DUBAI BUSINESS SCHOOL
1st International Conference on Advances in Business, Management and Law (2017) Volume 2017

Analytical Marketing with Collective Perception

Giuseppe D'Aniello
Dept. of Information and Electrical Engineering and Applied Mathematics,
Univeristy of Salerno, Fisciano, Italy. mail: gidaniello@unisa.it

Massimo de Falco
Dept. of Management Sciences - Management and Innovation Systems,
University of Salerno, Fisciano, Italy. mail: mdefalco@unisa.it

Matteo Gaeta
Dept. of Information and Electrical Engineering and Applied Mathematics,
Univeristy of Salerno, Fisciano, Italy. mail: mgaeta@unisa.it

Francesca Loia
Dept. of Management, Sapienza, University of Rome,
Rome, Italy. mail: francesca.loia@uniroma1.it

ABSTRACT

Social networks, forums and blogs are widely considered as a valuable source of information for many applications and in different domains. Being able to extract, analyze and use the knowledge, opinions and sentiments the users share on the Web can become a competitive advantage for any company or organization. Specifically, information about the feelings and the opinions of the users of a Web community with respect to a product or a service can be useful for marketing. In this context, the concept of collective perception is gaining momentum as a way to process, evaluate and quantify the perception and the sentiment that a community of users share about a given phenomenon. In this work, we propose an approach, based on Fuzzy Logic and Sentiment Analysis techniques, which allows to evaluate, also in a quantitative manner, the collective perception of a Web community with respect to a specific product or service.

Keywords: Collective Perception; Analytical Marketing; Fuzzy Logic; Sentiment Analysis.

http://dx.doi.org/10.30585/icabml-cp.v1i1.33

I. INTRODUCTION

One of the most interesting phenomena of the Knowledge Society is the possibility to access, analyze and measure opinions and characteristics of the collectivity by means of the aggregation and the analysis of the many pieces of information that people produce daily via web tools (forums, communities, social web sites, etc.). Such pieces of information represent a trace of the individual and collective behaviors in terms of opinion, relationships, interests, desires, and so on. This is a massive and pervasive phenomenon: the social relationships among individuals are stored in digital repositories; opinions, sentiments desires are stored in the graphs of the social web or in the repositories of search engines; new platforms like Delicious, Flickr, TripAdvisor stimulate and allow the users to easily leave their opinions on the Web and make them instantly available to the general public. Indeed, posts, hashtags, reviews, comments, likes, can all be considered manifestations of opinions, judgements, behaviors, which can be easily gathered, analyzed and used for different aims, with particular reference to the marketing sector. It is well known, indeed, that the number of people that buy products online and use the web in order to search for information about a product is rapidly increasing [1]. Specifically, the consumer product reviews, also referred as electronic word of mouth, have drawn much attention from both academics and the public as one of the most influential sources people rely on when making a purchase decision [1][2]. The valuable information contained in the customers' reviews is not useful only for other potential customers, but it represents a gold mine for marketing analysts and producers, since it gives an insight about what the customer likes or dislikes of a specific product, and it can reveal also possible latent defects of the product. Such kind of information can be used for different kind of applications. First of all, it is possible to discover the strong features that satisfy a majority of customers' needs as well as weak features that are undesirable to most customers. The ability to automatically identify successful and failing products along with their strong and weak product features could enable designers to refine next generation product designs prior to launch [3]. It can be also useful to support the post-sell customer service, helping in the identification of what people do not like about the product. Moreover, it can drive the definition of successful marketing campaigns that can be better tailored according to the users' opinions about the product and its features.

In this work, we propose a novel approach that is based on Fuzzy Logic and Sentiment Analysis for the analysis of product reviews created by the users of a web community. The objective is to define an approach able to process a set of reviews and extract from them the so called *collective perception* [4], which represents a quantitative indicator of the perception (in terms of positive or negative sentiment with a product or its feature) that a community of users has of a product.

The approach adopts the concept of *User Signature* proposed in [5], which is a fuzzy set representing the tagging activities of a user on a web site. We extend this concept to the context of products' reviews and by integrating it with a sentiment analysis technique which is able to characterize the user in term of its perception with respect to the product, according to the published review. The concept of collective perception based on the *User Signature* has been introduced in [4] and [6] in the context of Smart City and Smart Tourism Destination.

Therefore, the main contributions of this paper are:

- The extension of the concept of User Signature, User Perception and Collective Perception to the context of analytical marketing;
- An approach to analyze customers' reviews in order to identify the User Perception and Collective Perception and to use them to support marketing strategies;
- The definition of different mechanisms to define groups of users, according to different criteria, on which it is possible to compute the collective perception and the similarity with other groups or with single users. Indeed, the approach proposed in [5], although it proposes some measures of similarity that involves groups of users, it does not provide the criteria by which this groups can be defined. In the marketing domain, instead, it is really important the process of identification of different groups of users that can be treated with a same approach.

A case study conducted on a real web community, which contains reviews about the Samsung Galaxy S4, is proposed in this paper in order to demonstrate the capabilities of the proposed approach.

II. OVERALL APPROACH

In this work we propose an approach to the quantitative evaluation of the perception that a user, or a community of users, may have of a product or with respect to a specific feature of the product.

Specifically, the approach, starting from reviews, comments or posts regarding the product we want to analyze, gathered from a web site or a social web application, it is able to quantify the perception (in terms of its positive or negative sentiment) of the product. Such an approach is helpful for different analytical marketing applications, as for instance:

- Identification of the customers' opinions regarding a specific feature of the product;

- Identification of similar users with respect to their opinions about the product (or its feature)
- Identification of defects of the product according to the users
- Assessment of the variation of the trend of the perception about the product during the time
- Supporting the competitive intelligence **[9]**, i.e., monitoring the competitive environment to identify strengths and weaknesses of competitors' products or with respect to competitive brands.

