Customer sentiment appraisal from user-generated product reviews: a domain independent heuristic algorithm

Dilip Raghupathi, Bernard Yannou, Romain Farel, Emilie Poirson

Raghupathi D., Yannou B., Farel R., Poirson E. (**2015**) 'Customer sentiment appraisal from user-generated product reviews: A domain independent heuristic algorithm', *IJIDeM: International Journal on Interactive Design and Manufacturing*, doi: 10.1007/s12008-015-0273-4

Customer sentiment appraisal from user-generated product reviews: A domain independent heuristic algorithm

Dilip Raghupathi, Bernard Yannou[1] and Romain Farel, *Ecole Centrale Paris, Laboratoire Genie Industriel, Grande voie des Vignes – 92290 Chatenay-Malabry, France*

Emilie Poirson, *Ecole Centrale de Nantes, IRCCYN, 1 rue de la Noe, BP 92101, 44 321 Nantes Cedex 3, France*

Abstract: Social media give new opportunities in customer survey and market survey for design inspiration with comments posted online by users spontaneously, in an oral-near language, and almost free of biases. Opinion mining techniques are being developed, especially customer sentiment analysis. These techniques are most of the time based on a text parsing and costly learning techniques based on target or domain-dependent corpora for getting a fine understanding of users' preferences. On the contrary, in this paper, we propose an overall sentiment rating algorithm, accurate enough to deliver an overall rating on a product review, without a tedious customization to a product domain or customer polarities. The developed algorithm starts by a text parsing, uses a *Dictionary of Affect Language* to rate the word tree leaves and uses a series of basic heuristics to calculate backward an overall sentiment rating for the review. We validate it on the example of a commercial home theatre system, comparing our automated sentiment predictions with the one of a group of fifteen test subjects, resulting in a satisfactory correlation.

Keywords: user sentiment, sentiment rating, opinion mining, design inspiration, customer opinion, product appraisal, affective judgment

Introduction

To meet the demand of consumers, now very knowledgeable thanks to new numerical technologies, products must be placed on the market extremely quickly. That is become a fundamental rule of innovation [1]. Product designers always welcome feedbacks for the sake of design improvement. Spontaneous comments on new products posted by users or customers in the internet are an incredible source of unbiased information. They are testimonies of individual experiences with product usage, of preferences – complaints, satisfactions - about product features and of the overall appraisal of products. Unbiased feedback has been proven to be extremely hard to obtain. But, spontaneous customer comments on new products remain a valuable source for feedback on design. Resulted data from interviews, questionnaire, surveys and other similar methods suffer from the influence of the test situation [2]. With the rise of Social media, people express themselves without any influence of fear, pressure, intimidation or incentives while giving their opinion. These new media become the centre of attention for analytical purposes, both for industrial and academic research, design analytics for example [3].

A lot of event specific sentiment analyses have been carried out like stock market trends [4]. Real-time geo-localized tweet analysis has shown to develop efficient and inexpensive applications. For example, they

[1] Corresponding author, bernard.yannou@ecp.fr

have been effectively used to adapt the emergency situations in the wake of natural disasters [5]. In the same way, an epidemic can be detected based on a certain tweet trend [6]. The limitation of the use of tweet is its shortness. A consumer quickly limits his/her message to the binary answer of satisfaction or dissatisfaction. To have more explanations on the reasons why the product is liked or disliked, depending on the context of use, the designer has better to use the product online reviews. The product user motive is either to help others buy the product or make sure no one buys the product in future. So a major part of the review would talk about the salient features of a product linked to its method of usage. Analysing such micro blogs or product reviews carefully may provide a lot of details as to how people use it, in which scenarios and whether they are satisfied and happy about its usage values and features.

The domain of opinion mining is recently growing considerably in the literature, especially on sentiment rating of online tweets, reviews and dialogues (see [7] for a literature review). Dong et al [8] showed that the affective judgment of products, design process and people expressed during the design process was important to study. But Wang and Dong [9] showed that if one is interested in developing a sentiment classifier based on Product/Process/People categorization and a specific design domain, then one must devote considerable time and cost towards training the classifier on the target text. In the same manner, Vanrompay et al [10] showed that for extracting user opinions on products or services from spoken dialogues, data must be analysed in a tailored way adapted to user expectations. Cataldi et al [11] confirm that, in analyzing customer online reviews on hostels, a primary computation of customer polarities – they are the most salient features of a product or a service from the user's perspective - is needed to get a precise opinion of an individual customer represented as a word dependency graph, connected through syntactic and semantic dependency relations.

In this study, the mass market orientation view for product design in adapted. In this respect, an objective is be able to find a method to compute globally a set of online reviews, and to produce an overall sentiment rating without important details of individuals' opinions. In other words, this study aims for automatically compute or predict the overall sentiment rating from online reviews, with a good accuracy and without a tedious customization to a product domain or customer polarities. Indeed, in a second step, the study aims to correlate individual overall ratings with consumer data for clustering customer opinions. This is an alternative way of opinion mining which to the knowledge of the authors has not been yet completely explored.

The following section reviews a complementary literature on the user data analysis and the Natural language processing (NLP) method. Section 3 explains our provided framework: the SENTiment Rating ALgorithm (SENTRAL) that is used to rate the user reviews, isolate the usage scenarios, sacrifices and sarcasm into individual entities. Section 4 applies the proposed method on a case study, illustrating the use of SENTRAL on a commercial product. Section 5 goes through the validation procedure where the ratings obtained from our system are compared with those obtained from humans, before concluding in section 6.

