

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Emoji as a Proxy of Emotional Communication

Guillermo Santamaría-Bonfil

and Orlando Grabiel Toledano López

Abstract

Nowadays, emoji plays a fundamental role in human computer-mediated communications, allowing the latter to convey body language, objects, symbols, or ideas in text messages using Unicode standardized pictographs and logographs. Emoji allows people expressing more “authentically” emotions and their personalities, by increasing the semantic content of visual messages. The relationship between language, emoji, and emotions is now being studied by several disciplines such as linguistics, psychology, natural language processing (NLP), and machine learning (ML). Particularly, the last two are employed for the automatic detection of emotions and personality traits, building emoji sentiment lexicons, as well as for conveying artificial agents with the ability of expressing emotions through emoji. In this chapter, we introduce the concept of emoji and review the main challenges in using these as a proxy of language and emotions, the ML, and NLP techniques used for classification and detection of emotions using emoji, and presenting new trends for the exploitation of discovered emotional patterns for robotic emotional communication.

Keywords: emoji, machine learning, natural language processing, emotional communication, human-robot interaction

1. Introduction

Recently, in the episode “Smile” of the popular science fiction television program “Doctor Who,” a hypothetical off-earth colony is presented. This colony is maintained and operated by robots, which communicate and express emotions with humans and its pairs, through the usage of emoji. Sure, one may argue that such technology, besides being mere science fiction, is ridiculous since phonetic communication is much simpler and much easier to understand. While this is true for conventional information (e.g., explaining the concept of real numbers), communicating body emotional responses or gesticulation (e.g., to describe confusion) using only phonograms would require many more words to convey the same message than an emoji (e.g., 😊 or 😞). In this sense, emoji serve as a visual simplified form of (affective) communication that broadens the total amount of information (e.g., cues and gestures), which can be shared between humans and virtual/emodied artificial entities. If we consider that human languages, such as Chinese, Nahuatl [1], or even Sign Language, have evolved from ideographs and pictographs

lexicons, can we expect that in the near future, artificial entities (virtual or embodied) would employ emoji in their emotional communication?

The Japanese word *emoji* (e = picture and moji = word) literally stands for “picture word.” Although recently popularized, its older predecessors can be tracked to the nineteenth century, when cartoons were employed for humorous writing. *Smileys* followed in 1964 and were meant to be used by an insurance company’s promotional merchandise to improve the morale of its employee. The first to employ the emoticon :) in an online bulletin forum to denote humorous messages was Carnegie Mellon researchers in 1982, and 10 years later, the emoticons were already widespread in emails and Websites [2]. Finally, in 1998, Shigetaka Kurita devised emoji to improve emoticons pictorially, and became widespread by 2010. From this moment, the use of emoji has gained a lot of momentum, even achieving that the *word* namely “Face with Tears of Joy” (😄) was chosen by the Oxford Dictionary as the *Word of the Year* [2–4]. This choice was made under the assumption that the pictograph represented the ideas, beliefs, mood, and concerns of English speakers in 2015.

Since its origin, emoji undoubtedly have become a part of the mainstream communication around the globe allowing people, with different languages and cultural backgrounds, to share and interpret more accurately ideas and emotions. In this vein, it has been hypothesized that emoji shall become a *universal* language due to its generic communication features and its ever progressing lexicon [2, 5–7]. Although, this idea is controversial [8, 9] since emoji usage during the communication is influenced by factors such as context, users interrelations, users’ first language, repetitiveness, socio-demographics, among others [2, 5, 8]. This clearly adds ambiguity on how to employ them and its proper interpretation. Nevertheless, in the same fashion as sentiment analysis mines sentiments, attitudes, and emotions from text [10], we can employ billions (or perhaps more) of written messages within the Internet that contains emoji, to generate affective responses in artificial entities. More precisely, using natural language processing (NLP) along with machine learning (ML), we can extract semantics, express beat gestures, emotional states, and affective cues, add personality layers, among other characteristics from text. All this knowledge can be used to build, for instance, emoji sentiment lexicons [10] that will conform the emoji communication competence [2] that will power the engines of the emotional expression and communication of an artificial entity.

In the rest of this chapter, we first review the elements of the emoji code, and how emoji are used in the emotional expression and communication (Section 2). Afterward, in Section 3, we present a review of the state of the art in the usage of NLP and ML to classify and predict annotation and expression of emotions, gestures, affective cues, and so on, using written messages from multiple types of sources. In Section 4, we present several examples on how emoji are currently employed by artificial entities, both virtual and embodied, for expressing emotions during its interaction with humans. Lastly, Section 5 summarizes the chapter and discusses open questions regarding emoji usage as a source for robotic emotional communication.

