



Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality

Erik Cambria^{a,*}, Tim Benson^b, Chris Eckl^c, Amir Hussain^d

^a Temasek Labs, National University of Singapore, 117411, Singapore

^b Abies Ltd., Berkshire, RG18 9WL, United Kingdom

^c Sitekit Solutions Ltd., Portree, IV51 9HL, United Kingdom

^d COSIPRA Lab, University of Stirling, FK9 4LA, United Kingdom

ARTICLE INFO

Keywords:

AI
E-Health
Natural language processing
Opinion mining
Sentiment analysis

ABSTRACT

Barriers to use health related quality of life measuring systems include the time needed to complete the forms and the need for staff to be trained to understand the results. An ideal system of health assessment needs to be clinically useful, timely, sensitive to change, culturally sensitive, low burden, low cost, involving for the patient and built into standard procedures. A new generation of short and easy-to-use tools to monitor patient outcomes on a regular basis has been recently proposed. These tools are quick, effective and easy to understand, as they are very structured and rigid. Such structuredness, however, leaves no space to those patients who would like to say something more. Patients, in fact, are usually willing to express their opinions and feelings in free text, rather than simply filling in a questionnaire, for either speaking out their satisfaction or for cathartic complaining. *Sentic PROMs* allow patients to evaluate their health status and experience in a semi-structured way and accordingly aggregate input data by means of sentic computing, while tracking patients' physio-emotional sensitivity.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Public health measures such as better nutrition, greater access to medical care, improved sanitation and more widespread immunization, have produced a rapid decline in death rates in all age groups. Since there is no corresponding decline in birth rates, however, the average age of population is increasing exponentially. If we want health services to keep up with such monotonic growth, we need to automatize as much as possible the way patients access the health-care system, in order to improve both its service quality and timeliness. Everything we do that does not provide benefit to patients or their families, in fact, is waste.

To this end, a new generation of short and easy-to-use tools to monitor patient outcomes and experience on a regular basis has been recently proposed by Benson et al. (2010). Such tools are quick, effective and easy to understand, as they are very structured. However, they leave no space to those patients who would like to say something more.

Patients, in fact, are usually keen on expressing their opinions and feelings in free text, especially if driven by particularly positive or negative emotions. They are often happy to share their

health-care experiences for different reasons, e.g., because they seek for a sense of togetherness in adversity, because they benefited from others' opinions and want to give back to the community, for cathartic complaining, for supporting a service they really like, because it is a way to express themselves, because they think their opinions are important for others. When people have a strong feeling about a specific service they tried, they feel like speaking it out. If they loved it, they want others to enjoy it. If they hated it, they want to warn others away.

Standard patient reported outcome measures (PROMs) allow patients to easily and efficiently measure their health related quality of life (HRQoL) but, at the same time, they limit patients' capability and will to express their opinions about particular aspects of the health-care service that could be improved or important facets of their current health status. The framework developed within this work, in turn, exploits the ensemble application of standard PROMs and sentic computing (Cambria and Hussain, 2012), a novel approach to opinion mining and sentiment analysis, to allow patients to evaluate their health status and experience in a semi-structured way, i.e., both through a fixed questionnaire and through free text.

The structure of the paper is as follows: Section 2 provides some background about HRQoL measurement, Section 3 explains in detail sentic computing tools and techniques adopted in this work, Section 4 illustrates the processes for the extraction of cognitive and affective information from patient opinions, Section 5 shows how such information can be exploited for monitoring patients'

* Corresponding author. Tel.: +65 65165752.

E-mail addresses: cambria@nus.edu.sg (E. Cambria), tim.benson@abies.co.uk (T. Benson), chris.eckl@sitekit.net (C. Eckl), ahu@cs.stir.ac.uk (A. Hussain).
URL: <http://sentic.net> (E. Cambria).

physio-emotional sensitivity, Section 6 presents a preliminary evaluation of the system and Section 7 comprises concluding remarks and a description of future work.

2. Related work

In health-care, it has long been recognized that, although the health professional is the expert in diagnosing, offering help and giving support in managing a clinical condition, the patient is the expert in living with that condition. Next-generation patients are central to understanding the effectiveness and efficiency of services and how they can be improved. PROMs provide a means of gaining an insight into the way patients perceive their health and the impact that treatments or adjustments to lifestyle have on their quality of life.

Pioneered by Donabedian (1966), health status research began during the late 1960s with works focusing on health-care evaluation and resource allocation. In particular, early works mainly aimed to value health states for policy and economic evaluation of health-care programmes, but devoted little attention to the practicalities of data collection (Culyer et al., 1971; Fanshel and Bush, 1970; Torrance et al., 1972). Later works, in turn, aimed to develop lengthy health profiles to be completed by patients, leading to the term patient reported outcome (Bergner et al., 1976; Ware, 1976). PROMs can provide a new category of real-time health information, which enables every level of the health service to focus on continuously improving those things that really matter to patients. The benefits of routine measurement of HRQoL include helping to screen for problems, promoting patient-centric care, aiding patients and doctors to take decisions, improving communication amongst multi-disciplinary teams and monitoring progress of individual or groups of patients and the quality of care in a population. However, in spite of demonstrated benefits, routine HRQoL assessment in day-to-day practice remains rare as few patients are willing to spend the time needed to daily fill-in questionnaires, such as SF-36 (Ware and Sherbourne, 1992), SF-12 (Ware et al., 1996), Euroqol EQ-5D (Brooks, 1996) or the Health Utilities Index (Horsman, Furlong, Feeny, & Torrance, 2003).

To overcome this problem, *howRu*, a new generic PROM was recently proposed by Benson et al. (2010) for recording the level of each patient's physical and mental symptoms, limitations and dependency on four simple levels. The questionnaire was designed to take no more than a few seconds using electronic data collection and integration with electronic patient records as part of other routine tasks that patients have to do, such as booking appointments, checking in on arrival at clinic, or ordering or collecting repeat medication. The main aim of *howRu* is to use simple terms and descriptions, in order to reduce the risk of ambiguity and to ensure that as many people as possible could use the measure reliably and consistently without training or support.

The same approach has been employed to monitor also patient experience (*howRwe*) and staff satisfaction (*howRus*) on a regular basis. These questionnaires have been proved to be quick, effective and easy to understand, as they are short, rigid and structured. However, such structuredness can be very limiting, as it leaves no space to those patients who would like to say something more about their health or the service they are receiving. Patients, especially when driven by particularly positive or negative emotions, do want to express their opinions and feelings. *Sentic PROMs* allow patients to assess their health status and health-care experience in a semi-structured way by enriching the functionalities of the new PROM tools with the possibility of adding free text (Fig. 1).