Literature review

The notion of interactivity is fundamental in the development cycle of a product. This interaction is of several types: interaction between the expert designer and a digital model or global environment (virtual reality tools [12], intervention in a process of optimization rather than accept the result of a black box even set [13], interaction between several actors of the product development cycle (interactive facilities [14], co-design [15, 16]). For us, interactive design is also a creative activity dedicated to (re-)design products and services. Interactive design is seen as a co-design between user and designer: a participative design. It naturally involves the participation of the user.

Online customers' data analysis

Understanding the customer is a crucial issue for product design. The difficulty of capturing the voice of the customer orally in person can now be compensated with the opinions that customers leave on internet. The

analysis of opinions aims to provide professionals and developers with an overview of the customer experience and ideas that provide clues or evidence for designers to better interpret the voice of the customer [17]. User expressed himself in terms of preferences, which is a personal judgment of the product, often compared to his own experience. A common assumption is that the preference is largely perceptual in nature. According to [18], the perception of a product acts as stimuli on emotions, it is a multi-phase process in which sensation occupy an important role; the product's emotional impact is determined by our feelings in our interaction with the product. Research on consumer behaviour have shown that emotions and emotional states influence their purchasing decision [19,20], It seems thus interesting to consider the sentimental component of perception, determined by our feelings in our interaction with the product.

The first interest of analysis of opinions is to enrich the customer database, very useful in Customer Relationship Management for example [21]. The first domain using online reviews is the marketing to find the strategic goals and identify the customers [22] and customer service [23]. Increasingly, the design sector employs the weblogs and product review to target relevant information for designer [24] and [25]. The freedom given to the online reviewers allows them to express some feelings and sentiments. In public media it plays a big role in the decision making process of the end users [12] and [26], and hence collective sentiment in social media may influence consumer preferences and impact buying decision.

To analyze these online reviews, computer tools like the General Inquirer [27] are essential. Iker [28] proposes a method attempting to reduce the choice "a priori" word classes. After a phase of cutting and cleaning (determiners, prepositions...), the synonymous words are gathered. Sometimes when designers use search engines, they find themselves stuck with a lack of keywords to search. A tool called Tweetspiration [29] was created to provide designers alternative search paths and recommendations from recent twitter trend.Occurrences of the remaining words are calculated and presented as a matrix of correlation between each other. These interactions help to keep the meaning of the text underlining the main topics. In linguistics, POS tagging (Parts-Of-Speech) is the process of marking up a word in a text as corresponding to a particular part of speech based on its definition and context using a software tool [30]. Syntactic analysis can then be used to determine the combinations of words. It may be noticed that in all cases, the structure is similar: (1) Data retrieval and preparation (2) Text processing (3) Analysis.

All this tools are based on grammatical rules and statistical analysis of words and sentences. Halliday's theory [31] is very useful to give an "emotional sense" to the language theoretical analysis. In the recent years, studies where carried, based on Halliday's theory of emotion in language [32]. This study of language of appraisals takes into account the product, the process and the people without rules on interactions between them, thus limited to a non context-of-use oriented analysis.

Natural language processing (NLP)

Textual information in the world can be broadly categorized into two main types: *facts* and *opinions*. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe people's sentiments, appraisals or feelings toward entities, events and their properties [4]. Liu [11] created a model to classify data as subjective and objective. Sentiment analysis, the process of extracting the feelings expressed in a text, is considered as one of the methods of Natural Language Processing (NLP). This is an area of research that involves the use of computers to analyse and manipulate natural language with minimum human intervention for interpretation. In order to construct a program that understands human language, 3 main bases are required [41]. Thought Process, Linguistic representation, World Knowledge.

NLP is carried out in parts starting from word level to understand the Parts of Speech, then to sentence level in order to understand the word order and meaning of the sentence and then the entire text as whole to lift the underlying context.

Chowdary [33] explained that language is understood in 7 interdependent levels by humans and must be integrated in computer programs to replicate it. They are: (1) Phonetic level (2) Morphological level (3)

Lexical level (4) Syntactic level (5) Semantic level (6) Discourse level and Pragmatic level. Phonetics deals with the pronunciation, the smallest parts of a word like suffixes and prefixes are related to the morphology. Lexical level is the parts of speech and syntactic level deals with the structure of the sentence and the order of the words. Meanings of word and sentences are understood at the Semantic level where as knowledge exterior to the document is classified in the pragmatic level. Our system involves 4 of the 7 levels; Morphological, lexical, syntactic and semantic level. Several works had to be studied in order to understand these methodologies.

Though tweets are used for diverse reasons and the context of each tweet is different, they can primarily be grouped into two categories. One category shares personal issues while the other spreads information and creates awareness among the online community [34]. A number of biases are possible while conducting an opinion survey. The most prominent of them all is called the Bradley effect in which the responders are unwilling to provide accurate answers, when they feel such answers may reflect unpopular attitudes or opinions [35]. To overcome this effect, automated polling approaches, known as opinion mining were introduced. These automated polling approaches overcome most of these biases naturally. It was extended to sentiment analysis by Bollen et al. [36] using POMS (Profile of Mood States) and Hu et al. [37] using POS (Parts of Speech).

Methodology

We developed a methodology to analyse the online user review on products, looking forward to deal with the following challenges:

(1) Indicates features a customer is not pleased about
(2) Indicates features a customer is pleased about
(3) Outlines the overall satisfaction/dissatisfaction
(4) Provides keywords of appreciation
(5) Provides keywords of criticism
(6) Evaluate the modes of usage as described by the customer
(7) Detects possibility of sarcasm

The proposed methodology is depicted in Figure 1 and explained in detail as follows.

