking the spelling, and counting their words. But when it comes to interpreting sentences and extracting useful information for users, their capabilities are still very limited.

In this book, common sense computing techniques are further developed and applied to bridge the cognitive and affective gap between word-level natural language data and the concept-level opinions conveyed by these. In particular, two common sense knowledge bases are designed, together with a novel emotion categorisation model, and graph mining and multi-dimensionality reduction techniques are applied on them in order to infer cognitive and affective information from natural language text and, hence, develop opinion-mining systems in fields such as Social Web, HCI, and e-health. The structure of the book, specifically, is as follows: this chapter presents motivations, aims, and methodology of the proposed approach; Chap. 2 illustrates the state of the art of opinion mining and common sense computing; Chaps. 3, 4 and 5 explain in detail the developed techniques, tools, and applications, respectively; Chap. 6, finally, comprises concluding remarks and future work.

1.1 Sentic Computing

Sentic computing [4] is a multi-disciplinary approach to sentiment analysis that exploits both computer and social sciences to better recognise, interpret, and process opinions and sentiments over the Web. The approach specifically brings together

E. Cambria and A. Hussain, *Sentic Computing*, SpringerBriefs in Cognitive Computation,
DOI: 10.1007/978-94-007-5070-8_1, © The Author(s) 2012

lessons from both affective computing and common sense computing because, in the field of opinion mining, not only common sense knowledge, but also emotional knowledge is important to grasp both the cognitive and affective information (termed semantics and sentics) associated with natural language opinions and sentiments. Although scientific research in the area of emotion stretches back to the Ninteenth century when Charles Darwin and William James proposed theories of emotion that continue to influence thinking today [5, 6], the injection of affect into computer technologies is much more recent.

During most of the last century, research on emotions was conducted by philosophers and psychologists, whose work was based on a small set of emotion theories that continue to underpin research in this area. The first researchers to try linking text to emotions were actually social psychologists and anthropologists who tried to find similarities on how people from different cultures communicate [7]. This research was also triggered by a dissatisfaction with the dominant cognitive view centred around humans as 'information processors' [8].

Later on, in the 1980s, researchers such as Turkle [9] began to speculate about how computers might be used to study emotions. Systematic research programs along this front began to emerge in the early 1990s. For example, Scherer [10] implemented a computational model of emotion as an expert system. A few years later, Picard's landmark book affective computing [11] prompted a wave of interest among computer scientists and engineers looking for ways to improve human-computer interfaces by coordinating emotions and cognition with task constraints and demands. Picard described three types of affective computing applications:

1. Systems that detect the emotions of the user;
2. Systems that express what a human would perceive as an emotion;
3. Systems that actually 'feel' an emotion.

Although touching upon HCI and affective modelling, sentic computing primarily focuses on affect detection from text. Affect detection is critical because an affect-sensitive interface can never respond to users' affective states if it cannot sense their affective states. Affect detection need not be perfect, but must be approximately on target. Affect detection is, however, a very challenging problem because emotions are constructs (i.e., conceptual quantities that cannot be directly measured) with fuzzy boundaries and with substantial individual difference variations in expression and experience. To overcome such a hurdle, sentic computing builds upon a biologically-inspired and psychologically-motivated affective categorisation model [12] that can potentially describe the full range of emotional experiences in terms of four independent but concomitant dimensions, whose different levels of activation make up the total emotional state of the mind.

In sentic computing, whose term derives from the Latin *sentire* (root of words such as sentiment and sentience) and *sensus* (intended both as capability of feeling and as common sense), the analysis of natural language is based on affective ontologies and common sense reasoning tools, which enable the analysis of text not only at document, page, or paragraph level, but also at sentence and clause level. In particular, sentic computing involves the use of AI and Semantic Web techniques, for knowledge representation and inference; mathematics, for carrying out tasks such as graph mining and multi-dimensionality reduction; linguistics, for discourse analysis and pragmatics; psychology, for cognitive and affective modelling; sociology, for understanding social network dynamics and social influence; finally ethics, for understanding related issues about the nature of mind and the creation of emotional machines.

In this section, motivations for the development of sentic computing are illustrated (Sect. 1.1.1), together with the main aims of the proposed approach (Sect. 1.1.2), and the methodology adopted (Sect.1.1.3).

