

Additional Information on the Spanish Corpus for Sentiment Analysis towards Brands (second version)

María Navas-Loro¹, Víctor Rodríguez-Doncel¹, Idafen Santana¹, and Alberto Sánchez²

¹ Ontology Engineering Group, Universidad Politécnica de Madrid, Spain

² Havas Media, Madrid, Spain

Taxonomy

The four aspects of interest in the LPS BIGGER³ project for each tweet with respect to the brands, exposed graphically in Fig. 1 are the following:

- **Sentiment Analysis** identifies emotions towards a brand in a post beyond polarity, distinguishing between four sentiments and their direct opposite: *love* and *hate*, *happiness* and *sadness*, *trust* and *fear*, and *satisfaction* and *dissatisfaction*. Also secondary emotions are considered for each of these basic emotions (see Table 1), and a ninth category, called *NC2*, is added for those tweets expressing no feeling. It must be noted that these feelings are not exclusive: the same tweet can be considered *love*, *trust* and *sadness*, or even *satisfaction* and *dissatisfaction*.
- **Purchase Funnel** places the post’s writer within the consumer decision journey, determining whether the Twitter user has purchased a product or is thinking to do it. The opinion given by a customer that already acquired a product gives different information than that provided by a person just thinking about purchasing it: the former might focus on the qualities of the product whereas the latter probably would be related to its promotion. Five different stages are considered: *Awareness* (the user writes about the ad campaign of a brand), *Evaluation* (the user is comparing brands or considering a purchase), *Purchase* (the user is purchasing a product or waiting for its arrival), *Postpurchase* (there is evidence that the user has bought the product) and *Review* (an opinion on the product is given).
- **Marketing Mix** comprises the different marketing strategies the customers can evaluate, also known as the four Ps: *product* (including quality, design and guarantee), *price*, *promotion* (affiliate marketing, sponsorship and advertising) and *place* (point of sale and customer service)[1].
- **Meaningful Brands** a metric that measures the value of the brand, based on the customer’s well being. It is divided in *marketplace* (relating the product to performance such as quality and price), *personal wellbeing* (such as

³ <http://www.cienlpsbigger.es/>

self-esteem, healthy lifestyles, fitness or happiness) and *collective wellbeing* (the role brands play in communities).

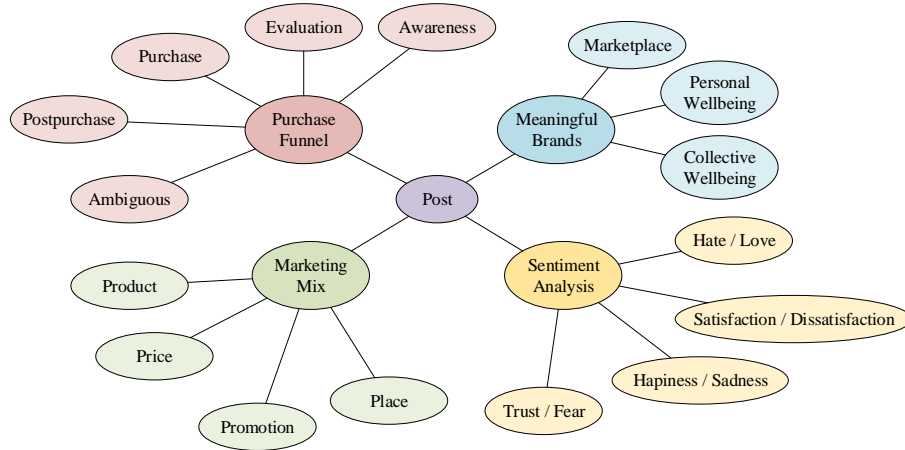


Fig. 1. The four aspects of interest for each tweet.

Tagging criteria

Sentiment Analysis

Three different people intervened in the emotion-tagging process of the first version of the corpus. Main emotions and their related secondary emotions are exposed in Table 1; the brands analyzed, derived from the LPS BIGGER project, can be found in Table 2. The criteria followed is exposed below:

- When the emotion is not obvious, the message must be marked as *NC2* (sometimes the meaning is not evident from the message alone). Messages that are evident advertising by the brand community managers, automatic messages or messages related to brand-events are also tagged as *NC2*.
- If you are sad and you have a *Mahou* (Spanish beer brand), the tag should not be *sadness*. The emotion must be sparked by the brand/product.
- A deep *dissatisfaction* is *hate*, a deep *satisfaction* is *love*. *Fear* is understood as the opposite of *trust*.
- The document “*I am satisfied with X but...*” is tagged as *satisfaction*, *trust* plus whichever negative sentiment follows.
- If a desired product is not found, *satisfaction* and *sadness* are tagged.