This way, when patients are happy with simply filling-in the questionnaire, they can just leave the input text box blank but, when they feel like speaking out their opinions and feelings, e.g.,

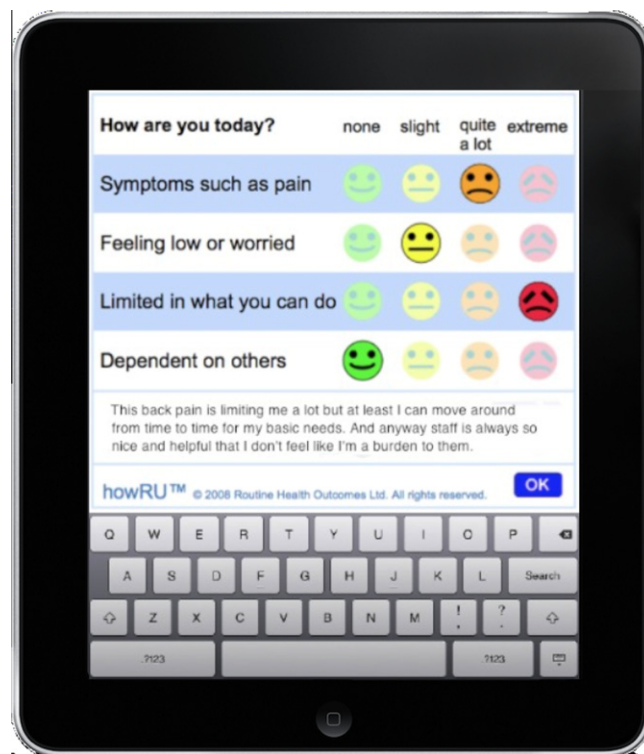


Fig. 1. *Sentic PROMs* prototype on iPad. The new interface allows patients to assess their health status and health-care experience both in a structured (questionnaire) and unstructured (free text) way.

in the occasion of a particularly positive or negative situation or event, they can now do it in their own words. Hence, *Sentic PROMs* input data, although very similar at concept level, are on two completely different structure levels – structured (questionnaire selection) and unstructured (natural language). As we would like to extract meaningful information from such data, the final aim of *Sentic PROMs* is to format the unstructured input and accordingly aggregate it with the structured data, in order to perform statistical analysis and pattern recognition. In particular, the gap between unstructured and structured data is bridged by means of sentic computing.

3. Sentic computing

Existing approaches to automatic identification and extraction of opinions and sentiments from text can be grouped into three main categories: keyword spotting (Elliott, 1992; Ortony et al., 1988; Wiebe et al., 2005), in which text is classified into categories based on the presence of fairly unambiguous affect words, lexical affinity (Rao and Ravichandran, 2009; Somasundaran et al., 2008; Stevenson et al., 2007; Wilson et al., 2005), which assigns arbitrary words a probabilistic affinity for a particular opinion or emotion, and statistical methods (Abbasi et al., 2008; Goertzel et al., 2000; Hu and Liu, 2004; Pang and Lee, 2005; Pang and Lee, 2002; Turney and Littman, 2003), which calculate the valence of keywords, punctuation and word co-occurrence frequencies on the base of a large training corpus. The problem with such approaches is that they mainly rely on parts of text in which opinions are explicitly expressed such as positive terms (e.g., good, nice, excellent, fortunate, correct, superior, best) and negative terms (e.g., bad, nasty, poor, unfortunate, wrong, inferior, worst). In general, in fact, opinions are expressed implicitly through context and domain dependent concepts, which make keyword-based approaches extremely ineffective.

Sentic computing is a multi-disciplinary approach to sentiment analysis that exploits both computer and social sciences to better recognize, interpret and process sentiments in natural language. In sentic computing, whose term derives from the Latin *sentire* (root of words such as sentiment and sentience) and *sensus* (intended both as capability of feeling and as common sense), the analysis of natural language is based on affective ontologies (Cambria, Havasi, & Hussain, 2012; Cambria, Grassi, Hussain, & Havasi, 2011) and common sense reasoning tools (Cambria, Hussain, Havasi, & Eckl, 2009; Cambria, Song, Wang, & Hussain, 2011), which enable the analysis of text not only at document, page or paragraph level but also at sentence and clause level.

Specifically, sentic computing involves the use of AI and Semantic Web techniques, for knowledge representation and inference; mathematics, for carrying out tasks such as graph mining and multi-dimensionality reduction; linguistics, for discourse analysis and pragmatics; psychology, for cognitive and affective modeling; sociology, for understanding social network dynamics and social influence; finally ethics, for understanding related issues about the nature of mind and the creation of emotional machines.

In this work, we exploit sentic computing tools and techniques to extract the semantics and sentics (i.e., the cognitive and affective information) associated with patient opinions and, hence, bridge the gap between the structuredness of questionnaire data and the unstructuredness of natural language. In particular, for the extraction of semantics, we use the following sentic computing tools and techniques:

1. A directed graph representation of common sense knowledge (Section 3.1).
2. A statistical method for the identification of common semantics (Section 3.2).
3. A technique that expands semantics through spreading activation (Section 3.3).

In turn, for the extraction of sentics, we use:

1. A language visualization and analysis system (Section 3.4).
2. A novel emotion categorization model (Section 3.5).
3. A technique for clustering sentics (Section 3.6).

3.1. ConceptNet

ConceptNet (Liu & Singh, 2004) is a semantic resource structurally similar to WordNet, but whose scope of contents is general world knowledge, in the same vein as Cyc. Instead of insisting on formalizing common sense reasoning using mathematical logic (Mueller, 2006), ConceptNet uses a new approach: it represents data in the form of a semantic network and makes it available to be used in natural language processing. The prerogative of ConceptNet, in fact, is contextual common sense reasoning: while WordNet is optimized for lexical categorization and word-similarity determination, and Cyc is optimized for formalized logical reasoning, ConceptNet is optimized for making practical context-based inferences over real-world texts.

In ConceptNet, WordNet's notion of node in the semantic network is extended from purely lexical items (words and simple phrases with atomic meaning) to include higher-order compound concepts, e.g., 'satisfy hunger', 'follow recipe', to represent knowledge around a greater range of concepts found in everyday life (see Table 1). Moreover WordNet's repertoire of semantic relations is extended from the triplet of synonym, is-a, and part-of, to a repertoire of twenty semantic relations including, for example, EffectOf (causality), SubeventOf (event hierarchy), CapableOf (agent's ability), MotivationOf (affect), PropertyOf, LocationOf. ConceptNet's knowledge is also of a more informal, defeasible and practi-

cally valued nature. For example, WordNet has formal taxonomic knowledge that 'dog' is a 'canine', which is a 'carnivore', which is a 'placental mammal'; but it cannot make the practically oriented member-to-set association that 'dog' is a 'pet'.

ConceptNet also contains a lot of knowledge that is defeasible, i.e., it describes something that is often true but not always, e.g., EffectOf ('fall off bicycle', 'get hurt'), which is something we cannot leave aside in common sense reasoning. Most of the facts interrelating ConceptNet's semantic network are dedicated to making rather generic connections between concepts. This type of knowledge can be brought back to Minsky's K-lines as it increases the connectivity of the semantic network and makes it more likely that concepts parsed out of a text document can be mapped into ConceptNet. ConceptNet is produced by an automatic process, which first applies a set of extraction rules to the semistructured English sentences of the Open Mind Common Sense (OMCS) corpus, and then applies an additional set of 'relaxation' procedures, i.e., filling in and smoothing over network gaps, to optimize the connectivity of the semantic network (Fig. 2).

In ConceptNet version 2.0, a new system for weighting knowledge was implemented, which scores each binary assertion based on how many times it was uttered in the OMCS corpus, and on how well it can be inferred indirectly from other facts in ConceptNet. In ConceptNet version 3.0 (Havasi et al., 2007), users can also participate in the process of refining knowledge by evaluating existing statements on Open Mind Commons (Speer, 2007), the new interface for collecting common sense knowledge from users over the Web.

By giving the user many forms of feedback and using inferences by analogy to find appropriate questions to ask, Open Mind Commons can learn well-connected structures of common sense knowledge, refine its existing knowledge, and build analogies that lead to even more powerful inferences. The pieces of common sense knowledge acquired through this interface are made publicly available in ConceptNet, which is released periodically both as an SQL database and through an API.