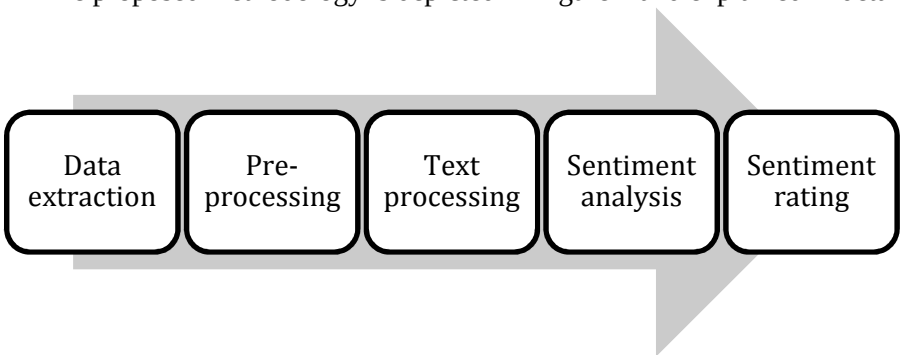


Fig. 1 Process Flow chart

The first step is the extraction of data from website. In step 2 (pre-processing), we carry out the reduction of the noise, classification of words with the aid of Perl script API and Stanford CoreNLP tokenizer. In the third step of Text processing, the noise free data is organised as a tree of dependency from the dependency list obtained with the aid of Stanford Parser and Probabilistic Context Free Grammar (PCFG). Thanks to DAL (Dictionary of Affect Language), the text is word by word analysis for extraction of sentiment in step 4 (sentiment analysis). To complete and evaluate globally the sentiments, we add a list of heuristics which give the sense depending on the context and mode of usage. The final rate is then given in step 5 of SENTRAL algorithm. Each step is described in the following sections.

Extraction of data from website and pre-processing

Data crawling

Three websites are selected to obtain data: Twitter, Amazon and Flipkart. The main reason is the publicly of their data, available with Perl script API's. Basically 2 types of data are obtained: Tweets and User review data. A tweet is a microblog, as shown in Figure 2, limited to 140 characters, containing normal text in addition to targets denoted with a "@" symbol, hash tags (#) to group words from different tweets and smileys (emoticons). Another place to express feelings is a product review on commercial websites without character constraint (example hereafter).

@jcdave The iPhone 5 is a waste of money, you end up paying 200 grand more than any other phone with same features ☹ #apple #dis-appointed	The new sound box by #Bose is an absolute marvel. Crystal clear sound :D I am so happy I decided to invest in this system ☺

Fig. 2 Example of tweets that review a product

Unlike tweets, there is no restriction to the size of a product review. The data are extracted with Perl script API from amazon.com and flipkart.com. A user review consists of the following information: the date of the review, the number of stars or rating in a scale of 0 to 5, the location of the user, the content of the review and also a count of the number users agreeing with the review to eliminate plagiarism and misleading customers.

Data pre-processing

As our objective is to find out the sentiments and usage objectives of the customer, there is a lot of noise in the data that are crawled and hence need to be filtered before it is taken forward in the process. This step is a filtration of the text extracted: each word is categorized thanks to an original list of acronyms (Stanford CoreNLP tokenizer [38, 39]). The tokenizer divides text into a sequences of "token", associated to "word". A table is defined matching each word to its "grammatical class". Every word of the text is assigned to a category. For example, NNP is a singular proper noun, VB is a verb on its basic form, PRP a personal pronoun, RB an adverb. All standard acronyms are expanded using this list and the ones not found in the dictionary are ignored and removed from the sentence. All URLs are removed as they do not help the performance of the system in any way.

The example below illustrates the data pre-processing for the sentence "This product is very good" where one can find a descriptive determiner (ND), a common name (NN), a verb VB2, an adverb RB and an adjective JJ.

Before: `This product is very good`
After: `This/ND product/NN is/VB2 very/RB good/JJ`

Text processing

Parsing and creation of dependency trees

Parsing is the process of breaking down the sentences to words and finding out the grammatical relations between these words. Probabilistic Context Free Grammar (PCFG) is based on the study of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences. A list of dependencies is obtained and a tree is created. This model proposes 55 kinds of possible grammatical dependencies

between words in the English language. A standard dependency is written as: Relation (governor, dependent). For instance, for the sentence “This product is very good”, "This" associated to "product" is a nominal group (NP). "is" is the verbal group (VP) and "very" and "good" is a qualificative group (ADJP). We define grammatical relations defined in a hierarchy so as to arrive at the intended meaning. Using the dependency list and the hierarchy, we are able to create the dependency. The result of the parsing, dependencies and tree is given Figure 3.

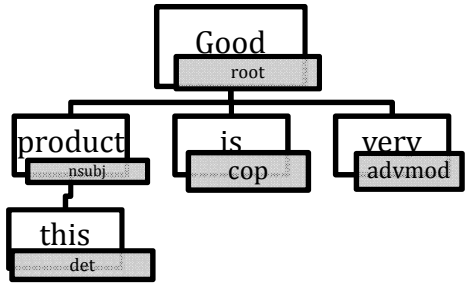
Parsing:	List of dependencies	Dependency tree
`(ROOT (S (NP (DT This) (NN product)) (VP (VBZ is) (ADJP (RB very) (JJ good)))))`	`det(product-2, This-1)` `nsubj(good-5, product-2)` `cop(good-5, is-3)` `advmod(good-5, very-4)` `root(ROOT-0, good-5)`	

Fig. 3 The stages of text processing

Extraction and analysis of the sentiments

Local sentiment analysis with DAL

In the dependency list, the relations are binary in nature. To carry out the process of finding the sentiment rating, we propose the SENTRAL algorithm that uses the Dictionary of Affect Language (DAL). The DAL [40] scores each of the 200,000 English words based on the pleasantness it evokes in the human mind. It is on a scale of 1 to 3 where 1 means the most unpleasant and 3 means the most pleasant. We normalize this score on a scale of 0-1 to suit out algorithm. Table 1 presents some words of tweet with their DAL score. For adjectives, the scores from the DAL can be directly assigned. The meaning of the adjective will change based on the presence of a modifier before or after it. For example, the word “good” and the word-cell “very good” evoke different levels of appreciation.