2. Competence, lexicon, and ways of usage of emoji

To study how emoji are employed and about its challenges, we cannot simply do it without specifying the *emoji competence* [2]. Loosely speaking, competence (either linguistic or communicative) stands for the rules (e.g., grammar) and abilities an individual owns to correctly employ a given language to convey a specific idea [11]. Hence, the emoji competence stands for an adequate usage of emoji within

messages, not only in their representation but also in exact position within the message, to address a specific function (e.g., emotional expression, gestures, maintain interest in the communication, etc.) [2]. Nevertheless, even while the emoji competence has not been formally defined yet, and it can only be developed through the usage of emoji themselves [2, 6], here, we elaborate several of its components.

A key element of the emoji competence is the *emoji lexicon*, which is the standardization of pictograms (i.e., figures that resemble the real-world object), ideograms (i.e., figures that represent an idea) and logograms (i.e., figures that represent a sound or words) into *anime-like* graphical representations that belong to the ever-growing Unicode computer character standard [2, 6, 12]. These are employed within any message in three different ways: adjunctively, substitutively, or providing mixed textuality. In the first case, emoji appear along text within specific points of the written message (e.g., at the end of it) conveying it with emotional tone or adding visual annotations; it requires an overall low emoji competence. In the second case, emoji replace words, requiring a higher degree of competence to understand, not only the symbols per se but also the layout structure of the message, for instance, if we consider *syntagms*, which are symbols sequentially grouped that together conform a new idea (e.g., I love coffee = 🧐❤️☕). The third case intertwines text with emoji in a substitutive form rather than adjunctively. This case is the one that requires the highest emoji competence degree, since its decoding requires sophisticated knowledge about rhetorical structures and the proper usage of signs and symbols.

The emoji lexicon possesses generic features such as *representationality*, which allows signs and usage rules to be combined in specific forms to convey a message. Similarly, any person who is well versed with code's signs and rules is capable of interpreting any message based on the code (i.e., *interpretability*). However, messages built using the emoji lexicon are affected by *contextualization*, allowing that references, interpersonal relationships, and other factors affect the meaning of the message [4, 5]. Besides these, the emoji code is composed by a core and peripheral lexicon [2, 5]. As in the *Swadesh list*, the core lexicon stands for those emoji whose meaning and usage is, somehow, universally accepted and used, even while the Unicode supports more than 1000 different emoji [10]. Within this stand, all facial emoji also contain those emoji that stand for Ekman's six basic emotions such as surprise (😱¹) or anger (😡) [2, 13]. On the other hand, the peripheral lexicon is constituted by specialized communication symbols such as the one required for marketing, education [14], promoting national identity, or cultural cues [2], among others. Nevertheless, it is worth mentioning that since emoji may be used as nouns, verbs, or other grammatical structure, even those in the core lexicon can be used as a peripheral element in accordance with users' first language, its position within message, or by concatenating several of them into a syntagm.

2.1 How do we use emoji?

Emoji within any message can have several functions; **Figure 1** summarizes these. As shown by the latter, one of the most important functions an emoji has is *emotivity*, which adds an emotional layer to plain text communication. In this sense, emoji serve as a substitute of face-to-face (F2F) facial expressions, gestures, and body language, to state oneself emotional states, moods, or affective nuances. When used in this manner, emoji take the role of discourse strategies such as intonation or