3.2. CF-IOF weighting

CF-IOF (concept frequency – inverse opinion frequency) (Cambria et al., 2010) is a technique that identifies common domain-dependent semantics, using an approach similar to TF-IDF weighting, in order to evaluate how important a concept is to a set of opinions concerning the same topic. Firstly, the frequency of a concept c for a given domain d is calculated by counting the occurrences of the concept c in the set of available d -tagged opinions and dividing the result by the sum of number of occurrences of all concepts in the set of opinions concerning d . This frequency is then multiplied by the logarithm of the inverse frequency of the concept in the whole collection of opinions, that is:

$$CF - IOF_{c,d} = \frac{n_{c,d}}{\sum_k n_{k,d}} \log \sum_k \frac{n_k}{n_c}$$

where $n_{c,d}$ is the number of occurrences of concept c in the set of opinions tagged as d , n_k is the total number of concept occurrences and n_c is the number of occurrences of c in the whole set of opinions. A high weight in CF-IOF is reached by a high concept frequency in a given domain and a low frequency of the concept in the whole collection of opinions. Therefore, thanks to CF-IOF weights, it is possible to filter out common concepts and detect relevant domain-dependent semantics.

3.3. Spectral association

Spectral association (Havasi et al., 2010) is a technique that involves assigning values, or activations, to 'seed concepts' and applying an operation that spreads their values across the ConceptNet

Table 1
Comparing WordNet and ConceptNet: while WordNet synsets contain vocabulary knowledge associated with concepts, ConceptNet assertions convey generic knowledge about what such concepts are used for.

| Term | WordNet hypernyms | ConceptNet assertions |
|---------------|---|---|
| Cat | Feline; Felid; Adult male; Man; Gossip; Gossiper; Gossipmonger; Rumormonger; Rumourmonger; Newsmonger; Woman; Adult female; Stimulant; Stimulant drug; Excitant; Tracked vehicle; ... | Cats can hunt mice; Cats have whiskers; Cats can eat mice; Cats have fur; cats have claws; Cats can eat meat; cats are cute; ... |
| Dog | Canine; Canid; Unpleasant woman; Disagreeable woman; Chap; Fellow; Feller; Lad; Gent; Fella; Scoundrel; Sausage; Follow, ... | Dogs are mammals; A dog can be a pet; A dog can guard a house; You are likely to find a dog in kennel; An activity a dog can do is run; A dog is a loyal friend; A dog has fur; ... |
| Language | Communication; Auditory communication; Word; Higher cognitive process; Faculty; Mental faculty; Module; Text; Textual matter; | English is a language; French is a language; Language is used for communication; Music is a language; A word is part of language; ... |
| iPhone | N/A; | An iPhone is a kind of a telephone; An iPhone is a kind of computer; An iPhone can display your position on a map; An iPhone can send and receive emails; An iPhone can display the time; ... |
| Birthday gift | Present; | Card is birthday gift; Present is birthday gift; Buying something for a loved one is for a birthday gift; ... |

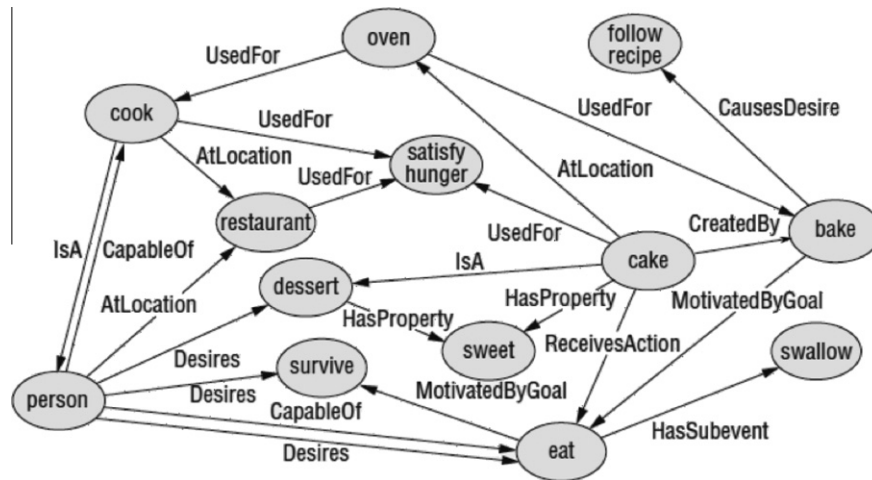


Fig. 2. ConceptNet represents the information in the Open Mind corpus as a directed graph where nodes are concepts and labeled edges are assertions of common sense that interconnect them.

graph. This operation, an approximation of many steps of spreading activation, transfers the most activation to concepts that are connected to the key concepts by short paths or many different paths in common sense knowledge. In particular, we build a matrix C that relates concepts to other concepts, instead of their features, and add up the scores over all relations that relate one concept to another, disregarding direction. Applying C to a vector containing a single concept spreads that concept's value to its connected concepts. Applying C^2 spreads that value to concepts connected by two links (including back to the concept itself). But what we would really like is to spread the activation through any number of links, with diminishing returns, so the operator we want is:

$$1 + C + \frac{C^2}{2!} + \frac{C^3}{3!} + \dots = e^C$$

We can calculate this odd operator, e^C , because we can factor C . C is already symmetric, so instead of applying Lanczos' method to CC^T and getting the singular value decomposition (SVD), we can apply it directly to C and get the spectral decomposition $C = V\Lambda V^T$. As before, we can raise this expression to any power and cancel everything but the power of Λ . Therefore, $e^C = Ve^{\Lambda}V^T$. This simple twist on the SVD lets us calculate spreading activation over the whole matrix instantly. As with the SVD, we can truncate these matrices to k axes and therefore save space while generalizing from similar concepts. We can also rescale the matrix so that activation values have a maximum of 1 and do not tend to collect in highly-connected concepts such as 'person', by normalizing the truncated rows of

$Ve^{\Lambda/2}$ to unit vectors, and multiplying that matrix by its transpose to get a rescaled version of $Ve^{\Lambda}V^T$.

3.4. AffectiveSpace

AffectiveSpace (Cambria et al., 2009) is a multi-dimensional vector space built by 'blending' (Havasi et al., 2009) ConceptNet with WordNet-Affect (WNA) (Strapparava and Valitutti, 2004), a linguistic resource for the lexical representation of affective knowledge. Blending is a technique that performs inference over multiple sources of data simultaneously, taking advantage of the overlap between them. It basically combines two sparse matrices linearly into a single matrix in which the information between the two initial sources is shared. When we perform SVD on a blended matrix, the result is that new connections are made in each source matrix taking into account information and connections present in the other matrix, originating from the information that overlaps. The alignment operation operated over ConceptNet and WNA yields a new matrix, A , in which common sense and affective knowledge coexist, i.e., a matrix $14,301 \times 117,365$ whose rows are concepts (e.g., 'dog' or 'bake cake'), whose columns are either common sense and affective features (e.g., 'isA-pet' or 'hasEmotion-joy'), and whose values indicate truth values of assertions. Therefore, in A , each concept is represented by a vector in the space of possible features whose values are positive for features that produce an assertion of positive valence (e.g., 'a penguin is a bird'), negative for features that produce an assertion of negative valence (e.g., 'a

Fig. 3. Affectively positive (bottom-left corner) and affectively negative (up-right corner) common sense concepts in AffectiveSpace.