There are basically 2 types of emotions; good and bad. The emotional guidance system [41] of humans indicates that a person is happy and satisfied if he is in alignment with his requirements. After the dependency tree is created, the words with the tags of *advmod* and *amod* are assigned the pleasantness score by comparing it with the DAL.

Table 1. Example of the pleasantness rating of words in the Dictionary of affect language

Word	DAL Score
Money	0.8889
Phone	0.4375
Waste	0.0000
Marvel	1.0000
Happy	1.0000
Investment	0.7222

Global sentiment rating with our SENTRAL algorithm

The SENTRAL algorithm uses the dependency tree, traversing from the last leaf till the root by progressively evaluating the grammatical relations encountered. To link the dependency tree to the local score given to each word by the DAL, we define 5 heuristics, a priori rules of language.

For each, we will give the idea, illustrated by an example, and we will describe its specification in language analysis.
The four first heuristics concern the AdvMod Tag, Adverbial Modifier. To take into account the effect of an adverb on a noun, we compare the DAL score of the 2 words.
For the governor of the couple, a DAL score less than 0.4 give a negative feeling. The words between 0.4 and 0.55 DAL score are neutral feel words and the words with score greater than 0.55 are said to be positive. The thresholds of 0.4 and 0.55 are being obtained from DAL directly.
For the dependent of the couple, there is not notion of neutrality. Its usage itself leads to boost or to attenuate another word. There is thus only one threshold between negative (<0.4) and positive (>0.4).

After this classification (positive, neutral, negative), we use simple rules of language, explained in table 2.

Table 2: Rules taking into account to study the effects of adverb

Word 1: Adjective	**Word 2: Adverb**	**Combined effect**
Positive *(ex: Good)*	Positive *(ex: Extremely)*	More Positive
Positive *(ex: Good)*	Negative *(ex: Rarely)*	More negative
Negative *(ex: Bad)*	Positive *(ex: Extremely)*	More negative
Negative *(ex: Bad)*	Negative *(ex: Rarely)*	More Positive

Effect of Advmod

For the dependency relation *advmod (adverbial modifier)*, we propose the specific sentiment rating algorithm defining 4 heuristics:

- **heuristic 1**: effect of a positive adverb on a positive adjective.
  The positive sentiment of the adjective will be emphasised by the positive adverb.
  $S_{adverb} > 0.55$ *and* $S_{adj.} > 0.4$ with S_{word} as the DAL score of a word
  $S_{group} = min(S_{adverb} + S_{adverb*} S_{adj.}, 1)$

- **heuristic 2**: effect of a positive adverb on a negative adjective.
  The positive sentiment of the adjective will be attenuated by the negative adverb.
  $S_{adverb} > 0.55$ *and* $S_{adj.} < 0.4$
  $S_{group} = S_{adverb} - S_{adverb*} S_{adj}$

- **heuristic 3**: effect of a negative adverb on a positive adjective.
  The negative sentiment of the adjective will be emphasized by the positive adverb.
  $S_{adverb\ .} < 0.4$ *and* $S_{adj.} > 0.4$
  $S_{group} = max(S_{adj} - S_{adverb*} S_{adj.}, 0)$

- **heuristic 4**: effect of a negative adverb on a negative adjective.
  The negative sentiment of the adjective will be attenuated by the adverb.
  $S_{adverb\ .} < 0.4$ *and* $S_{adj.} < 0.4$
  $S_{group} = S_{adverb} + S_{adverb*} S_{adj}$

Let us take an example: the tag “extremely easy”. Its definition in a sentence is: *advmod(easy-4, very-3)*
DAL scores : $S_{easy} = 0.6665$; $S_{very} = ,0.41665$
$S_{tag} = min(S_{easy} + (S_{easy} * S_{very}), 1)$
$S_{tag} = min(0.6665 + (0.6665*0.41665), 1) = 0,994$

Effect of Amod

The same relation between an adverbial modifier and an adjective is applied to the couple *(Adjectival modifier – Noun)*.

Effect of the ROOT

The third step is to check if the ROOT word's *POS tag is JJ (adjective) or adverb* and the DAL scores are assigned directly. If no such tags are found, it means no sentiment has been expressed and the sentence is ignored, represented by a N/A symbol in the algorithm.

Effect of NEG

Invert all scores of the calculated tags linked by a "*neg*" tag. So if the score of a tag is S_{tag}, and linked to a "*neg*" tag, the new score is $(1\text{-}Score_{tag})$

After this process we have the separate scores of all the related words, sentences and the paragraph. The score of the *jth* sentence is given by eq (1).

$$Sentence_j = \frac{\sum Dependency\ tag_{ij}}{i} \quad \text{(eq. 1)}$$

where "*dependency* tag_{ij}" denotes the score of the i^{th} tag in sentence j.
The score of the entire text is given by eq (2).

$$Sentiment\ score\ = \frac{\sum Sentence_j}{j} \quad \text{(eq. 2)}$$

The words that do not figure in the DAL are ignored since almost all words in the WordNet [42] dictionary are found in this and the probability of a common word missing is very weak. All nouns that have an adjective close to it are grouped together. Negations words like '*not*' '*cannot*' '*shouldn't*' are dealt in such a way that the scores are inverted for the words. For the non-English words, the list of words not found even in the WordNet dictionary is given, with a neutral value of 0.5.