The approach is complementary to other traditional approaches of customers' satisfaction analysis or sentiment analysis. It is useful to rapidly evaluate and estimate the perception of a community of users and is can drive further and deeper analysis using other techniques and approaches **[8]**.

The approach is depicted in Figure 1. Starting from a web community in which there are reviews and comments about the products, we extract the tags and keywords that represent the opinion of the user about the product. Such tags are related to particular feature of the product (e.g., "bad display", "good battery"). We construct a 3-D matrix containing the users, the tags they used and the features to which these tags refer. This matrix is useful to compute the User Signature, which is a fuzzy representation of the activities of the user with respect to his/her review about the product. The User Signature is the starting point to compute two measures: the User Perception which gives an insight about the positive or negative perception of the user with respect to the features or to the product, and the User Kindredness, which is a measure of similarity between users according to their activities and opinions. By aggregating the different User Signatures, it is possible to define different groups of users. We propose different mechanisms and strategies to define such groups. A group of users can be analyzed in order to obtain a measure of the Collective Perception and of the similarity between a user and a group (Group Kindredness). Such measures can be useful for a marketing analyst or a producer in order to understand the perceptions of the community about their product.

Next subsections provides further details about the approach.

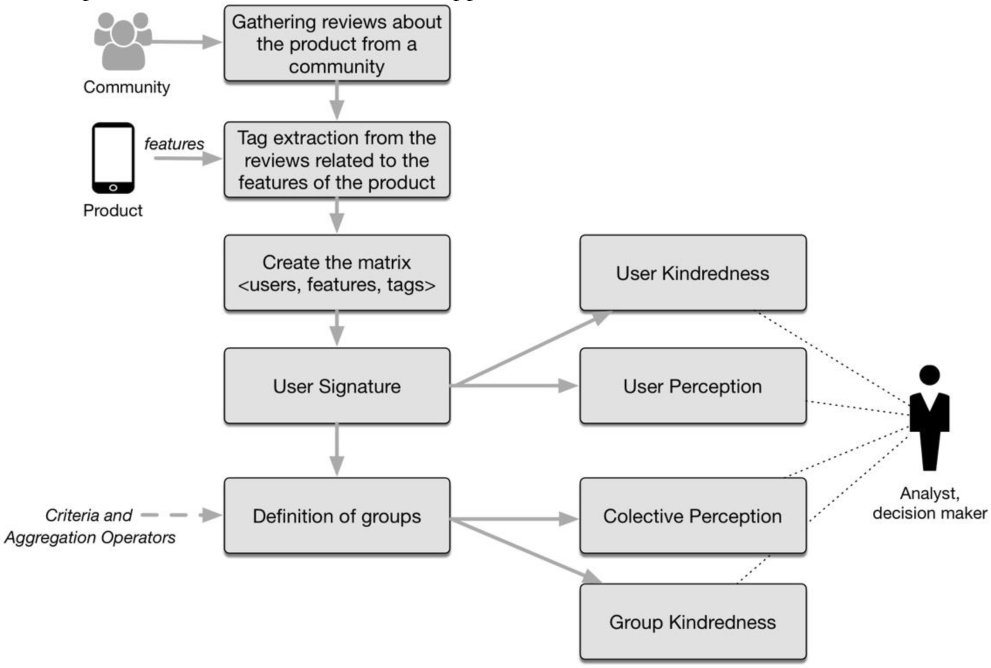

Figure 1. Overall Approach

A. User Signature and User Kindredness

Let us consider a product P on which we want to perform the analysis of the collective perception following the proposed approach. The product P is characterized by a set $F^P = \{f_1^P, f_2^P, \dots, f_m^P\}$ of m features (or characteristics or aspects of the product). For instance, if the product P is a smartphone, its features could be $F^P = \{battery, memory, display, \dots\}$. Moreover, let C be the community of users of whom we want to analyze the perception of the product P. This community C consists of a set $U^C = \{u_1^C, u_2^C, \dots, u_l^C\}$ of l users. Such users express their opinions about the product (and its features) by means of comments, reviews, tweets, or any other kind of textual message (in the following, we refer to this message as *opinion* without loss of generality) that, for instance, can be published on a forum, on a social web site, on a blog, etc.

Let be $O^P = \{o^P_{u_1^C,1}, o^P_{u_1^C,2}, \dots, o^P_{u_i^C,s}, \dots, o^P_{u_q^C,1}, \dots, o^P_{u_q^C,r}\}$, the set of opinions expressed by the users of the community C about the product P, in which $o^P_{u_i^C,s}$ is the *s-th* opinion of user u_i from the community C for the product P. Indeed, notice that each user $u_i^C \in U$ can publish more than one review in the community and that not all the users