1.1.1 Motivations

Opinions play a primary role in decision-making processes. Whenever people need to make a choice, they are interested in hearing others' opinions. When this choice involves consuming valuable resources (e.g., spending time and money to buy products or services), in particular, people strongly rely on their peers' past experiences. Just a few years ago, the main sources for collecting such information were friends, acquaintances and, in some cases, specialised magazine or websites. The advent of Web 2.0 has provided people with new tools, e.g., forums, blogs, social networks, and content sharing services, that allow them to create and share, in a time and cost efficient way, their own contents, ideas, and opinions with virtually the millions of people connected to the World Wide Web. This has made available by click a new and oceanic source of information and opinions and has provided a powerful communication medium to share knowledge and to get advantage from others' experiences [13].

Currently, over 75,000 new blogs are created daily, along with 1.2 million new posts each day, and more and more people in the modern world rely on opinions, reviews, and recommendations collected from these and related websites. The Web has made available the opinions of a vast pool of people that are neither our personal acquaintances nor well-known professional critics. People, in fact, are not just naturally keen on listening to others' advice, but also naturally inclined to give others advice. Web users are often happy to share both their positive and negative real-world experiences for different reasons, e.g., because they benefited from others' reviews and want to give back to the community, because they seek for a sense of togetherness in adversity, for cathartic complaining, for supporting a product they really like, because it is a way to express themselves, because they think their opinions are important for others.

When people have a strong feeling about a specific product or service they tried, they feel like expressing it. If they loved it, they want others to enjoy it. If they hated it, they want to warn others away. This huge amount of useful information, however, is mainly unstructured, that is in natural language, as it is specifically produced for human consumption and, hence, it is not directly machine-processable. The opportunity to capture the opinions of the general public about social events, political movements, company strategies, marketing campaigns, and product preferences has raised more and more interest both in the scientific community, for the exciting open challenges, and in the business world, for the remarkable fallouts in marketing and financial market prediction.

This has led to the emerging fields of opinion mining and sentiment analysis, which deal with information retrieval and knowledge discovery from text using data mining and natural language processing (NLP) techniques to distil knowledge and opinions from the huge amount of information on the World Wide Web. Mining opinions and sentiments from natural language, however, is an extremely difficult task as it involves a deep understanding of most of the explicit and implicit, regular and irregular, syntactical and semantic rules proper of a language.

Opinion mining and sentiment analysis are branches of the broad field of text data mining [14] and refer generally to the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents. They can be viewed as an extension of data mining or knowledge discovery from (structured) databases [15, 16]. Although commonly used interchangeably to denote the same field of study, opinion mining and sentiment analysis actually focus on polarity detection and emotion recognition, respectively. Since the identification of sentiment is often exploited for detecting polarity, however, the two fields are usually combined under the same umbrella or even used as synonyms.

As the most natural form of storing information is text, opinion mining is believed to have a commercial potential higher than that of data mining. Opinion mining, however, is also a much more complex task as it involves dealing with text data that are inherently unstructured and fuzzy. It is a multi-disciplinary research area that involves the adoption of techniques in fields such as text analysis, information retrieval and extraction, auto-categorisation, machine learning, clustering, and visualisation. Most of the existing approaches to opinion mining rely on the extraction of a vector representing the most salient and important text features, which is later used for classification purposes. Some of the most commonly used features are term frequency [17] and presence [18]. The latter is a binary-valued feature vectors in which the entries merely indicate whether a term occurs (value 1) or not (value 0) formed a more effective basis for review polarity classification.

This is indicative of an interesting difference between typical topic-based text categorisation and polarity classification. While a topic is more likely to be emphasised by frequent occurrences of certain keywords, overall sentiment may not usually be highlighted through repeated use of the same terms. Other term-based features are often added to the features vector. Position is one of these, in consideration of how the position of a token in a text unit can affect the way in which the token affect the sentiment of the text. Also presence n-grams, typically bi-grams and tri-grams, are often taken into account as useful features. Some methods also relies on the distance between terms. Part of speech (POS) information (nouns, adjectives, adverbs, verbs, etc.) is also commonly exploited in general textual analysis as a basic form of word sense disambiguation [19]. Certain adjectives, in particular, have been proved to be good indicators of sentiment and sometimes have been used to guide feature selection for sentiment classification. In other works, eventually, the detection of sentiments was performed through selected phrases, which were chosen via a number of pre-specified POS patterns, most including an adjective or an adverb [20]. All such approaches mainly rely on parts of text in which opinions and sentiments are explicitly expressed, e.g., polarity terms, affect words, and their co-occurrence frequencies. Opinions and sentiments, however, are often conveyed implicitly through context and domain dependent concepts, which make purely syntactical approaches ineffective.