We finally choose a 0-5 scale to globally rate the sentiment of the reviews through our SENTRAL algorithm in order to further compare with customer reviews which are most of the time appraised on such a scale.

Finally, once the score of a sentence calculated, one can consider that the feeling of the customer is approximately given by Table 3.

Table 3. Sentiment score legend

Scores	Conclusion
$0 \leq S_{review} < 2$	Sad and unsatisfied
$2 \leq S_{review} < 3$	Indifferent, happy to use with sacrifices
$3 \leq S_{review} \leq 5$	Happy and satisfied

Case demonstration: reviewing a home theatre

In this section we use the methodology proposed in the previous section to analyse the users review on a commercial home, shown in Figure 4.

In order to demonstrate the SENTRAL sentiment rating algorithm, a general usage product has been selected from an online product provider with an active feedback forum, in form of text and an overall note from 0 to 5. The selected product is a home theatre system (see Figure 4). Fifteen reviews (from different reviewers) are crawled from the feedback forum website (see for instance Figure 5). Here is how the methodology is applied.

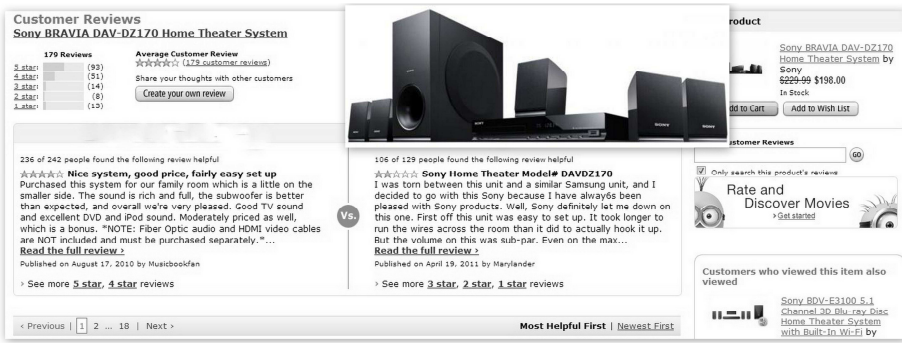


Fig. 4 Reviews of products on Amazon

Amazon Product Code: B003B8VBJ2
Product name: **Sony BRAVIA DAV-DZ170 Home Theatre System (Electronics)**

Review: Well, Sony definitely let me down on this one. First off this unit was easy to set up. It took longer to run the wires across the room than it did to actually hook it up. But the volume on this was sub-par. Even on the max level volume (35) it still wasn't that loud. The main problem was the amount of bass that it produces. The bass is so overpowering that you can barely even hear people talking in the movie, and there is no way to adjust the levels at all.

Fig. 5 Sampled review

Step 1. Data extraction

Extraction of data from website and pre-processing. The 15 comments are extracted and sequenced by sentences. Let us take the example of: "It took longer to run the wires across the room than it did to actually hook it up"

Step 2. Pre-processing

Text processing (organised as tree of dependency). The Stanford Parser is used to establish the dependencies network. For the line "It took longer to run the wires across the room than it did to actually hook it up" It gives: "It/PRP took/VBD longer/RB to/TO run/VB the/DT wires/NNS across/IN the/DT room/NN than/IN it/PRP did/VBD to/TO actually/RB hook/VB it/PRP up/RP ./.".

Step 3. Text processing

A dependency list is obtained from the parser again that arranges words in such a way that all grammatical relationships are established between the words. Following this step, a dependency tree is created as shown in Figure 6.

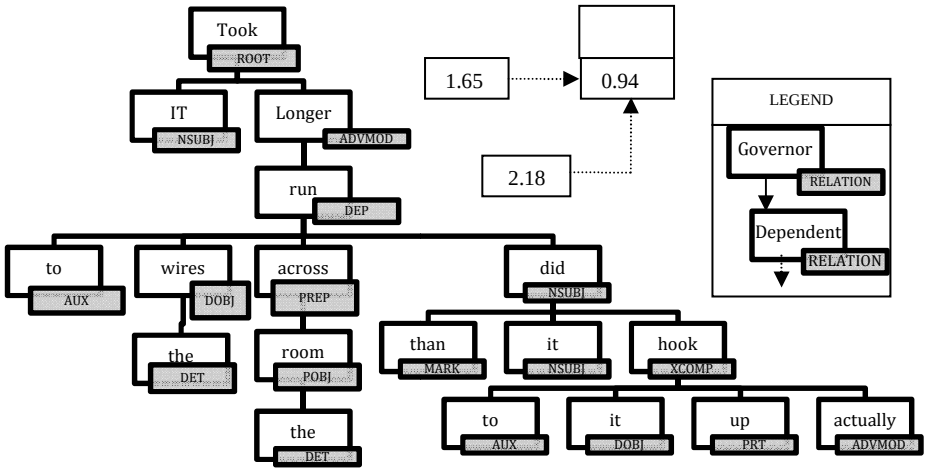


Fig. 6 Dependency tree of a sentence for a technical review on the home theatre system

Step 4. Sentiment analysis

In the dependence tree presented Figure 6, relations containing an ADVMOD or an AMOD are extracted. 2 relations are detected:

- *advmod(hook,actually)*
- *advmod(took,longer)*

Each word that we choose to consider is affected by a DAL score.

- *advmod(hook,actually)* → *advmod(0,55; 0,33)*
- *advmod(took,longer)* → *advmod(0,33; 0,4375)*

Step 5. Sentiment rating

The word "hook" has its individual score : 0.55. But this score is totally independent of the context and the influence of the other words around. In the strategy proposed, we use an heuristic (in this case, heuristic n° 2).