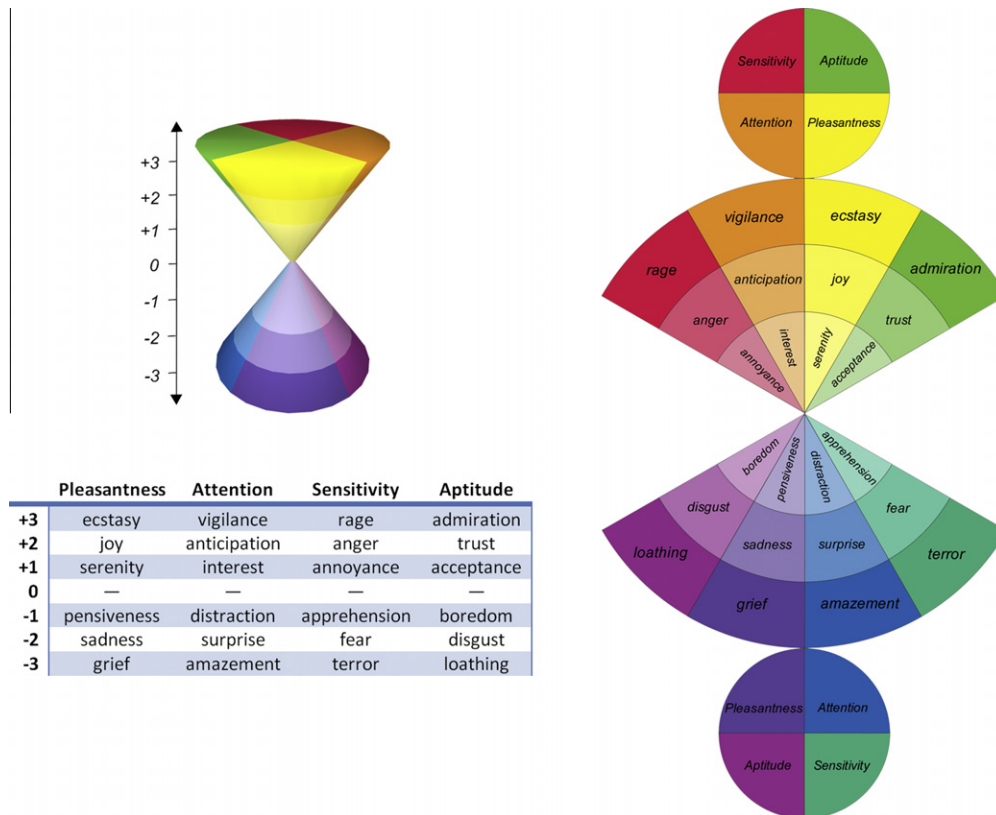


Fig. 4. The 3D model and the net of the Hourglass of Emotions. Dimensional and discrete forms of the different sentic levels are summarized in the proposed emotion categorization table.

3. The user is comfortable with the interface (Sensitivity).
4. The user is disposed to use the application (Aptitude).

Each affective dimension is characterized by six levels of activation, called 'sentic levels', which determine the intensity of the expressed/perceived emotion as an $int \in [-3, +3]$. These levels are also labeled as a set of 24 basic emotions (six for each of the affective dimensions) in a way that allows the model to specify the affective information associated with text both in a dimensional and in a discrete form. The dimensional form, in particular, is called 'sentic vector' and it is a four-dimensional *float* vector that can potentially express any human emotion in terms of Pleasantness, Attention, Sensitivity and Aptitude. Some particular sets of sentic vectors have special names as they specify well-known compound emotions. For example the set of sentic vectors with a level of Pleasantness $\in (+1, +2]$ ('joy'), a null Attention, a null Sensitivity and a level of Aptitude $\in (+1, +2]$ ('trust') are called 'love sentic vectors' since they specify the compound emotion of 'love'.

3.6. Sentic medoids

Sentic medoids (Cambria et al., 2011) is a clustering technique that adopts a *k*-medoids approach (Kaufman and Rousseeuw, 1990) to partition affective common sense concepts in Affective-Space into *k* clusters around as many centroids, trying to minimize a given cost function. Differently from the *k*-means algorithm (Hartigan and Wong, 1979), which does not pose constraints on centroids, *k*-medoids do assume that centroids must coincide with *k* observed points. The *k*-means approach finds the *k* centroids, where the coordinate of each centroid is the mean of the coordinates of the objects in the cluster and assigns every object to the nearest centroid. Unfortunately, *k*-means clustering is sensitive

to the outliers and a set of objects closest to a centroid may be empty, in which case centroids cannot be updated. For this reason, *k*-medoids are sometimes used, where representative objects are considered instead of centroids. In many clustering problems, in fact, one is interested in the characterization of the clusters by means of typical objects, which represent the various structural features of objects under investigation. Because it uses the most centrally located object in a cluster, *k*-medoids clustering is less sensitive to outliers compared with *k*-means.

Among many algorithms for *k*-medoids clustering, partitioning around medoids (PAM) is one of the most widely used. The algorithm, proposed by Kaufman and Rousseeuw (Kaufman and Rousseeuw, 1990), first computes *k* representative objects, called medoids. A medoid can be defined as that object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal. PAM determines a medoid for each cluster selecting the most centrally located centroid within the cluster. After selection of medoids, clusters are rearranged so that each point is grouped with the closest medoid. Compared to *k*-means, PAM operates on the dissimilarity matrix of the given dataset. It is more robust, because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances. A particularly nice property is that PAM allows clustering with respect to any specified distance metric. In addition, the medoids are robust representations of the cluster centers, which is particularly important in the common context that many elements do not belong well to any cluster. However, PAM works inefficiently for large data sets due to its complexity.

To this end, a modified version of the algorithm recently proposed by Park and Jun (2009) was used, which runs similarly to the *k*-means clustering algorithm. This has shown to have similar performance when compared to PAM algorithm while taking a significantly reduced computational time. In particular, we have *N*

concepts ($N = 14,301$) encoded as points $x \in \mathbb{R}^p$ ($p = 50$). We want to group them into k clusters and, in our case, we can fix $k = 24$ as we are looking for one cluster for each sentic level s of the Hourglass model. Generally, the initialization of clusters for clustering algorithms is a problematic task as the process often risks to get stuck into local optimum points, depending on the initial choice of centroids (Duda and Hart, 1973). However, we decide to use as initial centroids the concepts that are currently used as centroids for clusters, as they specify the emotional categories we want to organize AffectiveSpace into. For this reason, what is usually seen as a limitation of the algorithm can be seen as advantage for this approach, since we are not looking for the 24 centroids leading to the best 24 clusters but indeed for the 24 centroids identifying the required 24 sentic levels (i.e., the centroids should not be ‘too far’ from the ones currently used).