To this end, novel approaches that go beyond mere word-level sentiment analysis are needed. Such approaches should employ new techniques capable to better grasp the conceptual rules that govern sentiment and the clues that can convey these concepts from realisation to verbalisation in the human mind. Next-generation opinion mining systems need broader and deeper common sense knowledge bases and more cognitive and affective inspired reasoning methods, in order to better understand natural language opinions and sentiments and, hence, more efficiently bridge the gap between (unstructured) textual information and (structured) machine-processable data.

1.1.2 Aims

Today, opinion mining and sentiment analysis find applications in several different scenarios and there is a good number of companies, large and small, that include the analysis of opinions and sentiments as part of their mission. In current product review websites, such as Epinions,[1] Yelp,[2] and RateItAll,[3] feedback and reviews are explicitly solicited within the web interface. Opinion mining techniques can be exploited for the creation and automated upkeep of review and opinion aggregation websites, in which opinions are continuously gathered from the Web and not restricted to just product reviews, but also to wider topics such as political issues and brand perception. Opinion mining and sentiment analysis have also a great potential as sub-component technology for other system. They can enhance the capabilities of customer relationship management (CRM) and recommendation systems allowing, for example, to find out which features customers are particularly interested in or to exclude from the recommendations items that have received very negative feedbacks [21, 22].

Similarly they can be used in email or other types of communication to detect and exclude 'flames', i.e., overly heated or antagonistic language, and to enhance anti-spam systems. Also, online systems that display advertisements as sidebars can use opinion mining techniques to detect web pages that contain sensitive content inappropriate for ads placement [23]. Business intelligence is also one of the main factors behind corporate interest in the field of sentiment analysis [24]. Nowadays,

[1] http://epinions.com

[2] http://yelp.com

[3] http://rateitall.com

Table 1.1 List of most popular companies that are leveraging sentiment analysis tools to track and dissect how consumers feel about products and services of their own and also of the competition

Company	Founded	Headquarters	Web Link
Vocus	1992	USA	http://vocus.com
Kantar	1993	UK	http://www.kantar.com
Cymphony	1996	USA	http://www.cymfony.com
Alterian	1997	USA	http://alterian.com
Factiva	1999	USA	http://dowjones.com/factiva
Brandimensions	2001	Canada	http://brandprotect.com
Attensity	2000	USA	http://attensity.com
Converseon	2001	USA	http://converseon.com
Lithium	2001	USA	http://lithium.com
Lexalytics	2003	USA	http://lexalytics.com
MotiveQuest	2003	USA	http://www.motivequest.com
Visible Technologies	2003	USA	http://visibletechnologies.com
Evolve24	2004	USA	http://evolve24.com
Clarabridge	2005	USA	http://clarabridge.com
Collective Intellect	2005	USA	http://collectiveintellect.com
Radian6	2006	Canada	http://radian6.com
Rapid-I	2006	UK	http://rapid-i.com
Luminoso	2011	USA	http://lumino.so

companies invest more and more money in marketing strategies and they are constantly interested in both collecting and predicting the opinions and the attitudes of the general public towards their products and brands. The design of automatic tools capable to crawl reviews and opinions over the Web in real-time and to create condensed versions of them represents one of the most active research and development area. Several companies, in fact, already provide tools to track public viewpoints on a large scale by offering graphical summarisations of trends and opinions in the blogosphere (Table 1.1).

The development of such systems, moreover, is not only important for commercial purposes, but also for government intelligence applications able to monitor increases in hostile or negative communications [25]. All of these tools, however, are still mainly keyword based and, hence, often fail to meet the gold standards of human annotators. The fundamental aim of this research work is to go beyond such approaches by developing common sense knowledge bases to bridge the cognitive and affective gap between word-level natural language data and the concept-level opinions conveyed by these. Unlike keyword-based methods, sentic computing uses affective ontologies and common sense reasoning tools for a concept-level analysis of natural language text.

Specifically, the ensemble application of graph mining and multi-dimensionality reduction techniques is employed, together with a novel emotion categorisation model, on two common sense knowledge bases, in order to design an open-domain opinion mining engine capable to infer the cognitive and affective information associated with natural language text. Such engine has been exploited for the development of emotion-sensitive systems in fields such as social data mining, multimedia management, personalisation and persuasion, human-computer interaction, intelligent user interfaces, social media marketing, and patient-centred applications.