¹ <https://emojipedia.org/face-screaming-in-fear/>

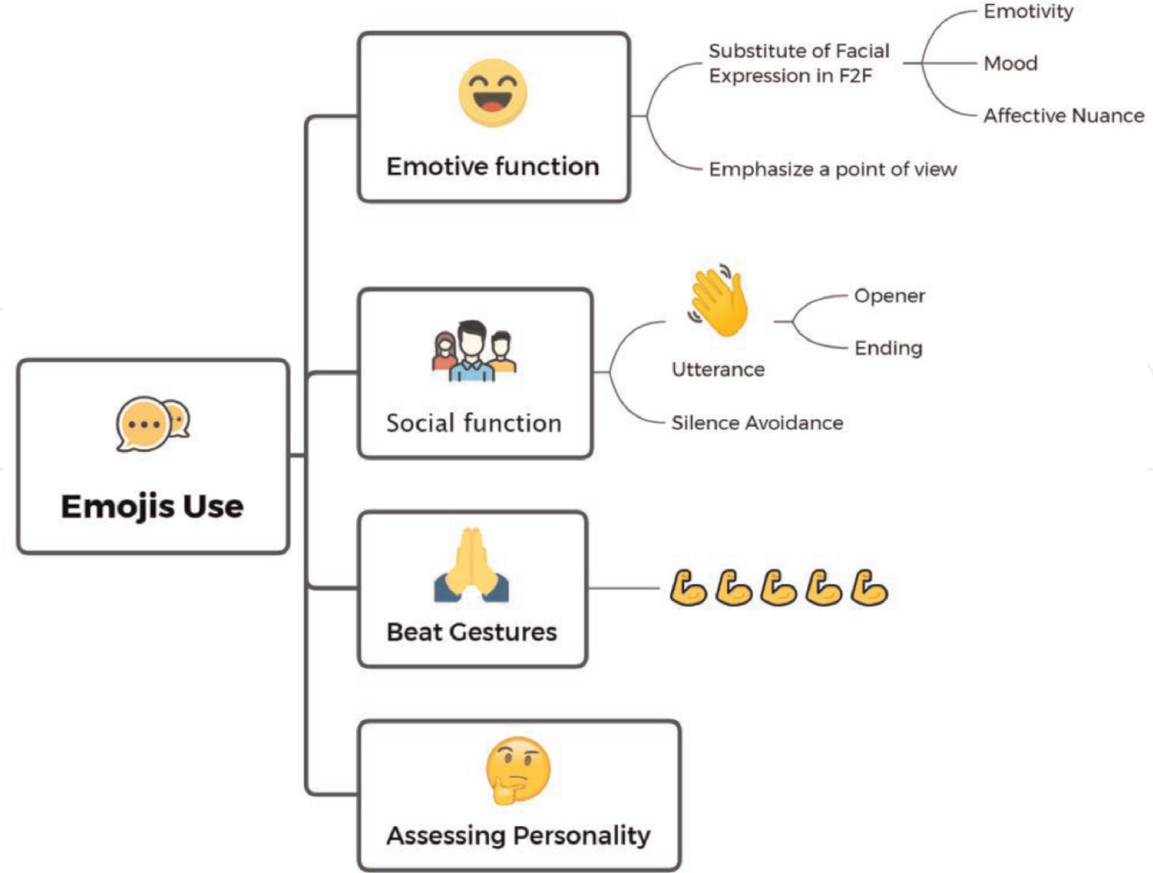


Figure 1.
Emoji functions within the computer-mediated communications.

phrasing [2, 4, 15]. Emoji *emotivity* mostly conveys positive emotions, hence it can be employed to emphasize an specific point of view, such as sarcasm, while softening the negative emotions associated with it (e.g., with respect to the one that is being sarcastic), allowing the receiver of the message to focus on the content instead of the negativity elicited [2, 14].

Another important role of emoji is as phatic instrument during communication [2, 16]. In this sense, they are employed as utterances that allow the flow of the conversation to unfold pleasantly and smoothly. In this sense, emoji serve as an opener or ending utterance (i.e., waving hand) to open or close a conversation, respectively, maintaining a positive dialog regardless of the content. Similarly, emoji can be used to fill uncomfortable moments of silence during a conversation avoiding its abrupt interruption. Beat gestures are another function of emoji; the former can be defined as a repetitive rhythmical co-speech gesture that emphasizes the *rhythm* of the speech [9]. For instance, in the same way that keeping nodding up and down during a conversation emphasizes agreement with the interlocutor, emoji can be repeated to convey the same meaning (e.g., 🙌🙌🙌🙌). Keeping in mind that although emoji, neither as utterance nor as beat gesture, explicitly stands for an emotional reaction, they implicitly convey an emotional (positive) tone to the conversation. Likewise, the other function of emoji, which is also implicitly related to emotion, is personality. The latter stands for basal characteristics that have pre-established effects on thoughts, behaviors, and emotions [17]. Been considered a genetic trait, it suffers less variability over time in contrast to emotions and moods [17]. In this sense, emoji can be used to elucidate the underlying personality traits of individuals, either by data mining or by replacing text-based items by their emoji equivalent in personality tests [18].