In particular, as the Hourglass affective dimensions are independent but concomitant, we need to cluster AffectiveSpace four times, once for each dimension. According to the Hourglass categorization model, in fact, each concept can convey, at the same time, more than one emotion (which is why we get compound emotions) and this information can be expressed via a sentic vector specifying the concept’s affective valence in terms of Pleasantness, Attention, Sensitivity and Aptitude. Therefore, given that the distance between two points in AffectiveSpace is defined as $D(a, b) = \sqrt{\sum_{i=1}^p (a_i - b_i)^2}$ (note that the choice of Euclidean distance is arbitrary), the used algorithm, applied for each of the four affective dimensions, can be summarized as follows:

1. Each centroid $C_n \in \mathbb{R}^{50}$ ($n = 1, 2, \dots, k$) is set as one of the six concepts corresponding to each s in the current affective dimension.
2. Assign each record x to a cluster Ξ so that $x_i \in \Xi_n$ if $D(x_i, C_n) \leq D(x_i, C_m) \forall m = 1, 2, \dots, k$.
3. Find a new centroid C for each cluster Ξ so that $C_j = x_i$:

$$\text{if } \sum_{x_m \in \Xi_j} D(x_i, x_m) \leq \sum_{x_m \in \Xi_j} D(x_h, x_m) \quad \forall x_h \in \Xi_j$$

4. Repeat step 2 and 3 until no changes on centroids are observed.

Note that condition posed on steps 2 and 3 may occasionally lead to more than one solution. Should this happen, our model will randomly choose one of them. This clusterization of AffectiveSpace allows to calculate, for each common sense concept x , a four-dimensional sentic vector that defines its affective valence in terms of a degree of fitness $f(x)$ where $f_a = D(x, C_j)C_j | D(x, C_j) \leq D(x, C_k) a = 1, 2, 3, 4 \quad k = 6a - 5, 6a - 4, \dots, 6a$.

4. Structuring the unstructured

Among the benefits of questionnaires’ structuredness, there are the quickness, effectiveness and ease to use and understand. However, such structuredness involves some drawbacks. A questionnaire, in fact, can limit the possibility to discover new important patterns in the input data and can constrain users to omit important opinions that might be valuable for measuring service quality. In the medical sphere, in particular, patients driven by very positive or very negative emotions are usually willing to detailedly express their point of view, which can be particularly valuable for assessing uncovered points, raising latent problems or redesigning the questionnaire. To this end, *Sentic PROMs* adopt a semi-structured approach that allows patients to assess their health status and health-care experience both by filling in a four-level questionnaire and by adding free text. The two different input methods are not mutually exclusive but complementary. When patients are happy with simply filling-in the questionnaire, they

can just leave the input text box blank but, when they feel like speaking out their opinions and feelings, e.g., in the occasion of a particularly positive or negative situation or event, they can do it in their own words.

As a result, the stored input data, although very similar at concept level, are on two completely different structure levels – structured (questionnaire selection) and unstructured (natural language). *Sentic PROMs* aim to format the unstructured input and accordingly aggregate it with the structured data in order to perform statistical analysis and pattern recognition on these and, hence, extract meaningful information. In order to bridge such gap between unstructured and structured patient data, the semantics and sentics associated with natural language text are extracted by means of sentic computing. In particular, semantics are built on the top of patient data and metadata, while sentics are built on the top of semantics, as they represent the emotions or the polarity conveyed by the detected concepts (Fig. 5). Specifically, given a free text patient input containing a set of opinions O about a set of topics T with different polarity $p \in [-1, 1]$, we extract, for each $t \in T$, the subset of opinions $o \subseteq O$ concerning t and determine p .

In other words, since each patient opinion can regard more than one topic and the polarity values associated with each topic are often independent from each other, in order to perform the mapping we need to extract, from each opinion, a set of topics and then, from each topic detected, the polarity associated with it. Since both the procedures work at semantic level, they can be combined in a unique process having patient opinions as input and both semantics and sentics as outputs.

The developed patient opinion mining engine consists of four main components: a pre-processing module, which performs a first skim of the opinion, a semantic parser, whose aim is to extract concepts from the opinionated text, the ConceptNet module, for the inference of the semantics associated with the given concepts, and the AffectiveSpace module, for the extraction of sentics. The pre-processing module firstly interprets all the affective valence indicators usually contained in opinionated text such as special punctuation, complete upper-case words, onomatopoeic repetitions, exclamation words, negations, degree adverbs and emoticons. Secondly, it converts text to lower-case and, after lemmatizing it, splits the opinion into single clauses according to grammatical conjunctions and punctuation.

The semantic parser deconstructs text into concepts using a lexicon based on sequences of lexemes that represent multiple-word concepts extracted from ConceptNet, WordNet and other linguistic resources. These n-grams are not used blindly as fixed word patterns but exploited as reference for the module, in order to extract multiple-word concepts from information-rich sentences. So, differently from other shallow parsers, the module can recognize complex concepts also when irregular verbs are used or when these are interspersed with adjective and adverbs, e.g., the concept ‘buy christmas present’ in the sentence ‘I bought a lot of very nice Christmas presents’. The semantic parser, additionally, provides, for each retrieved concept, the relative frequency, valence and status, that is the concept’s occurrence in the text, its positive or negative connotation and the degree of intensity with which the concept is expressed. For each clause, the module outputs a small

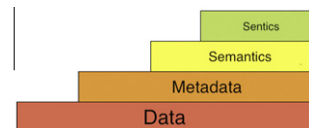


Fig. 5. The semantics and sentics stack. Semantics are built on the top of patient data and the metadata associated with these. Sentic, in turn, are built on the top of semantics, as they represent the emotions (or polarity) conveyed by the detected concepts.

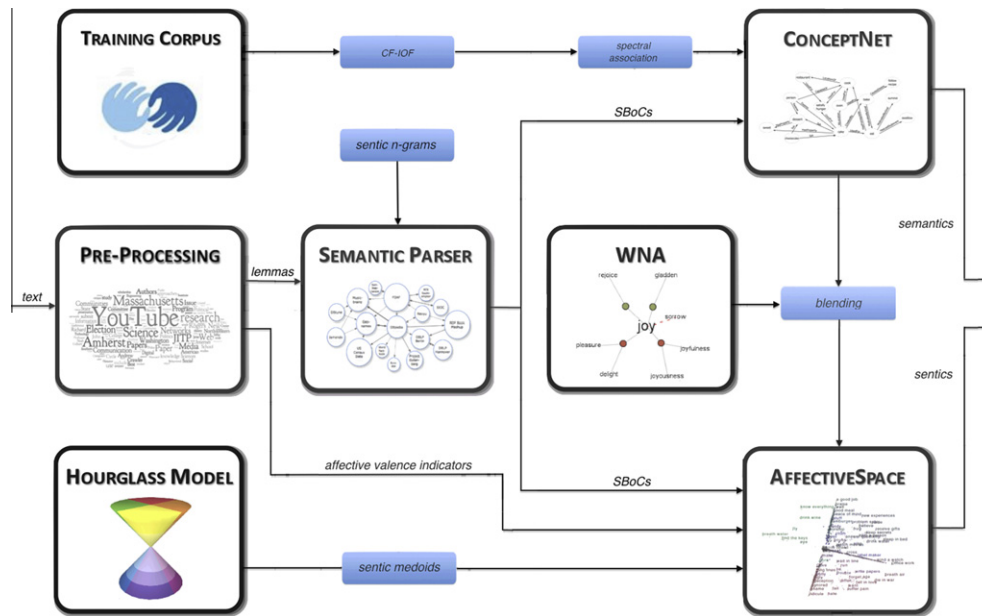


Fig. 6. The patient opinion analysis process basically consists in firstly skimming input text, secondly extracting small sets of concepts, and eventually inferencing semantics and sentics associated with these.

bag of concepts (SBoC), which is later on analyzed separately by the ConceptNet and AffectiveSpace modules to infer the cognitive and affective information associated with the input text, respectively (Fig. 6).