publish a review, i.e., $q \leq l$. Each opinion published by a user contains some keywords, adjectives, expressions or tags that express their sentiment and perception about the product P or about a specific feature f_j^P of the product. In what follows, we refer to such keywords, adjectives and expressions as *tags*. Let $T = \{t_1, \ldots, t_n\}$ be the set of *n* tags extracted from the opinions of the users of the community *C*.
We can define a relation $Z = U^C \times F^P \times T$. An element of such relation $\langle u_i^C, f_j^P, t_k \rangle$ indicates that it exists an opinion $o^P_{u_i^C,s} \in O^P$ in which the user $u_i^C \in U^C$ has used the tag $t_k \in T$ to express its idea or perception about the feature $f_j^P \in F^P$ of the product *P*.
The relation *Z* can be represented with a 3-D matrix with the following dimensions: users U^C, features F^P and tags T. If the user u_i^C used the tag t_k to describe the feature f_j^P, then the point in the 3-D matrix $\langle U^C, F^P, T \rangle$ with coordinates $\langle i, j, k \rangle$ is marked [5].
We can look at this matrix from different point of views, depending on the kind of analysis we want to perform. For instance, we can select a specific feature f_j^P thus to extract a slice of the 3-D matrix, obtaining a 2-D matrix with two dimensions, U^C and T, containing the information about the perception of the selected feature according to the community *C*. In what follows, we focus on the two dimensions F^P and *T*, by selecting the information related to a single user. This allows us to consider the perception of each single user according to the different features of the product we want to analyze.
According to the work of Yager and Reformat [5], it is possible to define some measures that gives a rapid insight on the perception and the interests of the user. If we consider the number of tags used for each feature, we obtain a "tag per feature" measure. Similarly, we can consider the frequency of each tag, if we consider how many times it is used for the different features. Interestingly, Yager and Reformat propose to consider these two measures in terms of fuzzy sets. Indeed we can consider the number of tags for each feature as an indicator of the attractiveness of such feature, as the higher is this number, the more attractive is the feature for the user. Thus, we define the fuzzy set "Feature Attractiveness" (*FeatAttract*), for which the membership degree represents the degree of interest of the user for that feature:

$$\boldsymbol{FeatAttract}_{u_i^C}(\boldsymbol{f^P}) = \left\{\frac{\boldsymbol{b_1}}{\boldsymbol{f_1^P}}, \frac{\boldsymbol{b_2}}{\boldsymbol{f_2^P}}, \ldots, \frac{\boldsymbol{b_j}}{\boldsymbol{f_j^P}}, \ldots, \frac{\boldsymbol{b_m}}{\boldsymbol{f_m^P}}\right\} \qquad (1)$$

with $b_j = \frac{\#\text{ of tags used for } f_j^P \textit{ and user } u_i}{\max \#\, of \text{ different tags used for a single feature by } u_i}$.
In the same way, for the frequency of each tag, it is possible to define the fuzzy set "Tag Popularity" (*TagPop*):

$$\boldsymbol{TagPop}_{u_i^C}(\boldsymbol{t}) = \left\{\frac{\boldsymbol{a_1}}{\boldsymbol{t_1}}, \frac{\boldsymbol{a_2}}{\boldsymbol{t_2}}, \ldots, \frac{\boldsymbol{a_k}}{\boldsymbol{t_k}}, \ldots, \frac{\boldsymbol{a_n}}{\boldsymbol{t_n}}\right\} \qquad (2)$$

with $a_k = \frac{\#\text{ of times } \textit{ is used by } u_i}{\max \#\, of \text{ resources tagged by } u_i \textit{ with a single tag}}$.
The higher is the membership of *TagPop*, the more popular is the tag for that user.
These two fuzzy sets, considered separately, are able to describe from two different perspectives, the activity and the interest of the user with respect to the analyzed product. In order to simultaneously analyze both the aspects of the user's opinions about the product, it is possible to define the fuzzy relation *UserSignature* (*US*) that represent the user herself [5]:

$$\begin{aligned}&\mathbf{UserSignature}_{\mathbf{u_i^C}}(\mathbf{f^P}, \mathbf{t})\\&= \mathbf{FeatAttract}_{\mathbf{u_i^C}}(\mathbf{f^P})\ \mathbf{X}\ \mathbf{TagPop}_{\mathbf{u_i^C}}(\mathbf{t})\end{aligned} \qquad (3)$$

and, by using the *min* operator, the value of the relation for a single tag t_k and a feature f_j^P is:

$$\begin{aligned}&\mathbf{UserSignature}_{\mathbf{u_i^C}}(\mathbf{f_j^P}, \mathbf{t_k})\\&= \mathbf{min}\left\{\mathbf{FeatAttract}_{\mathbf{u_i^C}}(\mathbf{f_j^P}), \mathbf{TagPop}_{\mathbf{u_i^C}}(\mathbf{t_k})\right\}\end{aligned} \qquad (4)$$

High value of the relation indicates the features in which the user has interest and tags (which express its opinion) he/she used often. For instance, we can consider the subset of the most frequent used tags and the most commented features. To do so, we consider the *alpha-cut* ($\alpha - cut$) operation of a fuzzy set. Given the user signature US_i of the user u_i, its $\alpha - cut$ is the crisp set of all the pairs (f_j^P, t_k) whose membership values $US_i(f_j^P, t_k)$ are greater than or equal to α, with $\alpha \in [0,1]$. Indeed, the $\alpha - cut$ of the User Signature extracts all the pairs *(feature, tag)* that are more frequent. In this way, we can obtain a stricter characterization of the user with greater values of α because we consider only the tags that are more used and the features that are more commented by the user.
The *UserSignature* is useful for different applications. It can be used to identify which is the features in which the users are most (or less) interested into, or to evaluate the similarity between users according to their interests and opinions. Indeed, it is possible to compute a value of similarity between users according to their *UserSignatures*.

Let US_1 and US_2 be the fuzzy sets representing the *UserSignatures* of user u_1^C and u_2^C, respectively. The similarity between these two sets is defined as *kindredness (K):*

$$\boldsymbol{K(US_1, US_2) = \frac{|US_1 \cap US_2|}{|US_1|}} \qquad (5)$$

where $|\cdot|$ is the cardinality of the set while the intersection of the two fuzzy sets can be computed by using a T-norm (generally the *min* operator is used). The kindredness represents commonality between User Signatures in the reference to the User Signature of the first user we are analyzing (notice that the kindredness is not symmetric).