$S_{(hook,\ actually)} = S_{hook} - S_{hook} * S_{actually} = 0{,}55 - 0{,}55 * 0{,}33 = 0{,}37$

The score of the *jth* sentence is given by eq (1), as an average of all the tags. This score is re-scaled on a 0-5 scale (multiplying the sentence score by 5.

The same procedure is carried out for all sentences iteratively (see scores in Table 4) and the score is obtained for the review as whole using equation 3.

Table 4. Sentence-wise scores in the review

Sentence	Score
Well, Sony definitely let me down on this one.	1.016
First off this unit was easy to set up.	2.325
It took longer to run the wires across the room than it did to actually hook it up.	1.39
But the volume on this was sub-par.	N/A
Even on the max level volume (35) it still wasn't that loud.	1.8052
The main problem was the amount of bass that it produces.	N/A
The bass is so overpowering that you can barely even hear people talking in the movie, and there is no way to adjust the levels at all.	1.0675

$$S_{review} = \frac{\sum S_{Sentences}}{Number\ of\ valid\ sentences} = \frac{1.016 + 2.325 + 1.39 + 1.8052 + 1.0675}{5} = 1.52074 \quad \text{(eq. 3)}$$

The total score of emotion found by our algorithm is then 1.52 on a scale of 5.

Validation

The model that we propose basically replaces the human function of understanding and interpreting a text. We propose to validate our model by asking 38 humans to do exactly the same task that our model, i.e. to perform 15 rate reviews on a scale of 0-5. For this, a poll was conducted online and administrated through a google form. A form containing all the fifteen reviews was made public, people were asked to read all the reviews and rate them on this scale based on what their mind evokes about the satisfaction. The question was the following: "*This questionnaire contains reviews about a Home Theatre system written by different users. After reading, please rate these reviews on a scale of 0-5 based on what you feel is the satisfaction level of each of these users. We request your kind patience and to help us with in our research work. Thanks a lot in advance :)*". The 15 reviews all concern true reviews found on internet about home theatre systems. The 38 human subjects have been selected from different gender, age and business areas but all with a satisfactory culture of Hi-Fi devices so as to be sure they understand most of technical descriptions.
The results obtained from the poll are summarized in Table 5.

In this table, each column denotes the number of persons who have voted for that particular rating, 1 being the least satisfied and 5 being the most satisfied based on their inference after reading the reviews. The two distributions of sentiment ratings are given as examples in Figure 7. The scores being well divided (unimodal repartition), the mean is calculated and given in Table 6. The weighted average is then compared with the score obtained from our model in Table 6 to find out the error (difference).

Table 5. Results from the online questionnaire

Rating/ Review	1	2	3	4	5
1	2	0	17	19	0
2	15	18	2	3	0
3	0	0	2	6	30
4	0	2	4	15	17
5	0	3	19	15	1
6	1	1	13	19	4
7	3	11	10	11	3
8	17	13	5	3	0
9	0	6	10	18	4
10	3	15	14	6	0
11	2	3	18	15	0

12

12	0	1	6	17	14
13	0	1	14	21	2
14	0	1	9	15	13
15	17	9	4	7	1

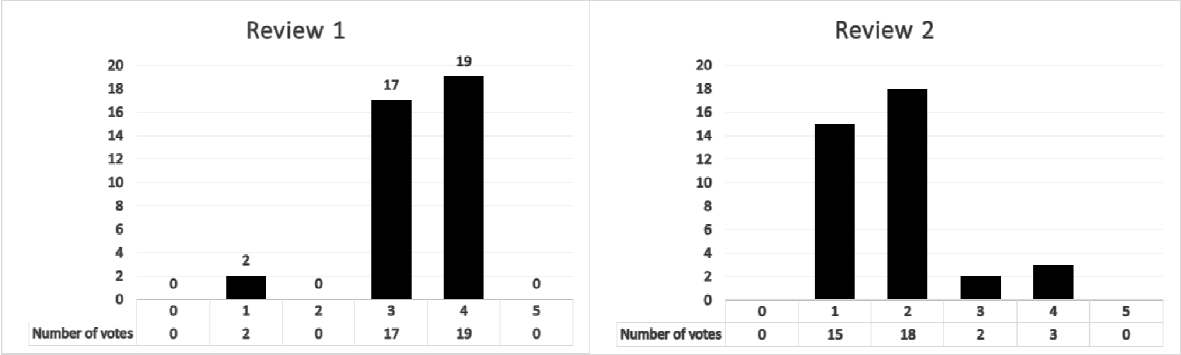


Fig. 7 Distributions of the sentiment ratings of the 38 subjects for reviews 1 and 2.

Table 6. Weighted scores of the votes

Review #	Average	Model's Score	Error	%Error
1	3.39	3.21	0.181	3.6%
2	1.81	1.07	0.748	15%
3	4.73	4.21	0.523	10.5%
4	4.24	4.05	0.187	3.7%
5	3.37	3.33	0.038	0.8%
6	3.63	3.48	0.154	3.1%
7	3	3.46	-0.457	-9.1%
8	1.84	1.88	-0.034	-0.7%
9	3.53	2.86	0.671	13.4%
10	2.61	2.43	0.172	3.5%
11	3.21	3.47	-0.257	-5.1%
12	4.16	3.95	0.212	4.2%
13	3.63	4.65	-1.014	-20.3%
14	4.05	4.21	-0.160	-3.2%
15	2.11	2.10	0.003	0.1%

This error is rather weak (see Table 6 and Figure 8) since the average of errors is 1.3% (over 5 points) and the average of absolute error values is 6.42%.