Evidence of the impact of the approach is found in the presence of sentic computing in high impact factor journals and top AI conferences, and in its adoption by several leading American, British, and Asian companies, including: Zoral Inc., Luminoso Inc., Abies Ltd., Patient Opinion Ltd., Sitekit Solutions Ltd., HP Labs India, and Microsoft Research Asia. For these reasons, sentic computing has also been recently put forward as impact case study to the UK Research Excellence Framework (REF) by the University of Stirling.

1.1.3 Methodology

Relying solely on traditional methods to develop computer systems with a new set of affect-sensitive functionalities is insufficient [26] because today user emotions are still far from being on the radar of computing methods. This is where insights gleaned from a century and a half of scientific study on human emotions can become useful for the development of affect-sensitive interfaces. Despite the extensive literature in emotion research, however, the affective computing literature has been primarily driven by computer scientists and AI researchers who have remained agnostic to the controversies inherent in the underlying psychological theory. Instead, they have focused their efforts on the technical challenges of developing emotion-sensitive computer interfaces. However, ignoring the important debates has significant limitations because a functional affective computing application can never be completely divorced from underlying emotion theory [27].

Blending scientific theories of emotion with the practical engineering goals of analysing sentiments in natural language text and developing affect-sensitive interfaces is one of the main contributions of this book. Recently, many research activities focusing on the extraction of cognitive and affective information from natural language text have gained ground under the umbrella of opinion mining and sentiment analysis. The reason of this trend lies on the ever-growing amount of valuable data available through the Web in the form of news, reviews, blogs, chats, tweets, etc. Sentiment analysis, however, is a multi-faceted and multi-disciplinary problem that requires a deep understanding of natural language. Existing reported solutions and currently available systems are still far from perfect or fail to meet the satisfaction level of the end users. The main issue may be that there are many conceptual rules that govern sentiment and the possibly unlimited clues that can convey these concepts from realisation to verbalisation in the brain.

Recent efforts in this context have been carried about through research works published in reputed conferences through special tracks and workshops, e.g., the *TREC-BLOG* tracks since 2006, the *Sentiment and Subjectivity in Text* workshop in COLING-ACL 2006, the *SemEval 2007 Task#14: Affective Text*, the *TAC 2008 Opin-*

ion Summarisation task, *Emotion, Metaphor, Ontology and Terminology* in LREC 2008, the *Social Data on the Web* (SDoW) workshop from 2008, the *Workshop on Opinion Mining and Sentiment Analysis* (WOMSA) in 2009, *Topic Sentiment Analysis for Mass Opinion Measurement* (TSA) in CIKM 2009, *Computational Approaches to Analysis and Generation of Emotion in Text* in NAACL 2010, the *Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* (WASSA) in ECAI 2010 and in ACL 2011, and the ICDM workshop series on *Sentiment Elicitation from Natural Text for Information Retrieval and Extraction* (SENTIRE) since 2011.

Sentic computing extended the findings expressed in such research works to develop new strategies for open-domain opinion mining, based on the implicit features associated with natural language concepts. The novelty of the approach, in particular, lies in:

1. its multi-disciplinarity (not only computational, but also biologically-inspired and psychologically-motivated);
2. its semantic-based analysis (not only based on word co-occurrence frequencies, but also on the cognitive and affective information associated with natural language);
3. its fine-grained classification (not only at document, page, or paragraph level, but also at sentence and clause level).

In order to evaluate the different facets of the engine from different perspectives, three different resources, namely a Twitter[4] hashtag repository, a LiveJournal[5] database and a PatientOpinion[6] dataset, were used and results obtained using Princeton's WordNet,[7] MIT's ConceptNet[8] and Microsoft's Probase[9] were compared. The first resource is a collection of 3,000 tweets crawled from Bing[10] web repository by exploiting Twitter hashtags as category labels, which is useful to test the engine's target spotting performance. In particular, hashtags about electronics (e.g., IPhone, XBox, Android and Wii), companies (e.g., Apple, Microsoft and Google), countries, cities, operative systems and cars were selected. In order to test the resource's consistency and reliability, a manual evaluation of 100 tweets was performed, which showed that hashtags are accurate to 89 %. The second resource is a 5,000 blogpost database extracted from LiveJournal, a virtual community of more than 23 millions users who keep a blog, journal or diary.