3. Studying emoji usage using formal frameworks

Emoji usage has had a deep impact on humans' computer-mediated communication (CMC). With the increasing use of social media platforms such as Facebook, Twitter, or Instagram, people now massively interchange messages and ideas through text-based chat tools that support emoji usage, imbuing these with semantic, emotional, and meaningful meaning. In order to analyze and extract comprehensive knowledge from emoji-embedded message data sets, many methods have been developed through the usage of a multidisciplinary approach, which involves ML along with NLP, psychology, robotics, and so on. Among the tasks developed with ML algorithms for the analysis of emoji usage stand sentiment analysis [5, 19], polarity analysis [10, 20], sentiment lexicon buildage [10], utterance embeddings [21], personality assessment [18], to mention a few. These applications are summarized in **Table 1**.

The following section shows an analysis from the point of view of the use of ML algorithms to support tasks related to the sentiment analysis through the use of emoji, classification, comparison, polarity detection, data preprocessing from tweets with emoji embeddings, and computer vision techniques for video processing to detect facial expression.

3.1 Emoji classification and comparison

In recent years, algorithms such as deep learning (DL) have emerged as a new area of ML, which involve new techniques for signal and information processing. This type of algorithms employ several nonlinear layers for information processing through supervised and unsupervised feature extraction, and transformation for pattern analysis and classification. It also includes algorithms for multiple levels of representation attaining models that can describe the complex relations within data. Particularly, if data sets are considerably large, a deep-learning approach is the best option for reaching a well-trained model regarding if data are labeled or not [25, 26]. Until our days, ML algorithms that use shallow architecture show a good performance and effectiveness for solving simple problems, for instance, linear regression (LR), support vector machines (SVM), multilayer perceptron (MPL) with a single hidden layer, decision trees like random forest or ID3, among others. These architectures have limitations for extracting patterns from a wide complex problem's variety, such as signals, human speech, natural language, images, and sound processing [25]. For this reason, a deep-learning approach allows to solve these limitations showing good results.

Emoji classification and comparison constitute two important tasks for discriminating several kinds of emoji, including those with similar meaning. Deep-learning models have been used for this goal in texts where emoji are embedded, producing better result than softmax methods, such as logistic regression, naive Bayes, artificial neural networks, and others. For example, Xiang Li et al. developed a deep neural network architecture for getting a trained model that could predict the correct emoji for its corresponding utterance [21]. This approach provides the possibility that machines generate an automatic message for humans during a conversation with the use of implicit sentiment and better semantic on ideas.

In Li et al.'s [21] proposal, the system receives as input an utterance set $Y = \{y_1, y_2, \dots, y_n\}$ and an emoji set $X = \{x_1, x_2, \dots, x_n\}$. The main goal is to train a classification model, which could predict the correct emoji for an utterance given.

The architecture used in this work has two parts. The first is a convolutional neural network (CNN) for giving a sentence embedding that represents an

Related papers	Problems addressed	Method	Emoji use	Emoji competence
[21]	Emoji classification correct emoji prediction Matching utterance embeddings with emoji embeddings Emoji for sentiment analysis	One-hot vector Sliding windows CNN Cosine similarity Dynamic Pooling	Emotive Social	Adjunctive
[20]	Sentiment analysis Polarity detection	10-fold cross validation ANN CNN NLP Shallow classifiers: SVM and LR Search-based classifier	Emotive Social	Adjunctive Substitutive Mixed textuality
[22]	Image processing & computer vision to detect facial expression Emoji embeddings	Haar classifier AdaBoost Canny algorithms	Emotive	Adjunctive
[19]	Sentiment analysis Auto-labeling using emoji sentiment Emoji classification Emoticons as a heuristic data Tweets data preprocessing	Tf-idf Word2Vec NLP Ensemble classifiers Deep learning	Emotive Social	Adjunctive
[10]	Sentiment analysis Emoji sentiment map & lexicon Polarity detection Emoji sentiment ranking	Discrete probability distribution approximation Welch's t-test Krippendorffs Alpha-reliability	Emotive Social	Adjunctive Substitutive
[5]	Sentiment analysis Automated analysis of social media contents Emoji classification Correlation analysis among languages	Nearest neighbors	Emotive Social	Adjunctive Substitutive
[23]	Emotions detection Emoji classification	LR SVM Adaptive Boosted Decision Trees (ADT) 10-fold cross validation Random Forests (RF)	Emotive Social	Adjunctive Substitutive
[18]	Emotions representation using emoji Big 5 personality assessment test using emoji	Exploratory Factor Analysis (EFA) Confirmatory Factor Analysis (CFA) Bonferroni correction	Emotive Personality assessment	Adjunctive Substitutive
[9]	Emoji as co-speech element Emoji-based measures	NA	Beat gestures Social	Adjunctive Substitutive
[24]	Facial expressions recognition Emoji embeddings Emoji usage for peer communication Emoji as social cues	Haar classifier AdaBoost	Emotive Social	Adjunctive

Table 1.
Comparative table of the articles analyzed.

utterance, and the second one is the embedding of emoji and this part should be trained. In order to join both parts, a matching structure was created due to embeddings in continuous vector space that could well represent emoji, consequently performing better than discrete softmax classifier.