B. Perception of the User

The *User Signature* is useful to evaluate the interests of the users with respect to the different features of the products (or for different products). But it does not give information about the perception that the users have of the different features, which means that we do not know if the interest of a given user in a feature is positive (e.g., “the display is wonderful”) or negative (e.g., “the battery is not good enough”). It is self-evident that such kind of information is crucial in the context of analytical marketing. Consequently, the proposed approach use a sentiment analysis technique in order to understand if the perception of each tag (i.e., opinion) related to a feature is positive or negative.

Given a tag $t_k \in T^C$, it is possible to evaluate the sentiment convoyed by such tag in terms of how much objective, positive and negative a term is. Specifically, we indicate with $Pos(t_k) \in [0.0,1.0]$ the value of positiveness, with $Neg(t_k) \in [0.0,1.0]$ the value of negativeness and with $Obj(t_k) \in [0.0,1.0]$ the value of objectiveness of the tag $t_k \in T^C$. Moreover, $Pos(t_k) + Neg(t_k) + Obj(t_k) = 1.0$ [10].

In order to obtain a single value for the collective perception, which is useful to evaluate and compare different features, products or users, also during different time intervals, first we need to define a unique value which is able to synthetize the value of positiveness, negativeness and objectiveness of a tag. Consequently, we define the Tag Perception $TP(t_k)$ in the following way:

$$\boldsymbol{TP(t_k)}: [\boldsymbol{Pos(t_k)} \times \boldsymbol{Neg(t_k)} \times \boldsymbol{Obj(t_k)}] \rightarrow [\boldsymbol{0.0, 1.0}] \qquad (6)$$

Specifically, we use the Fuzzy Inference System (FIS) shown in Figure 1 to calculate $TP(t_k)$. In the Figure, P represents the fuzzy membership of $Pos(t_k)$, N of $Neg(t_k)$ and O of $Obj(t_k)$. Table 1 shows some of the rules implemented in the FIS to achieve the Tag Perception.

Table 1.
IF-THEN Rules with OBJ=Low of the Fuzzy Inference System

POS	OBJ	NEG	TP
H	L	L	VP
H	L	M	P
H	L	H	N
M	L	L	P
M	L	M	N
M	L	H	B
L	L	L	N
L	L	M	B
L	L	H	VB

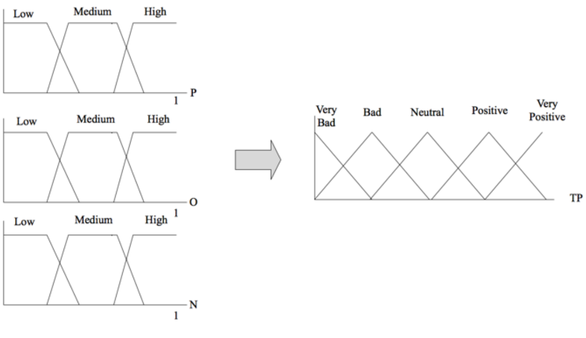

Figure 2. Fuzzy Inference System for computing the Tag Perception

The value of a given tag perception depends only on the tag and on the way the FIS is defined, while it is completely independent form the user and the features to which it refers.
In this work, we use SentiWordNet [10], a lexical resource which allows to associate to each word *w* (specifically, to each *synset* included in WordNet [11]), the three values for positiveness POS(w), negativeness NEG(w) and objectiveness OBJ(w). These represent the input for the FIS in order to compute the value for TP(w).
The same approach used to define the Tag Perception can be applied to other elements to obtain useful insights on the opinions of the users.
Let $T_{f_j^P} = \{t_1 \dots t_r\} \subseteq T$ the subset of tags a given user $u_i \in U^C$ has used to express his/her opinion on the feature $f_j^P \in F^P$. It is possible to define the *Feature Perception* $FP_{u_i}(f_j^P)$ of the feature $f_j^P \in F^P$ for the user $u_i \in U^C$ in the following way:

$$FP_{u_i}(f_j^P) = \varphi(TP(t_k)), \forall t_k \in T_{f_j^P} \qquad (7)$$

where φ is an aggregation operator (e.g., average, weighted average, median, mode, etc.). This measure can be useful to evaluate which is the perception of the user with respect to a single feature of the product we are analyzing, or to compare the opinions of two or more users with respect to the feature f_j^P.
Lastly, it is possible to consider the matrix $\langle F^P, T\rangle$ related to the user $u_i \in U^C$ to give a characterization of the perception of the user with respect to the product. In this case we can consider all the tags used by the user for all the features, or just a subset of them by using the $\alpha - cut$ operation. Considering only the tags of the crisp set obtained by performing the $\alpha - cut$, we obtain a bag of words which have been used by the user for describing the product. On this set of tags we define the *User Perception* $UP_\alpha(u_i)$, which is a measure of the perception the user has of the product, in the following way:

$$UP_\alpha(u_i) = \varphi(TP(t_k)) | t_k \in T, f_j^P \in F^P, US_i(f_j^P, t_k) \geq \alpha \qquad (8)$$

where φ is an aggregation operator.