An interesting feature of this website is that bloggers are allowed to label their posts with both a category and a mood tag, by choosing from predefined categories and mood themes. Since the indication of mood tags is optional, posts are likely to

[4] http://twitter.com

[5] http://livejournal.com

[6] http://patientopinion.org.uk

[7] http://wordnet.princeton.edu

[8] http://conceptnet5.media.mit.edu

[9] http://research.microsoft.com/probase

[10] http://bing.com

Chapter 2
Background

The good opinion of mankind,
like the lever of Archimedes,
with the given fulcrum, moves the world.
Thomas Jefferson

The World Wide Web represents one of the most revolutionary applications in the history of computing and human communication, which is keeping on changing how information is disseminated and retrieved, how business is conducted and how people communicate with each other. As the dimension of the Web increases, the technologies used in its development and the services provided to its users are developing constantly. Even if just few years have passed, in fact, Web 1.0's static and read-only HTML pages seem now just an old memory. Today the Web has become a dynamic and interactive reality in which more and more people actively participate by creating, sharing, and consuming contents. In this way, the World Wide Web configures itself not only as a 'Web of data', but also as a 'Web of people' where data and users are interconnected in an unbreakable bond.

This chapter shows how and why online opinions are important in the Web 2.0 era (Sect. 2.1) and illustrates existing approaches and depths of analysis in mining and characterising such opinions (Sect. 2.2). Eventually, the chapter comprises a background section on common sense knowledge representation and reasoning, which later will be further developed and applied to go beyond merely syntactical approaches to sentiment analysis (Sect. 2.3), and some concluding remarks (Sect. 2.4).

2.1 Opinion Mining and Sentiment Analysis

The passage from a read-only to a read-write Web made users more enthusiastic about interacting, sharing, and collaborating through social networks, online communities, blogs, wikis, and other online collaborative media. In the last years, this collective intelligence has spread to many different areas of the Web, with particular focus on fields related to our everyday life such as commerce, tourism, education, and health. The online review of commercial services and products, in particular, is an action that users usually perform with pleasure, to share their opinions about services they

E. Cambria and A. Hussain, *Sentic Computing*, SpringerBriefs in Cognitive Computation,
DOI: 10.1007/978-94-007-5070-8_2, © The Author(s) 2012

have received or products they have just bought, and it constitutes immeasurable value for other potential buyers.

This trend opened new doors to enterprises that want to reinforce their brand and product presence in the market by investing in online advertising and positioning, that is, in social media marketing. The reasons why opinion mining is attracting so much attention from both the academic and the business world, in particular, can be found in the dynamics behind the buzz mechanism (Sect. 2.1.1), in the motivating factors that gave birth to the field (Sect. 2.1.2), and in the sub-tasks that make it different from standard information retrieval (Sect. 2.1.3).

2.1.1 The Buzz Mechanism

What mainly makes social media marketing work is the buzz mechanism [1]. A buzz replicates a message through user-to-user contact, rather than purchasing some advertising or promoting a press release. The message does not have to necessarily deal with the product. Many successful viral campaigns, in fact, have spread thanks to a compelling message, with the company logo included incidentally. At the heart of buzz is an understanding that the natural, spontaneous networks that comprise the social universe are the most effective means of reaching people in a meaningful way. The power of marketing lies, therefore, not in pushing information to the masses, but in effectively tapping those individuals who wield influence over others.

The marketers who are winning are the ones using consumers and culture to their advantage, crafting messages with consumers rather than throwing messages at them. In confirmation of the growing interest in this novel approach to marketing, several academic and commercial tools, e.g., OASYS[1] [2], ESSE [3], Luminoso[2] [4], Factiva,[3] NM Incite,[4] Attensity,[5] and Converseon,[6] have been developed to provide companies (and users) with a way to analyse the blogosphere on a large scale, in order to extract information about the trend of the opinions relative to their products. Nevertheless most of the existing tools and the research efforts are limited to a polarity evaluation or a mood classification according to a very limited set of emotions.
In addition, such methods mainly rely on parts of text in which emotional states are explicitly expressed and, hence, are unable to capture opinions and sentiments that are expressed implicitly.