The bottom of CNN is a word embedding layer for tasks of NLP. This provides semantic information about a word using real vector that represents its features. For an utterance that represent a sequence of words, for each word w_i is a one-hot vector of dictionary dimension, a bit from w_i takes value 1 if it corresponds to word on the dictionary and 0 for remaining bits. In Eq. (1), the embedding matrix is defined such that [21]:

$$E_1 \in R^{D \times V}, \quad (1)$$

where D and V are word embedding and word dictionary dimensions, respectively. Each $e_1(w_i) \in E_1$ is the embedding for word in a dictionary. The convolutional layer uses sliding windows to get information from word embeddings; for this process, the following function is used (see the Eq. (2) [21]):

$$Y_1 = f(W_1[e_1(w_1); e_1(w_2); \dots; e_1(w_t)] + b_1), \quad (2)$$

where t is the size of window and b_1 is the bias vector. Hence, the parameter to be trained is W_1 .

Once obtained a series continuous representations of local features from convolutional layer, theory of dynamic pooling is used for sensitizing these embeddings into one vector of the whole utterance. This produces by output the max pooling. The hidden layer uses the sentence embedding of the utterance obtained as y_2 and returns finally the vector to represent the utterance.

Similarly to the word embedding layer, the emoji embedding layer uses a matrix defined as $E_2 \in R^{D \times V}$ to obtain $e_2(x_i)$, where $\$K\$$ is the one-hot vector's length that represents each emoji x_i . Each $e_2(x_i)$ of E_2 is one parameter of neural network. The process of training is a forward propagation for computing the matching score between the given utterance and the correct emoji, and matching score between the given utterance and the negative emoji. Backward propagation is used to update model parameters. For calculating the matching score, the cosine similarity measure is used, whereas for training the neural network, the Hinge Loss function was used. It is worth mentioning that the latter is very useful for carrying out pairwise comparison to identify similar emoji types.

Finally, the author obtains an architecture that uses a CNN and a matching approach for classifying and learning emoji embeddings. The importance of the aforementioned work regarding the field of robotics is the possibility of producing a facial gesture as a result of the introduction of a statement, conversation, or idea to a machine, employing the semantic and emotional relation of emoji.

3.2 Emoji sentiment analysis

In the area of decision making, it is being relevant to know how the people think and what they will do in the future. These produce the needs of grouping peoples in accordance with their interaction on Internet and social networks. Sentiment analysis or opinion mining is the study of people's opinions, sentiments, or emotions, using an NLP approach, which includes, but is not limited to, text mining, data mining, ML, and deep learning [20]. For instance, the CNN's usage has been employed to predict the tweets' emoji polarities. These techniques have showed to

be more effective than shallow models in image recognition and text classification where they reach better results [19].

Tweets processing for mining opinion and text analysis tasks play a crucial role for different areas in the industry because these produce relevant result for feed-back the design of products and services. As Twitter is a platform where user interactions are very informal and unstructured and people use many languages and acronyms, it becomes necessary to build a model language-independent and nonsupervised learning. We can see the use of emoji or emoticons in this scenario through heuristic labels for a system; for this, the feature's extraction process was developed by unsupervised techniques. The emoji/emoticons are the final result that represents a sentiment that a tweet contains. According to Mohammad Hanafy et al. in order to get a trained model for text processing, it is essential to do a data preprocessing for obtaining the data sets, where noisy elements are removed such as hashtags and other stranger characters like "at," reduction of words by removing duplicated words, and very important, reemphasizing the emoticons with their scores. Each emoticon has a raw data that contain a sentiment classified as negative, neutral, or positive. For each classification, a continuous value is recorded. This representation is used in auto-label phase, for generating the training data using the score for determining emoji [19].