C. Group Signature and Collective Perception

The proposed approach is also useful to evaluate the perception of a group of users. Analyzing groups of users allows to identify users that share the same opinions on some features or on the whole product. This enables the possibility to employ specific marketing techniques and to propose specific offerings according to the characteristics of the considered group.
Given a group of users $G \subseteq U^C$, we define the Group Signature GS_G as a way to represent the whole group of users by aggregating their User Signature:

$$\mathrm{GS_G(f^P, t)} = \psi_{\mathrm{u_i \in G}}\left(\mathrm{US_i(f^P, t)}\right) \qquad (9)$$

where ψ is an aggregation operator. An Ordered Weighted Average (OWA) [7] [12] operator can be used for the aggregation operator ψ. Indeed, an OWA operator allows to obtain a finer control on the aggregation of the signatures by using different linguistic quantifiers, obtaining more detailed or coarser grained representations of the group.
Having a representation of a group of users with the Group Signature, it is possible to evaluate how much a user is similar (or compatible) with the group. We define the User Group Kindredness:

$$UGK(u_i, G) = \frac{|US_i(\mathbf{f^P}, \mathbf{t}) \cap GS_G(\mathbf{f^P}, \mathbf{t})|}{|US_i(\mathbf{f^P}, \mathbf{t})|} \qquad (10)$$

We extend the definition of *Feature Perception* and *User Perception* with reference to a group G. We define the Collective Feature Perception as the perception of a feature of the product according to a group of user:

$$CFP_G(f_j^P) = \theta\left(FP_{u_i}(f_j^P)\right), \forall u_i \in G \qquad (11)$$

where θ is an aggregation operator.
The *Collective Perception* (or Group Perception) represents the opinion of a set of users with respect to a product:

$$CP_\alpha(G) = \theta(UP_\alpha(u_i)), \forall u_i \in G \qquad (12)$$

D. Group selection mechanisms

In previous section, we defined a way to represent a group of users and to evaluate their perception. The group G can be defined by selecting different users according to different criteria. The way by which we define the group have a great impact on the analysis of the perception regarding a product or its feature. For instance, we can select all the users that have a bad opinion of the product, in order to understand the motivations of their perception.
In this section, we propose different ways to define the groups according to different criteria which uses the above defined measures.

Group defined by Tag Perception

The Tag Perception (TP) defines a value indicating the sentiment of a given tag. The higher is TP, the more positive is the tag. Accordingly, it is possible define a group by selecting only the users that have used at least one tag whose TP value is higher (or lower) than a given threshold γ_{TP}:

$$G_{>\gamma_{TP}} = \{u_i^C \in U^C \big| \exists \langle u_i^C, f_j^P, t_k \rangle \in (U^c \times F^P \times T) \ \wedge TP(tk) > \gamma_{TP}\} \quad (13)$$

The group $G_{>\gamma_{TP}}$ contains the users that have used at least one tag with a Tag Perception greater than a given threshold, in order to select all the users that have at least a positive opinion about the product. Similarly, we can define the group $G_{<\gamma_{TP}}$ to select all the users that have used at least a tag with a Tag Perception smaller than a given threshold, usually to identify all the users that have used at least a negative tag with respect to the product.

Group defined by Feature Perception

In order to support the analysis related to a specific feature of the product, it can be useful to define different groups according to the perception the users have of each feature. Accordingly, we define the group by considering the values of the Feature Perception $FP_{u_i}\left(f_j^P\right)$ and by setting a threshold γ_{FP}:

$$G_{>\gamma_{FP}} = \{u_i^C \in U^C \big| \exists \langle u_i^C, f_j^P, t_k \rangle \in (U^c \times F^P \times T) \ \wedge FP_{u_i}\left(f_j^P\right) > \gamma_{FP}\} \quad (14)$$

In this way, we can select all the users that share a positive opinion on a given feature. This gives also the opportunity to compare their perception on the specific feature with the perception of the whole product. Similarly, we can also define groups that have a bad perception of a feature, by considering the group $G_{<\gamma_{FP}}$ in which we took only the users whose Feature Perception is below the threshold.

Group defined by Kindredness

The kindredness between two users expresses a measure of their similarity according to the tags they use and the features in which they are interested into. A group can be defined for containing the users that have a high (or low) level of kindredness, thus to select all the users that are very similar (or dissimilar). Given a threshold $\gamma_K \in [0,1]$, we define the group in the following way:

$$G_{>\gamma_K} = \{u_i^C \in U^C \big| \exists \, u_j \in U^C \wedge K\left(US_i, US_j\right) > \gamma_K\} \quad (15)$$

III. CASE STUDY: SMARTPHONE

In this section, we provide an illustrative example of application of the proposed approach for the analysis of the perception of users and groups regarding a specific product. In this case study, we focus on the smartphone market, due to the proliferation of forums and communities of users on the Web wherein there are many reviews and comments about a same product. Moreover, the feedbacks and judgements of the customers have a big impact on the success of the brand as they strongly influence other potential customers, and thus the producers are really interested in collecting and analyzing such comments, even also to further develop and improve their next models of smartphone.

A. Data

The case study is about the smartphone Samsung Galaxy S4. First, we need to define the features of the smartphone that we want to analyze and use to compare the opinion of the users. We refer to the work of Hu and Liao [13] which proposes a set of features of a smartphone, identified by means of an approach based on the Analytical Hierarchy Process. Using this work, we identified 6 main features of a smartphone:

1. Body (cover material, weight, style design)
2. Platform (CPU, operating system, build-in memory, e-mail service, personal information management, word processing)
3. Autonomy (battery life)
4. Camera&Sound (photo functions, mobile TV, multimedia, sound recording)

5. Connectivity (high speed internet access, GPS)
6. Display (touch screen, screen size, quality screen)

The community of users that we select is GSMArena[1]. We choose this community because it has many detailed reviews, it is independent from a specific brand, and it provides many technical specifications about the smartphones.