Specifically, the ConceptNet module employs spectral association for assigning activation to key concepts, that is nodes of the semantic network which are used as seeds for classification. Such seeds are found by applying CF-IOF on a set of 2000 topic-tagged posts extracted from PatientOpinion¹, a social enterprise providing an on-line feedback service for users of the UK National Health Service (NHS). Thanks to CF-IOF weights, it is possible to filter out common concepts and detect domain dependent concepts that identify topics typically found in patient opinions, e.g., cleanliness, food, kindness of staff and timeliness. Such concepts represent seed concepts for spectral association, which spreads their values across the ConceptNet graph and, hence, detects semantically related concepts concerning the same topic, which are stored in a database to be accessed at run-time by the opinion analysis process. Therefore, the concepts of each SBoC provided by the semantic parser are projected on the matrix resulting from spectral association in order to calculate their semantic relatedness to each seed concept and, hence, their degree of belonging to each different class. Such classification measure is directly proportional to the degree of connectivity between the nodes representing the retrieved concepts and the seed concepts in the ConceptNet graph.

As for the extraction of sentics, the concepts output by the semantic parser are also given as input to the AffectiveSpace module, which employs dimensionality reduction techniques to infer the affective information associated with them. To this end, the concepts of each SBoC are projected into AffectiveSpace and, according to their position in the vector space representation of affective common sense knowledge, they are assigned to one of the 24 affective clusters defined through the sentic medoids technique. In case any of the detected concepts is found more than once in the vector space (that is, any of the concepts has multiple senses), all the SBoC concepts are exploited for a context-dependent coarse sense disambiguation. In particular, to represent the

expected semantic value of the clause as a whole, we can average together the vectors corresponding to all concepts in the clause (in their ambiguous form). The resulting vector does not represent a single meaning but the ‘ad hoc category’ of meanings that are similar to the various possible meanings of concepts in the clause (Havasi et al., 2010). Then, to assign the correct sense to the ambiguous concept, we find the sense of each concept that has the highest dot product (strongest similarity) with the clause vector.

As well as in the ConceptNet module, the categorization does not consist in simply labeling each concept but also in assigning a confidence score to each emotional label, which is directly proportional to the degree of belonging to a specific affective cluster (dot product between the given concept and the relative sentic medoid). Such affective information, expressed in terms of sentic vector, can also be exploited to calculate a polarity value associated with each SBoC and, hence, detect the overall polarity associated with the opinionated text by averaging out all the SBoCs’ polarity values, according to the following formula:

$$p = \sum_{i=1}^N \frac{\text{Pleasantness}(o_i) + |\text{Attention}(o_i)| - |\text{Sensitivity}(o_i)| + \text{Aptitude}(o_i)}{9N}$$

where N is the total number of retrieved concepts and 9 is the normalization factor (as the maximum and minimum values of the numerator are given by the sentic vectors $[3, \pm 3, 0, 3]$ and $[-3, 0, \pm 3, -3]$, respectively). In the formula, Attention and Sensitivity are taken in absolute value since, from the point of view of polarity rather than affection, all of their sentic values represent positive and negative values respectively (e.g., ‘anger’ is positive in the sense of level of activation of Sensitivity but negative in terms of polarity and ‘surprise’ is negative in the sense of lack of Attention but positive from a polarity point of view).

As an example of how the software engine works, we can examine intermediate and final outputs obtained when a natural language opinion is given as input to the system. We choose the patient opinion “This back pain is limiting me a lot but at least I can move around from time to time for my basic needs. And anyway staff is always so nice and helpful that I don’t feel like a burden to them”. After the pre-processing and semantic parsing operations, we obtain the following SBoCs:

¹ <http://patientopinion.org.uk>.

Table 2

Structured output example of the opinion mining engine: for each clause, the engine detects opinion target, category and affective information associated with it both in a textual form (emotional label) and in a dimensional form (polarity).

| SBoC | Opinion Target | Category | Moods | Polarity |
|------|----------------|-------------------------|-----------------------------|----------|
| #1 | 'back pain' | 'clinical', 'symptom' | 'annoyance', 'apprehension' | −0.85 |
| #2 | 'back pain' | 'limitation', 'symptom' | 'acceptance', 'serenity' | +0.39 |
| #3 | 'staff' | 'staff', 'service' | 'joy', 'trust' | +0.75 |
| #4 | 'staff' | 'dependency' | 'joy' | +0.56 |

SBoC#1:

< Concept: 'back pain'>

< Concept: 'limit'++>

SBoC#2:

< Concept: 'able move'>

< Concept: 'basic need'>

SBoC#3:

< Concept: 'staff'>

< Concept: 'nice'++>

< Concept: 'helpful'++>

SBoC#4:

< Concept: 'feel like burden'>

These are then concurrently processed by the ConceptNet and the AffectiveSpace modules, which output the cognitive and affective information associated with each SBoC, both in a discrete way, with one or more labels, and in a dimensional way, with a polarity value $\in [-1, +1]$ (as shown in Table 2).

5. Monitoring patients' physio-emotional sensitivity

The importance of physio-emotional sensitivity in humans has been proven by recent health research, which has shown that individuals who feel loved and supported by friends and family, or even by a loving pet, tend to have higher survival rates following heart attacks than other cardiac patients who experience a sense of social isolation. Such concept is also reflected in natural language as we use terms such as 'heartsick', 'broken-hearted' and 'heartache' to describe extreme sadness and grief, idioms like 'full of gall' and 'venting your spleen' to describe anger, and expressions such as 'gutless', 'yellow belly' and 'feeling kicked in the gut' to describe shame.

Human body contracts involuntarily when it feels emotional pain such as grief, fear, disapproval, shock, helplessness, shame, terror, in the same reflex it does if physically injured. Such gripping reflex normally releases slowly, but if a painful experience is intense, or happens repeatedly, the physio-emotional grip does not release, and constriction is retained in the body. Any repeated similar experience then layers on top of the original unreleased contraction, until we are living with layers of chronic tension, which constricts our bodies. The mind, in fact, may forget the origin of pain and tension, but the body does not.

Besides HRQoL measurement, *Sentic PROMs* aim to monitor also users' physio-emotional sensitivity on a regular basis, as a means of patient affective modeling. In particular, the dimensional affective information coming from both questionnaire data (*howRu* aggregated score) and natural language data (sentic vectors) is stored separately by the system every time patients conclude a *Sentic PROMs* session and plotted on four different bi-dimensional diagrams (Fig. 7). Such diagrams represent the pairwise fusion of the four dimensions of the Hourglass model and allow to detect more complex (compound) emotions that can be particularly relevant for monitoring patients' physio-emotional sensitivity, e.g., frustration, anxiety, optimism, disapproval and rejection.

6. Evaluation

Since *Sentic PROMs* interface is still under-development, at the time of writing this article we had no consistent dataset of aggregated patient data available for thoroughly testing the system. Hence, a preliminary evaluation of the system had to be performed separately at two different levels: structured level (questionnaire data) and unstructured level (natural language data). As for the structured-level evaluation, a validation study was undertaken to examine the psychometric properties and construct validity of *howRu* and to compare these with *SF-12*. In particular, 2751 subjects with long-term conditions (average age 62, female 62.8%), were classified by *howRu* score, primary condition, number of conditions suffered, age group, duration of illness and area of residence. Across all six classifications, the correlation of the mean *howRu* scores with the mean values of the Physical Components Summary (*PCS-12*), the Mental Components Summary (*MCS-12*) and the sum of *PCS-12* + *MCS-12* were generally very high (0.91, 0.45 and 0.97 respectively) (Benson et al., 2010).