Emotion	Related emotions
Trust	Optimism, Hope, Security
Satisfaction	Fulfillment, Contentment
Happiness	Joy, Gladness, Enjoyment, Delight, Amusement, Joviality, Enthusiasm, Jubilation, Pride, Triumph
Love	Passion, Excitement, Euphoria, Ecstasy
Fear	Nervousness, Alarm, Anxiety, Tenseness, Apprehension, Worry, Shock, Fright, Terror, Panic, Hysteria, Mortification
Dissatisfaction	Dislike, Rejection, Revulsion, Disgust, Irritation, Aggravation, Exasperation, Frustration, Annoyance
Sadness	Depression, Defeat, Hopelessness, Unhappiness, Anguish, Sorrow, Agony, Melancholy, Dejection, Loneliness, Humiliation, Shame, Guilt, Regret, Remorse, Disappointment, Alienation, Isolation, Insecurity
Hate	Rage, Fury, Wrath, Envy, Hostility, Ferocity, Bitterness, Resentment, Spite, Contempt, Vengefulness, Jealousy

Table 1. Main emotions and their secondary emotions.

- Love emoticons are tagged as *love* and musical as *happiness* (unless irony happens). *Love* typically implies *happiness*.
- *Happiness* is only given to products already acquired, not to future purchases. The wish for a brand (“*I would like to have X*”) is tagged as *trust* and *satisfaction*, and possibly *love*.
- If the purchase is recurrent, then the document is tagged as *trust*; a future purchase is also typically labeled as *trust*.

Marketing Mix

Each tweet can be labeled as *NC2* or with one or several marketing mix Ps. Criteria for tagging, and examples of each of them, can be found below:

Product This category encompasses texts related to the features of the product (quality, performance, taste...), along with references to design (size, colors, packing, presentation of the product...) or guaranty, such as in the following examples:

Examples

I find the new iPhone too big for my pocket.

Nike Pegasus are not comfortable, to tight!

*I love the new mix Milka Oreo!**

(*) Note that when someone loves/likes something (such as food for instance), we assume it refers to some feature of a product (such as its taste), so we tag it as *Product*.

Sector	Brand
BEBIDAS (BEVERAGES)	Cruzcampo, Heineken, Estrella Galicia, Mahou
AUTOMOCIÓN (AUTOMOTIVE)	Citroën, Fiat, Hyundai, Kia, Peugeot, Toyota
BANCA (BANKING)	Bankia, Bankinter, BBVA, Sabadell, ING, La Caixa/Caixabank, Santander
ALIMENTACIÓN (FOOD)	Auchan, Bimbo, Hacendado, Milka, Pascual, Puleva
RETAIL	Alcampo, Carrefour, Decathlon, Ikea, Leroy Merlin, Mediamarkt, Mercadona
TELCO (TELECOM)	Amena, Lowi, Movistar, Orange, Vodafone, Yoigo
DEPORTES (SPORTS)	Adidas, Nike, Reebok

Table 2. Sectors and Brands in the corpus.

Promotion Texts referring to all the promotions and programs of the brand channeled to increase sales and ensure visibility to their products or the brand, such as advertisements, sponsorships (prizes, competitions, sport teams, events...), special offers, work offers, promotional articles...

Examples

Freaking out with the new 2x1 @Ikea!
La Liga BBVA is the best league in the world.
3x2 is back at @carrefour.

Price Includes economical aspects of a product, such as references to its value or promotions involving discounts or price drops (that must also be tagged as *Promotion*). Examples of texts that should be tagged as *Price* would be the following:

Examples

I'm afraid that I can't afford the new Toyota.
Yesterday I saw the same Adidas for just 40e!
Carrefour sells T-shirts for 3e each... great deal!

Place Aspects related to commercialization, such physical places of distribution of the products (for instance, if a product is difficult to find) and customer service (in every stage of the purchase: information, at the point of sale, postpurchase, technical support...).

Examples

I love the new Milka McFlurry at McDonalds

*Already three malls and unable to find the new Nike Pegasus!
Mad at @Movistar, customer service is terrible... waiting for 10 minutes now*

NC2 Impersonal messages of the brand, news or texts that include none of the aspects mentioned before.