[1] http://oasys.umiacs.umd.edu/oasys

[2] http://lumino.so

[3] http://dowjones.com/factiva

[4] http://nmincite.com

[5] http://attensity.com

[6] http://converseon.com

(eds.) Storage and Retrieval for Still Image and Video Databases, vol. 2670, pp. 76–87. SPIE, Bellingham (1996)
9. Porkaew, K., Chakrabarti, K.: Query refinement for multimedia similarity retrieval in MARS. In: ACM International Conference on Multimedia, pp. 235–238. ACM, New York (1999).
10. Nakazato, M., Manola, L., Huang, T.: ImageGrouper: search, annotate and organize images by groups. In: Chang, S., Chen, Z., Lee, S. (eds.) Recent Advances in Visual Information Systems, Lecture Notes in Computer Science, vol. 2314, pp. 93–105. Springer, Berlin Heidelberg (2002)
11. O'Hare, N., Lee, H., Cooray, S., Gurrin, C., Jones, G., Malobabic, J., O'Connor, N., Smeaton, A., Uscilowski, B.: MediAssist: using content-based analysis and context to manage personal photo collections. In: CIVR, pp. 529–532. Tempe (2006).
12. Sebe, N., Tian, Q., Loupias, E., Lew, M.S., Huang, T.S.: Evaluation of salient point techniques. In: International Conference on Image and Video Retrieval, pp. 367–377. Springer-Verlag, London (2002).
13. Urban, J., Jose, J.: EGO: a personalized multimedia management and retrieval tool. Int. J. Intell. Syst. **21**(7), 725–745 (2006)
14. Datta, R., Wang, J.: ACQUINE: Aesthetic quality inference engine–real-time automatic rating of photo aesthetics. International Conference on Multimedia Information Retrieval. Philadelphia, In (2010)
15. Bianchi-Berthouze, N.: K-DIME: an affective image filtering system. IEEE Multimed. **10**(3), 103–106 (2003)
16. Smith, J., Chang, S.: An image and video search engine for the world-wide web. Symposium on Electronic Imaging. Science and Technology, In (1997)
17. Frankel, C., Swain, M.J., Athitsos, V.: WebSeer: an image search engine for the world wide web. University of Chicago, Technical Report (1996)
18. Lempel, R., Soffer, A.: PicASHOW: pictorial authority search by hyperlinks on the web. In: WWW. Hong Kong (2001).
19. Jing, F., Wang, C., Yao, Y., Deng, K., Zhang, L., Ma, W.Y.: IGroup: web image search results clustering. In: ACM Multimedia. Santa Barbara (2006).
20. Lieberman, H., Rosenzweig, E., Singh, P.: ARIA: an agent for annotating and retrieving images. IEEE Comput. **34**(7), 57–62 (2001)
21. Chi, P., Lieberman, H.: Intelligent assistance for conversational storytelling using story patterns. IUI. Palo Alto, In (2011)
22. Cambria, E., Hussain, A.: Sentic album: content-, concept-, and context-based online personal photo management system. Cogn, Comput (2012)
23. Lieberman, H., Selker, T.: Out of context: computer systems that adapt to, and learn from, context. IBM Syst. J. **39**(3), 617–632 (2000)
24. Damasio, A.: Descartes' Error: Emotion, Reason, and the Human Brain. Grossett/Putnam, New York (1994)
25. Vesterinen, E.: Affective computing. Digital Media Research Seminar. Helsinki, In (2001)
26. Pantic, M.: Affective computing. In: Encyclopedia of Multimedia Technology and Networking, vol. 1, pp. 8–14. Idea Group Reference (2005).
27. Bonanno, G., Papa, A., O'Neill, K., Westphal, M., Coifman, K.: The importance of being flexible: the ability to enhance and suppress emotional expressions predicts long-term adjustment. Psychol. Sci. **15**, 482–487 (2004)
28. Richards, J., Butler, E., Gross, J.: Emotion regulation in romantic relationships: the cognitive consequences of concealing feelings. J. Soc. Pers. Relatsh. **20**, 599–620 (2003)
29. Burke, A., Heuer, F., Reisberg, D.: Remembering emotional events. Mem. Cogn. **20**, 277–290 (1992)
30. Christianson, S., Loftus, E.: Remembering emotional events: the fate of detailed information. Cogn. Emot. **5**, 81–108 (1991)
31. Wessel, I., Merckelbach, H.: The impact of anxiety on memory for details in spider phobics. Appl. Cogn. Psychol. **11**, 223–231 (1997)
32. Reisberg, D., Heuer, F.: Memory for emotional events. In: Reisberg, D., Hertel, P. (eds.) Memory and Emotion, pp. 3–41. Oxford University Press, New York (2004)