As for the unstructured-level evaluation, in turn, the engine's capability to extract cognitive and affective information from natural language opinions was tested. In particular, in order to calculate statistical classifications such as precision and recall of the semantics and sentics extraction process, we evaluated the system both with the PatientOpinion database and with a corpus of topic and mood tagged blogs from LiveJournal² (LJ). LJ is a virtual community of more than 23 million users who keep a blog, journal or diary.

One of the interesting features of this website is that LJ bloggers are allowed to label their posts not only with a topic tag but also with a mood label, by choosing from more than 130 predefined moods or by creating custom mood themes. Since the indication of the affective status is optional, the mood-tagged posts are likely to reflect the true mood of the authors and, hence, form a good test-set for the opinion engine. However, since LJ mood themes do not perfectly match the sentic levels, we had to consider a reduced set of 10 moods, i.e., 'ecstatic', 'happy', 'pensive', 'surprised', 'enraged', 'sad', 'angry', 'annoyed', 'scared' and 'bored'. Moreover we could not consider non-affective web-posts since mood-untagged blog entries do not necessarily lack emotions. As for the topic tags, in turn, we selected the LJ labels that match PatientOpinion topic-tags, e.g., 'food', 'cleanliness' or 'communication', in order to collect natural language text that is likely to have the same semantics as the cognitive information usually associated with patient opinions. All LJ accounts have Atom, RSS and other data feeds which show recent public entries, friend relationships and interests. Unfortunately the current LJ API allows retrieval of posts by topic only so, in order to also get mood-tagged posts, we had to design our own web crawler.

After retrieving and storing relevant data and metadata from 10,000 LJ posts, we extracted semantics and sentics through the opinion analysis process and compared the output with the relative topic and mood tags, in order to calculate precision, recall and F-measure rates. On average, each post contained around

² <http://livejournal.com>.

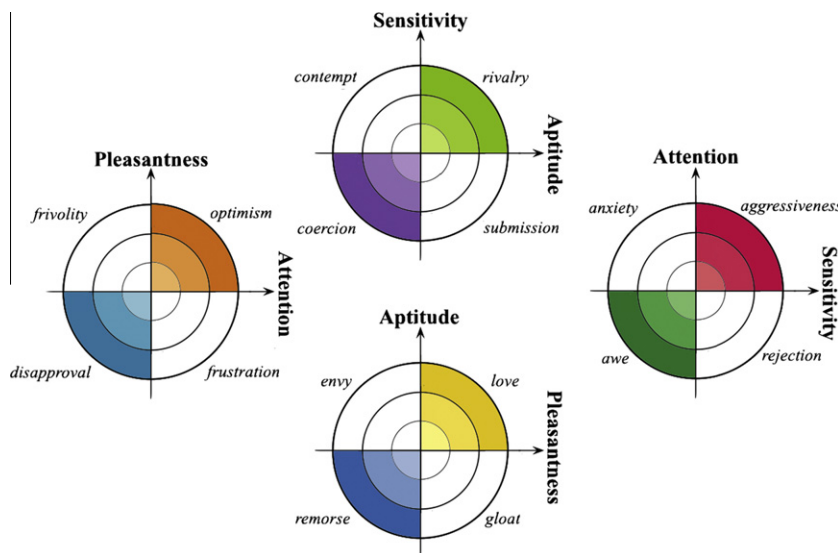


Fig. 7. Hourglass pairwise fusion diagrams. Plotting *howRu* aggregated score and sentic vectors allows to detect compound emotions that are particularly relevant for monitoring patients' physio-emotional sensitivity.

140 words, from which about 12 affective valence indicators and 60 concepts were extracted. From the retrieved concepts we inferred semantics and sentics associated with each of the selected posts and, hence, tagged them with topic and mood labels. We then compared these labels with the corresponding topic and mood LJ tags, obtaining very good accuracy in terms of both semantics and sentics extraction.

As for the detection of moods, for example, 'happy' and 'sad' posts were identified with particularly high precision (89% and 81% respectively) and good recall rates (76% and 68%). The F-measure values obtained, hence, were significantly good (82% and 74% respectively), especially if compared to the corresponding F-measure rates of existing approaches to automatic identification of emotions in text such as keyword spotting (53% F-measure for 'happy' posts and 51% for 'sad' posts), lexical affinity (63% and 58% F-measure rates respectively) and statistical methods (69% and 62% F-measure for 'happy' and 'sad' posts respectively). As for the detection of topics, in turn, the classification of 'food' and 'communication' posts was performed with a precision of 75% and 69% and recall rates of 65% and 58% respectively. The total F-measure rates, hence, were considerably good (70% for 'food' posts and 63% for 'communication' posts), particularly if compared to the corresponding F-measure rates of the baseline methods (44% and 35% for keyword spotting, 53% and 39% for lexical affinity, 61% and 52% for statistical methods).

We also performed an evaluation test on the 2000 topic- and polarity-tagged posts of the PatientOpinion database. Although CF-IOF has been applied on the database for extracting spectral association seeds, the polarity values of the dataset have not been exploited during the training phase. Hence, we could test the engine's capability to infer the sentics associated with each of the 2000 patient opinions and, in particular, calculate polarity detection accuracy, which, thanks to a full correspondence of concepts, resulted to be very high (91%). As soon as a new version of PatientOpinion database is released, we plan to perform a more comprehensive evaluation at both topic and polarity level in order to test the system's capability to discover all the different facets of the expressed opinion and the affective valence associated with each of these (feature-based sentiment analysis (Liu, 2010)). Further results will be submitted elsewhere for publication.

7. Conclusion and future work

Medicine is finally waking up to the use of novel technologies to listen to the 'wisdom of the patient'. Health-care of the future will be based on community, collaboration, self-caring, co-creation and co-production using technologies delivered via the Web. Engaging patients in their health-care and encouraging people to take responsibility for protecting their health, in fact, are seen as the best way to ensure the sustainability of health systems (WHO, 2000). Patients can play a distinct role in their own care by diagnosing and treating minor, self-limiting conditions and by preventing occurrence or recurrence of disease or harm, by selecting the most appropriate form of treatment for acute conditions in partnership with health professionals, and by actively managing chronic diseases.

Traditional paternalistic practice styles undermine people's confidence in their ability to look after themselves, so replacing paternalism with a partnership approach could help to enhance a sense of self-efficacy (Coulter, 2011). A growing body of evidence demonstrates that patient engagement in treatment decisions and in managing their own health-care can lead to more appropriate and cost-effective utilization of health services and better health outcomes (Coulter and Ellins, 2006).

This shift in emphasis to e-health does not replace traditional health care models but rather complements them and will ideally become the prevailing model. To aid such process, we proposed *Sentic PROMs* as a new framework for measuring health care quality that exploits the ensemble application of standard PROMs and sentic computing to overcome the common barriers to the use of HRQoL measurement systems, such as the respondent burden (the time needed to complete the forms) and the need for staff to be trained to understand the results.

Sentic PROMs, in fact, aim to be clinically useful and timely, sensitive to change, culturally sensitive, low burden, low cost, involve the patient and built into standard procedures and needs to meet the requirements of regulators, payers and continuous quality improvement. In particular, to bridge the gap between the structuredness of questionnaire data and the unstructuredness of natural language data, which are different at structure-level yet similar at concept-level, *Sentic PROMs* exploit both the semantics and sentics associated with patient opinions to accordingly aggregate such

data and, hence, evaluate patients' health status and experience in a semi-structured way, while tracking their physio-emotional sensitivity.