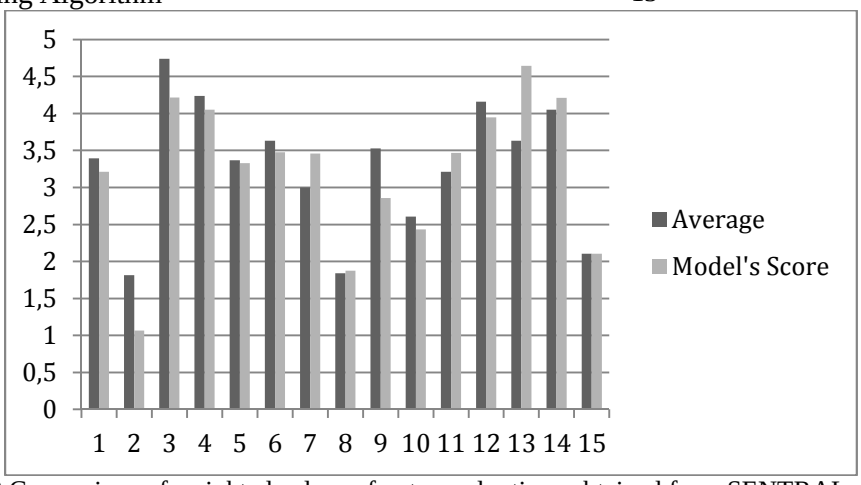


Fig. 8 Comparison of weighted values of votes and ratings obtained from SENTRAL

Human-computer interaction research often involves experiments with human participants to test one or more hypotheses. We use ANOVA (Table 8) to test the hypothesis of whether the difference between results obtained from SENTRAL and the online poll to rate the sentiments (Table 6, columns 2 and 3) are significant (H1) or not (H0).

The ANOVA result is reported as an F-statistic and its associated degrees of freedom and p-value. The individual means for SENTRAL and Human rating were 3.29 and 3.22 respectively. The grand mean for both types of sentiment rating is 3.255. As evident from the means, the difference is only 1.92%. The difference is statistically insignificant with ($F_{1, 28} = 0.034093$, $p > .005$). Hence the null hypothesis H0 was accepted and H1 was rejected, which by extension, validates our model.

Table 7. Student-t test for correlation

Correlation test(student t test)	
Correlation coefficient	0.896425516
tTab	0.063928134
tcal	7.292754614
Correlation	YES

Table 8. ANOVA results

Anova: Single Factor
H0: The difference between SENTRAL's score & human ratings is not significant
H1: The difference is significant.

SUMMARY

Groups	Count	Sum	Average	Variance
Weighted Average obtained from human ranking	15	49.31	3.287333	0.775278
Model's Score	15	48.36	3.224	0.989483

ANOVA

Source of Variation	SS	df	MS	F	P-value
Between Groups	0.03008333	1	0.030083	0.034093	0.85483920
Within Groups	24.7066533	28	0.88238		
Total	24.7367366	29			

ANOVA Result: **F crit = 4.195971819** > F (0.034093) Accept hypothesis H0

Conclusion

Today the user reviews for many products are available online and almost for free. Obtaining feedback from online evaluation of products provides an enormous value for the different services of a company, such as marketing, design, engineering, etc. However, the huge amount of data and the complexity of the analysis limit their usability. This paper is a first step toward automatically analysing user appraisal of products and services with sentiment rating. This analysis is combined with correlating the sentiment rating to data related to customers for clustering their overall opinions. The developed methodology is demonstrated with a case and is evaluated against a sample human rating.

Either conversation in person or expressed in online text form, subjectivity and sentiment add richness to the shared information. Customer's sentiment can easily go beyond facts and rumours and convey unbiased mood, opinion and emotion particularly in online expression. This may bring an immense business value. Listening for brand mentions, complaints and concerns is the first step in social engagement program for any company. Businesses that can listen, could potentially uncover sales opportunities, measure satisfaction, channel reactions to marketing campaigns, detect and respond to competitive threats

An algorithm like SENTRAL, which is domain-independent, can help companies offering a diversity of products and services to save a lot of time in quickly analysing text information from internal and online data sources. Compared to other sentiment analysis models discussed earlier, SENTRAL provides lesser computing complication with a rating algorithm based on simple heuristics. These heuristics in turn are just the mathematical captives of the human process of comprehending a text. This algorithm can be used to find out the global satisfaction of a particular product in the market by comparing the satisfaction scores of similar products. It can possibly be used to find out the trend of a product and to predict its performance in the future as well.

The future improvements in SENTRAL will be on proving the robustness of this domain-independent heuristic algorithm for other categories of products and services, as well as its robustness in terms of the quality of input data: presence of acronyms, typographical errors, ironic and sarcastic expressions. More design oriented works will develop comparison facilities between products of the same category and evolution facilities for studying success propagation and word-of-mouth phenomena.

References

1. Petiot J.-F., Furet B. (2010) Product, process and industrial system: innovative research tracks. *International Journal on Interactive Design and Manufacturing (IJIDeM)* Volume 4, Issue 1, pp 211–213
2. McGue M., Bouchard T. J. (1998). Genetic and environmental influences on human behavorial differences. *Annual review of neurosciences*, 21, 1-24.
3. Lewis K., van Horn D. (2013) Design Analytics In Consumer Product Design: A Simulated Study, ASME International Design Engineering Technical Conferences, Portland, Oregon.
4. Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts stock market. *Journal of Computational Science*, 2(1), 1-6.
5. Caragea C., McNeese N., Jaiswal A., Traylor G., Kim H.W., Mitra P., Wu D., Tapia A.H., Giles L., Jansen B.J. (2011) Classifying text messages for the haiti earthquake. In *Proceedings of the 8th International Conference on In- formation Systems for Crisis Response and Management (ISCRAM2011).*
6. Culotta A. (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In Proceedings of the First Workshop on Social Media Analytics (SOMA '10). ACM, New York, NY, USA, 115-122.
7. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
8. Dong, A., Kleinsmann, M., & Valkenburg, R. (2009). Affect-in-cognition through the language of appraisals. Design studies, 30(2), 138-153.
9. Wang, X., & Dong, A. (2008). A case study of computing appraisals in design text Paper presented at the DCC'08: International Conference on Design Computing and Cognition.