From all the available reviews about the Samsung Galaxy S4, we select 201 reviews from April 2013 to October 2016, with more than 150 words each. From the selected reviews, we need to extract the tags in order to construct the 3-D matrix $\langle U^C, F^P, T \rangle$. We use a semi-automatic approach to extract the adjectives and expressions (e.g., predictive nominals) that are related to the features we want to analyze. Specifically, we use the CONCEPTUM system [14] in order to identify a set of concepts and synonyms related to the features we have identified. Next, we scan the text of each reviews in order to identify the presence of a reference to the smartphone or its features (i.e., by using the concepts and synonyms identified by the CONCEPTUM system). Then, we manually check and analyze the tags extracted by such approach. Figure 3 shows an example of a review about the Samsung Galaxy S4 from GSMArena, in which the tags have been highlighted.

In Figure 4 we report an example of the tags used for each feature by some users of the community.

*"I've been using this phone for 4 months now, here's a quick review. I find the **battery life** to be **excellent** when using applications (heavy apps as well)/gaming(fifa14, tapped out etc)/camera+video/music/talk time/standby, its only drawback is when you start heavy browsing it drains really quickly, no idea why but it's something i can live with. A fully charged phone lasts me an entire day (9-9) which i find quite good for a heavy user like me. I don't understand why people complain so much about the design, i feel its sleek, and for that **beautiful** **5inch screen size** the **design** fits wonderfully. I used both the white and black colour, i like the white one slightly better cause it's more striking and the glossy **plastic** looks **good** imo. Also spigen has come up with some fantastic cases (the best one being the one with the s view and the spigen hard case) which basically render any people who complain about its design, useless. **Performance** **wise**, its a joy to play with and use. All **applications** run **smooth**, and the multi task option is awesome, u can listen to youtube videos and browse at the same time, which makes me very happy. My laptop usage has dropped considerably, and the credit for that goes to this monster of a phone that basically does almost everything my laptop does at an incredible pace and it fits in my pocket. if you are looking to buy a new phone honestly look no further than this **beautiful** beast. You can watch movies in hd and its **colors** are **delightful**. Its also got an **awesome** **camera** with some nifty features to boast, and the **video** recording is **awesome**. Shame that it lacks ios which would've made it just perfect. Hopefully s5 will get that. Cons - battery life while browsing - lack of ios - slightly worried about the region lock (although i've been told it shouldn't be a problem)"*

Figure 3. Example of a review about the Samsung Galaxy 4. In blue the concepts related to the smartphone, in red the extracted tags.

User	Body	Platform	Autonomy	Camera&Sound	Connectivity	Display
1		perfect				
2		small			malfunctioning	
3		decent	degraded			
4			problematic			
5	worst	worst	worst	worst, worst	worst	worst
6						problematic
7		superfluous				
8		unrespoonsive				
9						unsteady
10		unrespoonsive	inflamed			
11		incorrect				
12	liked		inflamed	better		

Figure 4. Example of the tags used by some users to comment the features of the smartphone.

B. Method

User Signature and User Kindredness

The data gathered by the GSMArena community are organized in the 3-D matrix $\langle U^C, F^P, T \rangle$ with dimension $201 \times 6 \times 89$. To perform the analysis on single users, we extract a 2-D matrix for each user (of dimensions 6x89). As an example, Figure 5 shows the tags extracted from the review of the user u_18 related to each feature. From this set of tags, we construct the User Signature of user u_18 as shown in Figure 6. In this figure, on the columns there are the tags represented with an ID: in this case, tag_22=plastic, tag_23= durable; tag_24=light. On the rows, there are the 6 features: Autonomy, Body, Camera&Sound, Connectivity; Display, Platform. Each cell shows the value of the UserSignature according to eq. *(4)*.

[1] www.gsmarena.com

User	Body	Platform	Autonomy	Camera&Sound	Connectivity	Display
18	plastic,durable, light, slim	unrespoonsive, unsmoooth	inflamed,rapid		better	

Figure 5. Tags extracted from the review of user u_{18}.

	19	20	21	22	23	24	25	26	27	28	29
1	0	0	0	0	0	0	0	0	0.5000	0	0
2	0	0	0	1	1	1	0	0	0	1	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0.5000

Figure 6. Excerpt of the User Signature of user u_{18}

Considering all the User Signatures it is possible to evaluate the value of user kindredness between each pair of users. This can be useful to identify the users that share similar opinions about the features of the smartphone as they basically use the same tags for the same features. We report the values of the user kindredness between two users in the matrix $\langle U^C \times U^C \rangle$ of Figure 7. The cell (u_i, u_j) reports the value of the kindredness between user u_i and u_j according to the eq. *(5*).

	57	58	59	60	61	62	63
44	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0
46	0	0	0.1111	0	0	0	0
47	0	0	0	0	0	0	0
48	0	0	0.5000	0	0	0	0
49	0	1	0	0	1	0	0.5000
50	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0
53	0	0	0.6667	0	0	0	0
54	0	0	0	0	0	0	0
55	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0
57	0	0	0.2857	0.2143	0	0	0
58	0	0	0	0	1	0	0.5000
59	0.4444	0	0	0	0	0	0
60	0.1538	0	0	0	0	0	0

Figure 7. Excerpt of the matrix reporting the user kindredness between pairs of users.

Higher values of kindredness indicate pairs of users that potentially share the same opinion about the product (even if we do not know if such opinion is positive or negative until we consider the User Perception measure). For instance, in Figure 8 are highlighted the pairs of users (u_81, u_96) and (u_101, u_91) that have a value of kindredness equal to 1.