Examples

*Nike is paying no tax!
I can't decide between Puleva and Pascual.*

Purchase Funnel

Each tweet can belong to just one stage in the purchase funnel, be ambiguous or be related to a brand without the author being involved in the purchase (such as is the case of posts of the brand itself). Different phases (and concrete examples for each one) are tagged in the corpus as follows:

Awareness This phase covers the first contact of the client with the brand (either showing a willingness to buy or not). The author tends to comment, criticize or refer to his knowledge of the brand and/or publicity campaigns. They are usually subjective and expressed in first person and tend to mention advertising, videos and publicity campaigns.

Examples

*I just discovered a new BMW ad listening to the radio.
I just loved last Movistar ad.
I like the videos in Nike's YouTube channel.
I liked a video from @Youtube @todoautos youtu.be/mENukeroivs?a Hyundai Santa Fe 2011 - Características y equipamiento.
Citroen ad is fuckin' awesome.*

Evaluation The post implies some research on the brand (questions, seek of confirmation...) or comparison to others (by showing preferences among them, for instance) beyond just knowing it. It does not require previous information on the brand, but some interest in acquiring a product or service. Examples of evaluation would be the following:

Examples

*I prefer Citroen to more expensive brands, such as Mercedes or BMW.
I prefer Nike over Adidas.
I'm looking for a second-hand Kia Sorento in NY, please send me a DM.
BMW is the car of my dreams.*

Purchase There is a direct reference to the moment of a purchase or to a clear intention of purchase (usually in first person). Is less frequent than the other classes, and tends to include synonyms of "purchase" (such as buy, acquire...). Some examples:

*I've finally decided to switch to Movistar.
 I'm buying a Kit Kat in 3,2,1...
 Buying my brand new blue Citroen right now!
 My new phone has just arrived, tomorrow I'll will finally buy it! @Movistar*

Postpurchase Texts referring to a past purchase or to a current experience implying to own or use a product. They are usually expressed in first person, both in present and past. This class presents a special complexity, since interpretation on the same linguistic patterns change depending on the kind of product, as exposed in [2] and exemplified in the sentences below:

- (1) *I like Heineken, the taste is so good.
 I would love a Heineken!
 I need a Heinken!
 I would kill for a Heineken!*
- (2) *I like BMWs, they are so classy!
 I would love a BMW!
 I need a BMW!
 I would kill for a BMW!*

In (1), the client has likely tasted that beer brand before; people does not tend to like or want beverages they have no experience with (at least without mentioning, such as in "*I want to taste the new Heineken.*"). But the same fact is not derived from more expensive items, even when expressed the same way, such as happens in (2): someone can like a car (its appearance, its engine...) without having used it or intending to. This is kind of expressions must be tagged therefore:

- As *ambiguous* for more expensive products of which people tend to give opinion without acquiring them. In our case, this applies just to the *AUTOMOTIVE* sector.
- As *postpurchase* for products that are easier to acquire, and are likely to have been tried before talking about them. This include the sectors *BEVERAGES*, *BANKING*, *FOOD*, *RETAIL*, *TELECOM* and *SPORTS*.

Examples

*We drove there in our Fiat.
 I acquired my Adidas three years ago and they still look new.
 In love with my Nike coat, so warm!
 Space for legs in @Renault Cactus is too little...
 I would love a Kit Kat right now!*

Ambiguous This category includes critical posts, suggestions and recommendations, along with posts where it is not clear in which stage the costumer is (such as the case mentioned above).

Examples

*Do not buy Milka!
 Loving the new Kia!
 Kia Rio is big, like it!*

NC2 Includes impersonal messages without opinions (such as corporative news or responses of the brand to clients), questions implying no personal evaluation or intention (for instance, involving a third person), texts with buy or rental offers with no mention to real use experience...

Examples

2008 Hyundai for sale.

My aunt didn't like the Kia.

Nike Race is comming!

References

1. Neil H Borden. The concept of the marketing mix. *Journal of advertising research*, 4(2):2-7, 1964.
2. Silvia Vázquez, Óscar Muñoz García, Inés Campanella, Marc Poch, Beatriz Fisas, Nuria Bel, and Gloria Andreu. A classification of user-generated content into consumer decision journey stages. *Neural Networks*, 58(Supplement C):68 – 81, 2014. Special Issue on “Affective Neural Networks and Cognitive Learning Systems for Big Data Analysis”.