Soon, we plan to conduct on-field usability and performance tests on different case-mixes in order to thoroughly evaluate the system and possibly discover new interesting patterns. We also plan to further develop sentic computing techniques in order to enable the system to make more sense of the collected health-care data and, hence, be adaptive, in order to pave the way for the development of next-generation intelligent personal e-doctors.

Acknowledgments

This work has been part-funded by the UK Engineering and Physical Sciences Research Council (EPSRC Grant Reference: EP/G501750/1), the UK Technology Strategy Board (TSB Grant Reference: 12074-75246) and Sitekit Solutions Ltd. (a Company Registered in Scotland: SC116007).

References

- Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, 26(3), 1–34.
- Benson, T., Sizmur, S., Whatling, J., Arian, S., McDonald, D., & Ingram, D. (2010). Evaluation of a new short generic measure of health status. *Informatics in Primary Care*, 18(2), 89–101.
- Bergner, M., Bobbitt, R., Kressel, S., Pollard, W., Gilson, B., & Morris, J. (1976). The sickness impact profile: Conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services*, 6, 393–415.
- Brooks, R. (1996). EuroQoL – the current state of play. *Health Policy*, 37, 53–72.
- Cambria, E., Hussain, A., Havasi, C., & Eckl, C. (2009). AffectiveSpace: Blending common sense and affective knowledge to perform emotive reasoning. In *Proceedings of CAEPIA, Seville*, (pp. 32–41).
- Cambria, E., Hussain, A., Durrani, T., Havasi, C., Eckl, C., & Munro, J. (2010). Sentic computing for patient centered application. In *Proceedings of IEEE ICSP, Beijing*, (pp. 1279–1282).
- Cambria, E., Song, Y., Wang, H., & Hussain, A. (2011). Isanette: A common and common sense knowledge base for opinion mining. In *Proceedings of ICDM, Vancouver*.
- Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. In *Proceedings of FLAIRS, Marco Island*.
- Cambria, E., Grassi, M., Hussain, A., & Havasi, C. (2011). Sentic computing for social media marketing. *Multimedia Tools and Applications* <http://dx.doi.org/10.1007/s11042-011-0815-0>.
- Cambria, E., & Hussain, A. (2012). *Sentic computing: Techniques, tools, and applications*. Berlin Heidelberg: Springer.
- Cambria, E., Hussain, A., Havasi, C., & Eckl, C. (2009). Common sense computing: From the society of mind to digital intuition and beyond. In J. Fierrez, J. Ortega, A. Esposito, A. Drygajlo, & M. Faundez-Zanuy (Eds.), *Biometric ID management and multimodal communication. Lecture notes in computer science* (Vol. 5707, pp. 252–259). Berlin Heidelberg: Springer.
- Cambria, E., Mazzocco, T., Hussain, A., & Eckl, C. (2011). Sentic medoids: Organizing affective common sense knowledge in a multi-dimensional vector space. In *Advances in neural networks. Lecture notes in computer science* (Vol. 6677, pp. 601–610). Berlin Heidelberg: Springer-Verlag.
- Coulter, A. (2011). *Engaging patients in healthcare*. McGraw Hill.
- Coulter, A., & Ellins, J. (2006). The quality enhancing interventions project: Patient-focused interventions. *The Health Foundation*.
- Culyer, A., Lavers, R., & Williams, A. (1971). Social indicators: Health. *Social Trends*, 2, 31–42.
- Donabedian, A. (1966). Evaluating the quality of medical care. *The Millbank Memorial Fund Quarterly*, 44, 166–203.
- Duda, R., & Hart, P. (1973). *Pattern classification and scene analysis*. New York: John Wiley & Sons Inc.
- Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3), 211–218.
- Elliott, C. D. (1992). The affective reasoner: A process model of emotions in a multi-agent system, Ph.D. thesis. Northwestern University, Evanston.
- Fanshel, S., & Bush, J. (1970). A health status index and its application to health-services outcomes. *Operations Research*, 18, 1021–1066.
- Goertzel, B., Silverman, K., Hartley, C., Bugaj, S., & Ross, M. (2000). The Baby Webmind project. In *Proceedings of AISB, Birmingham*.
- Hartigan, J., & Wong, M. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, 28(1), 100–108.
- Havasi, C., Speer, R., & Alonso, J. (2007). ConceptNet 3: A flexible, multilingual semantic network for common sense knowledge. In *Proceedings of RANLP, Borovets*.
- Havasi, C., Speer, R., Holmgren, J. (2010). Automated color selection using semantic knowledge. In *Proceedings of AAAI CSK, Arlington*.
- Havasi, C., Speer, R., & Pustejovsky, J. (2010). Coarse word-sense disambiguation using common sense. In *Proceedings of AAAI CSK, Arlington, USA*.
- Havasi, C., Speer, R., Pustejovsky, J., & Lieberman, H. (2009). Digital intuition: Applying common sense using dimensionality reduction. *IEEE Intelligent Systems*, 24(4), 24–35.
- Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The health utility index (HUI): Concepts, measurement, properties and applications. *Health and Quality of Life Outcomes* 1 (54).
- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of AAAI, San Jose*.
- Kaufman, L., & Rousseeuw, P. (1990). *Finding groups in data: An introduction to cluster analysis*. Wiley-Interscience.
- Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*.
- Liu, H., & Singh, P. (2004). ConceptNet: A practical commonsense reasoning toolkit. *BT Technology Journal*.
- Minsky, M. (2006). The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind, *Simon & Schuster*.
- Mueller, E. (2006). *Commonsense reasoning*. Morgan Kaufman.
- Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP, Philadelphia*, (pp. 79–86).
- Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL, Ann Arbor*, (pp. 115–124).
- Park, H., & Jun, C. (2009). A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2), 3336–3341.
- Plutchik, R. (2001). The nature of emotions. *American Scientist*, 89(4), 344–350.
- Rao, D., & Ravichandran, D. (2009). Semi-supervised polarity lexicon induction. In *Proceedings of EACL, Athens*, (pp. 675–682).
- Somasundaran, S., Wiebe, J., & Ruppenhofer, J. (2008). Discourse level opinion interpretation. In *Proceedings of COLING, Manchester*.
- Speer, R. (2007). Open mind commons: An inquisitive approach to learning common sense. In *Proceedings of the Workshop on Common Sense and Interactive Applications, Honolulu*.
- Stevenson, R., Mikels, J., & James, T. (2007). Characterization of the affective norms for english words by discrete emotional categories. *Behavior Research Methods*, 39, 1020–1024.
- Strapparava, C., & Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In *Proceedings of LREC, Lisbon*.
- Torrance, G., Thomas, W., & Sackett, D. (1972). A utility maximisation model for evaluation of health care programs. *Health Services Research*, 7, 118–133.
- Turney, P., & Littman, M. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4), 315–346.
- Wall, M., Rechtsteiner, A., & Rocha, L. (2003). Singular value decomposition and principal component analysis. In D. Berrar, W. Dubitzky, & M. Granzow (Eds.), *A practical approach to microarray data analysis* (pp. 91–109). Springer.
- Ware, J. (1976). Scales for measuring general health perceptions. *Health Services Research*, 11, 396–415.
- Ware, J., Kosinski, M., & Keller, S. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care*, 34(3), 220–233.
- Ware, J., & Sherbourne, C. (1992). The MOS 36-item short-form health survey (SF-36). Conceptual framework and item selection. *Medical Care*, 30, 473–483.
- WHO. The world health report 2000: Health systems - improving performance, Technical report, World Health Organization (2000).
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2), 165–210.
- Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT/EMNLP, Vancouver*.