	90	91	92	93	94	95	96	97	98	99	100	101	102
80	0	0	0	0	0	0	0	0	0	0	0	0	0
81	1	0	0.5000	0	0	0	1	0	0	1	0	0	0
82	0	0	0	0	0	0	0	0	0	0	0	0	0
83	0.5000	0	0.2500	0	0	0	0.5000	0	0	0.5000	0	0	0
84	0	0	0	0	0	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0	0	0	0	0	0
86	0	1	0.1250	0	0	0	0	1	0	0	0	0.5000	0
87	0	1	0.1250	0	0	0	0	1	0	0	0	0.5000	0
88	0	1	0.1250	0	0	0	0	1	0	0	0	0.5000	0
89	0.5000	0	0.2500	0	0.5000	0.5000	0.5000	0	0	0.5000	0	0	0
90	0	0	0.2500	0	0.5000	0	0.5000	0	0	0.5000	0	0	0
91	0	0	0.1250	0	0	0	0	1	0	0	0	0.5000	0
92	0.1667	0.0833	0	0	0	0	0.1667	0.0833	0	0.1667	0.1667	0.0833	0
93	0	0	0	0	0	0	0	0	0	0	0	0	0
94	0.3333	0	0	0	0	0.3333	0	0	0.1667	0	0	0	0
95	0	0	0	0	0.5000	0	0	0	0	0	0	0	0
96	1	0	0.5000	0	0	0	0	0	0	1	0	0	0
97	0	0.5714	0.0714	0	0	0	0	0	0	0	0	0.2857	0
98	0	0	0	0	0.1667	0	0	0	0	0	0	0	0
99	0.1905	0	0.0952	0	0	0	0.1905	0	0	0	0	0	0
100	0	0	0.5000	0	0	0	0	0	0	0	0	0	0
101	0	1	0.2500	0	0	0	0	1	0	0	0	0	0
102	0	0	0	0	0	0	0	0	0	0	0	0	0
103	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 8. Excerpt of the matrix of kindredness. Some pairs of users have a value of kindredness equals to 1. Notice that the User Kindredness measure is not symmetric.

Observing Figure 9, we notice that the user u_{81} and u_{96} use the same tag ("better") for the feature platform. Due to the fact that they do not use any other tags, their kindredness is equal to 1.

User	Body	Platform	Autonomy	Camera&Sound	Connectivity	Display
80		poor, best				
81		better				
82		cheap				
83		better, unsmoooth				
84		horrible				
85		customisable				
86			inflamed, dead			
87			dead, inflamed			
88			inflamed, dead			
89		better				better
90		better		better		
91			dead, inflamed			
92		better, smooth, unimproove	dead		smooth	
93		malfunctioning	swelled			
94		functional	removable	better		better
95		great				better
96		better				
97	cheap	extendible	inflamed, dead	horrible		

Figure 9. Tags used by the users (on the rows) for the different features (on the columns). Users u_{81} and u_{96} use the same tag for the same feature.

A particular case happens when the tags used by a user are included in the tags used by another one, as highlighted in Figure 10. In such cases, due to the fact that the User Kindredness is not symmetric, the value of kindredness for u_{101} with respect to u_{91} is $K(u_{101}, u_{96}) = 1$, while $K(u_{96}, u_{101}) = 0.5$.

User	Body	Platform	Autonomy	Camera&Sound	Connectivity	Display
90		better		better		
91			dead, inflamed			
92		better, smooth, unimprooved, dim	dead		smooth	
93		malfunctioning	swelled			
94		functional	removable	better		better
95		great				better
96		better				
97	cheap	extendible	inflamed, dead	horrible		
98	plastic, slim	problematic	removable			
99		better, accurate, manageable, transferable		good		liked, unreal
100		dim				
101			dead			
102			normal			
103		weird				

Figure 10. Users u_96 and u_101 share a tag for the feature autonomy.

The $\alpha - cut$ operation can be applied also on the user kindredness. In this case, for instance, the analyst may easily find all the pairs of users that have a kindredness above a specific threshold, thus to identify all the users that basically share the same opinions.

User Perception and Group Perception

The computation of the user signatures and the evaluation of user kindredness is useful to identify like-minded users that thus may be analyzed and treated as similar, but do not give any insights on the fact that the user are similar because they like or are disappointed by the smartphone or some of its features. This kind of analysis can be realized by considering the user perception and the group perception.

Figure 11 shows the values of the Tag Perception (TP) for some of the tags extracted from the reviews, computed as indicated in eq. *(6)* by means of the Fuzzy Inference System.

	P	O	N	PT
unreal	0,04	0,575	0,385	0,276
awesome	0,875	0	0,125	0,92
dead	0,08	0,418	0,502	0,25
malfunctioning	0	0,25	0,75	0,173
plastic	0,091	0,909	0	0,5
durable	0	1	0	0,5
light	0,092	0,711	0,197	0,5
normal	0,08	0,66	0,26	0,416
extendible	0,625	0,375	0	0,753
rapid	0,167	0,833	0	0,5
slim	0,333	0,625	0,042	0,659
unsmooth	0	0,625	0,375	0,29
best	0,75	0,25	0	0,827
cool	0,291	0,587	0,122	0,616
friendly	0,33	0,53	0,14	0,656
weird	0,167	0,333	0,5	0,25
cracked	0	0,795	0,205	0,428
nice	0,75	0,209	0,041	0,827
fine	0,37	0,617	0,013	0,703
wonderful	0,75	0,25	0	0,827
superb	0,875	0,125	0	0,92
dim	0,173	0,349	0,478	0,25
excellent	1	0	0	1
poor	0,051	0,413	0,536	0,25

Figure 11. Tag Perception.

It is possible to define different groups according to a threshold value for the TP by using eq. *(13)*. Such groups will contain all the users that have used at least a tag with a TP greater (or lower) than the threshold.