10. Vanrompay, Y., Cataldi, M., Le Glouanec, M., Aufaure, M.-A., & Lamolle, M. (2014). Sentiment Analysis for Dynamic User Preference Inference in Spoken Dialogue Systems. Paper presented at the First Workshop on Semantic Sentiment Analysis (SSA) at ESWC2014.
11. Cataldi, M., Ballatore, A., Tiddi, I., & Aufaure, M.-A. (2013). Good location, terrible food: detecting
12. Weidlich D., Cser L., Polzin T., Cristiano D., Zickner H. (2009) Virtual reality approaches for immersive design, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Volume 3, Issue 2, pp 103-108
13. Bénabès J., Bennis F., Poirson E., Ravaut Y. (2010). Interactive optimization strategies for layout problems. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Volume 4, Issue 3, pp 181-190
14. Mobach, M.P. (2012). Interactive facility management, design and planning. *International Journal on Interactive Design and Manufacturing (IJIDeM)* Volume 6, Issue 4, pp 241-250.
15. Serna L., Merlo C., Zolghadri M., Minel S. (2011) Actors' networks management for design co-ordination. *International Journal on Interactive Design and Manufacturing (IJIDeM)* Volume 5, Issue 1, pp 67-71
16. Giannini F., Monti M., Biondi D., Bonfatti F., Moanari P.D. (2002) A modelling tool for the management of product data in a co-design environment. Computer Aided Design 34, 1063–1073
17. Liu, B. (2010). Sentiment Analysis and Subjectivity. In F. J. N. Indurkhya, *Handbook of Natural Language Processing.* Chicago.
18. Fenech, O.C. and Borg, J.C. (2007). Exploiting Emotions for Successful Product Design. Proceedings of International Conference of Engineering Design ICED'07.
19. Holbrook, M. et Hirshchman, E. (1982). The Experiential Aspects of Consumption: Consumer Fantasies, Feelings and Fun. Journal of Consumer Research, Vol. 9, No. 2, pp. 132-140
20. Richins M. (1997). Measuring Emotions in the Consumption Experience. *Journal of Consumer Research. Vol. 24, No. 2, pp. 127-146*
21. Buttle, F. (2003). Customer relationship management. Butterworth-Heinemann.
22. Berry, M. J., & Linoff, G. (1997). Data mining techniques: For marketing, sales, and customer support. New York, NY, USA: John Wiley & Sons.
23. Bennekom, F. C. V. (2002). Customer surveying: A guidebook for service managers. Customer Service Press.
24. D. Kushal, S. Lawrence, and D. Pennock. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In WWW2003, May 20–24, 2003, Budapest, Hungary.
25. Tucker C., Kim H. (2011) Predicting emerging product design trend by mining publicly available customer review data. In *Proceedings of the 18th International Conference on Engineering Design (ICED11)*, 6, 43–52.
26. OConnor B., Balasubramanyan R., Routledge B.R, Smith N.A. (2010) From tweets to polls: Linking text sentiment to public opinion time series. In Proceedings of the International AAAI Conference on Weblogs and Social Media, 122–129, 2010.
27. Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). The General Inquirer: A Computer Approach to Content Analysis. MIT Press.
28. Iker, H. P. (1974): SELECT: A computer program to identify associationally rich words for content analysis. I. Statistical results , Computers and the Humanity, 8, 313-319.
29. Scarlett R. Herring, Christina M. Poon, Geoffrey A. Balasi, and Brian P. Bailey. (2011) TweetSpiration: leveraging social media for design inspiration. CHI Extended Abstracts, page 2311-2316. ACM, 2011.
30. Nazarenko, A., B. Habert & C. Reynaud (1995): "Open response" surveys: from tagging to syntactic and semantic analysis. In Proceedings of JADT (3rd International Conference on Statistical Analysis of Textual Data), Vol. II, 29-36, Rome, Italy.
31. [Halliday] Halliday, M. A. K. (1985). *An Introduction to Functional Grammar, 1rst Edition*. London: Arnold.
32. Pak, A., Paroubek, P. (2010). Twitter as corpus for sentiment analysis and opinion mining. *LREC conference*, 24-37.
33. Chowdary, G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37, 51-89.
34. Liddy, E. (1998). Enhanced text retrieval using natural language processing. *Bulletin of the American Society for Information Science*, Apr/May 1998, 14-16.
35. Naman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me? Message content in social awareness streams. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 189-192.

36. Bollen J., Mao, H., Pepe, A. (2011). Modelling public mood and sentiment: Twitter Sentiment and Socio-Economic Phenomena. *AAAI conference on weblogs and Media*. Michigan, 450-453.
37. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *SIGKDD*, 168–177.
38. Manning, Klein, D., & D., C. (2003). Accurate unlexicalized parsing. *41st Meeting of the Association for Computational Linguistics*, 423-430.
39. de Marneffe, M.-C., Manning, C. D. (2008). *The Stanford typed dependencies representation*, CrossParser '08 Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation, 1-8, Association for Computational Linguistics Stroudsburg, PA, USA.
40. Whissel, C. (1989). *The dictionary of Affect in Language*. London: Acad Press.
41. Bryne, R. (Director). (2006). *The Secret* [Motion Picture].
42. Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39-41, ACM New-York.