In this illustrative example, we consider instead the values of Feature Perception (FP) thus to obtain group of users that have a similar perception of a given feature. For instance, we can set a threshold $\gamma_{FP} = 0.75$ for the feature $f_2^P = Platform$, and by using eq. *(14)*, we obtain a group of user that have a good perception of the platform of the Samsung Galaxy S4. Such group contains 23 users. Specifically, by considering the tags used by the users of Figure 9 and the values of PT of Figure 11, the users u_{81}, u_{95}, u_{96} belongs to the group $G_{>\gamma_{FP}}$, while users u_{80}, u_{84}, u_{92} belongs to $G_{<\gamma_{FP}}$ for the feature *Platform*.

The identification of such groups is helpful for the marketing analyst, as he/she can, for instance, propose some special offerings to the users that do not like a specific feature, in order to not lose a customer. Moreover, having specific marketing actions for treating each group, allows also to use the correct actions for new customers that are similar to the group. In this case, we can compute the kindredness between a user and the group with the eq. *(10)*.

IV. CONCLUSIONS, IMPLICATIONS AND LIMITATIONS OF THE APPROACH

The proposed approach allows to evaluate the perception the users has regarding specific features of a product by analyzing their comments and reviews. The integration of Sentiment Analysis technique with the concept of User Signature V, allows to process, at least at a high level of abstraction, the perception of users in a community regarding specific features of a product. Moreover, the capability of creating groups of similar users in order to treat them according to different marketing strategies can be really helpful for marketing analysts.

From a strategic point of view, the different parameters that can be used in the proposed approaches (e.g., $\alpha - cuts$, groups, different fuzzy sets, etc.) allows to perform different kind of analysis with different levels of details, according to the needs of the decision maker. The identification of users that have a positive perception of some features is an important feedback for the producer, which can thus focus on such features of the product that should be improved. Moreover, the capability of identifying groups of users with a negative perception, drives the analysts in the identification of the real causes and possible defects of the products that thus can improve the development of new products. These groups can also become the target of some marketing strategies which propose them a different product with can best fit the users' needs.

The case study shows that the approach is able to give an insight on the similarity between the opinions of different users. Indeed, we have manually analyzed the content of each pair of reviews with a high value of kindredness, thus to verify the reliability and efficacy of the proposed measure.

What emerges from the analysis is also that some posts are very similar, if not equals in terms of content and adopted tags. This may be due to the fact that a restricted number of users is able to influence the opinions of other people (the so-called influencers). The presence of such kind of users may have a great impact on the results of the analysis by means of the proposed approach.

In future works, we will propose other approaches to define the groups of users. Moreover, we will conduct further experimentations on a wider audience. Furthermore, from a technological point of view, we will work in order to improve the phases of text analysis and tags/keywords extraction from the reviews. Lastly, we will consider also the time dimension in the analysis of the users' perception.

V. REFERENCES

[1] Eun-Ju Lee, Soo Yun Shin, *When do consumers buy online product reviews? Effects of review quality, product type, and reviewer's photo*, In Computers in Human Behavior, Volume 31, 2014, Pages 356-366, ISSN 0747-5632

[2] Hu, N., Pavlou, P. A., & Zhang, J. (2006). *Can online reviews reveal a product's true quality?: Empirical findings and analytical modeling of online word-of-mouth communication*. In J. Feigenbaum, J. Chuang, & D. M. Pennock (Eds.), Proceedings of the 7th ACM conference on electronic commerce (pp. 324–330).

[3] Tuarob S, Tucker CS. Quantifying Product Favorability and Extracting Notable Product Features Using Large Scale Social Media Data. ASME. J. Comput. Inf. Sci. Eng. 2015;15(3):031003-031003-12. doi:10.1115/1.4029562.

[4] G. D'Aniello, A. Gaeta, M. Gaeta, V. Loia and M. Z. Reformat, "Collective awareness in Smart City with Fuzzy Cognitive Maps and Fuzzy sets," 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, 2016, pp. 1554-1561. doi: 10.1109/FUZZ-IEEE.2016.7737875

[5] R. R. Yager and M. Z. Reformat, "Looking for like-minded individuals in social networks using tagging and e fuzzy sets," IEEE Transactions on Fuzzy Systems, vol. 21, no. 4, pp. 672–687, 2013.

[6] G. D'Aniello, M. Gaeta and M. Z. Reformat, "Collective Perception in Smart Tourism Destinations with Rough Sets," 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), Exeter, 2017, pp. 1-6. doi: 10.1109/CYBConf.2017.7985765

[7] Yager, R. R. and J. Kacprzyk. (1997). *The Ordered Weighted Averaging Operators: Theory and Applications*. Norwell, MA: Kluwer.

[8] Wu He, Harris Wu, Gongjun Yan, Vasudeva Akula, Jiancheng Shen, A novel social media competitive analytics framework with sentiment benchmarks, Information & Management, Volume 52, Issue 7, 2015, Pages 801-812, ISSN 0378-7206, http://dx.doi.org/10.1016/j.im.2015.04.006.

[9] L. Kahaner, Competitive Intelligence: How to Gather, Analyze and Use Information to Move Your Business to the Top, Touchstone, New York, NY, 1998.

[10] Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In LREC (Vol. 10, pp. 2200-2204).

[11] George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41

[12] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," Systems, Man and Cybernetics, IEEE Transactions on, vol. 18, no. 1, pp. 183–190, 1988.

[13] Hu, Y. C., & Liao, Y. L. (2013, January). Utilizing analytic hierarchy process to analyze consumers' purchase evaluation factors of smartphones. In Proceedings of World Academy of Science, Engineering and Technology (No. 78, p. 1029). World Academy of Science, Engineering and Technology (WASET).

[14] D. Toti and M. Rinelli, "On the road to speedreading and fast learning with CONCEPTUM," in Proceedings - 2016 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2016, 2016, pp. 1–6.