Feature extraction stage uses the Tf-idf approach; it indicates the importance of a word in the text through its frequency in the text or text's set. Using Eq. (3), we can calculate this as follows [19, 27]:

$$TfIDF(t, d, F) = tf(t, d) \cdot \log \frac{n_d}{df(d, t) + 1} \quad (3)$$

where t is the word and d is the tweet. Term frequency in the document is tf , df is document frequency where word exists, and n_d is the number of tweets.

Other feature-extracting methods employed were bag-of-words (BOW) and Word2Vec. BOW selects a set of important words in tweets, and then each document is represented as a vector of the number of occurrences of the selected words. Word2Vec uses a two-layer neural network to represent each word as a vector of certain length based on its context. This feature extraction model computes a distributed vector representation of the word, been its main advantage that similar words are close in the vector space. Moreover, it is very useful for named entity recognition, parsing, disambiguation, tagging, and machine translation. In the area of big data processing, the library Spark ML within the Apache Spark engine uses skip-gram-model-based implementation that seeks to learn vector representation that take into account the contexts in which words occur [27].

Skip-gram model learns word vector representations that are good at predicting its context in the same sentence or sequence of training words denoted as $W = \{w_1, w_2, \dots, w_t\}$, where T is $\|W\|$. The objective function is to maximize the average log-likelihood, which is defined by Eq. (4) [27]:

$$\frac{1}{T} \cdot \sum_{t=1}^T \sum_{j=-k}^{j=k} \log p(w_{t+j} | w_t), \quad (4)$$

where k is the size of training windows. Each w is associated with two vectors u_w as word and v_w as context, respectively. Using Eq. (5), given the word w_j , the probability of correctly predicting w_i is computed as [27]:

$$p(w_i | w_j) = \frac{\exp(u_{w_i}, u_{w_j})}{\sum_{l=1}^V \exp(v_l, v_{w_j})}, \quad (5)$$

where V is the vocabulary length. The cost of computing $p(w_i|w_j)$ is expensive; consequently, Spark ML uses hierarchical softmax with computational cost of $O(\log(V))$ [27].

These feature extractor models were used with other classifiers, such as SVM, MaxEnt, voting ensembles, CNN, and LSTM to extend the architecture of recurrent neural network (RNN). As solution proposal, a weighted voting ensemble classifier is used that combines the output of different models and its classification probabilities. For each model, a different weight when voting is assigned. The proposed model reaches a considerable accuracy in comparison with other models. This approach is very important in scenarios where we need no human intervention and any information about the used language; it is very useful to apply a good combination between classical and deep-learning algorithm to achieve better accuracy [19].

3.3 From video to emoji

As consequence of the semantic meaning that emoji carriers, there are some applications and researches that involve the image processing for generating emoji classification or an utterance with emoji embeddings. For that purpose, Chembula et al. have created an application that receives as input a stream of video or images from a person and create an emoticon based on image face. The solution detects the facial expression at the time that message is being generated. Once that facial expression is detected, the device generates a message with the suitable emoticon [28].

This system performs a facial detection, facial feature detection, and classification task to finally identify the facial expression. Although the initial processing proposed by Chembula and Pradesh [22] was not specified on the general description, we can use open source solutions in order to aim this job.

OpenCV is an open source library for computer vision, and it includes classifiers for real-time face detection and tracing like Haar classifiers and Adaptive Boosting (AdaBoost). We can download trained model for performing this task; the model is an XML file that can be imported inside the OpenCV project. For featuring extraction, the library includes algorithm for detecting region of interest in human face like eyes, mouth, and nose. For this propose, drop information from image stream using gray scale convert and afterward using Gaussian Blur for reducing noise is important. Canny algorithms may be used for tracking facial features with more precision than others like Sobel and Laplace [29].

In [24], Microsoft's emotion API is used as a tool to detect facial images from the Webcam image capture of the computer. Once the image is captured, the detected face is classified into seven emotion tags. Although the process is not specified exactly, the API mentioned works on an implementation of the OpenCV library for .NET [30], so the algorithms used for face detection should be the same as those described above.

For classification task, we can use nearest neighbor classifier, SVM, logistic classifier, Parzen density classifiers, normal density Bayes classifiers, and Fisher's linear discriminant [31]. Finally, when the classification is done, the output layer consists a group of types of emoji according to the meaning for each type of emotion detected in the image face. The importance of this contribution lies in the possibility of introducing new forms of human-computer interaction through the use of emotions. This can be useful for intelligent assistants both physical and visual that are able to react or are current according to the mood of people who use a particular intelligent ecosystem.

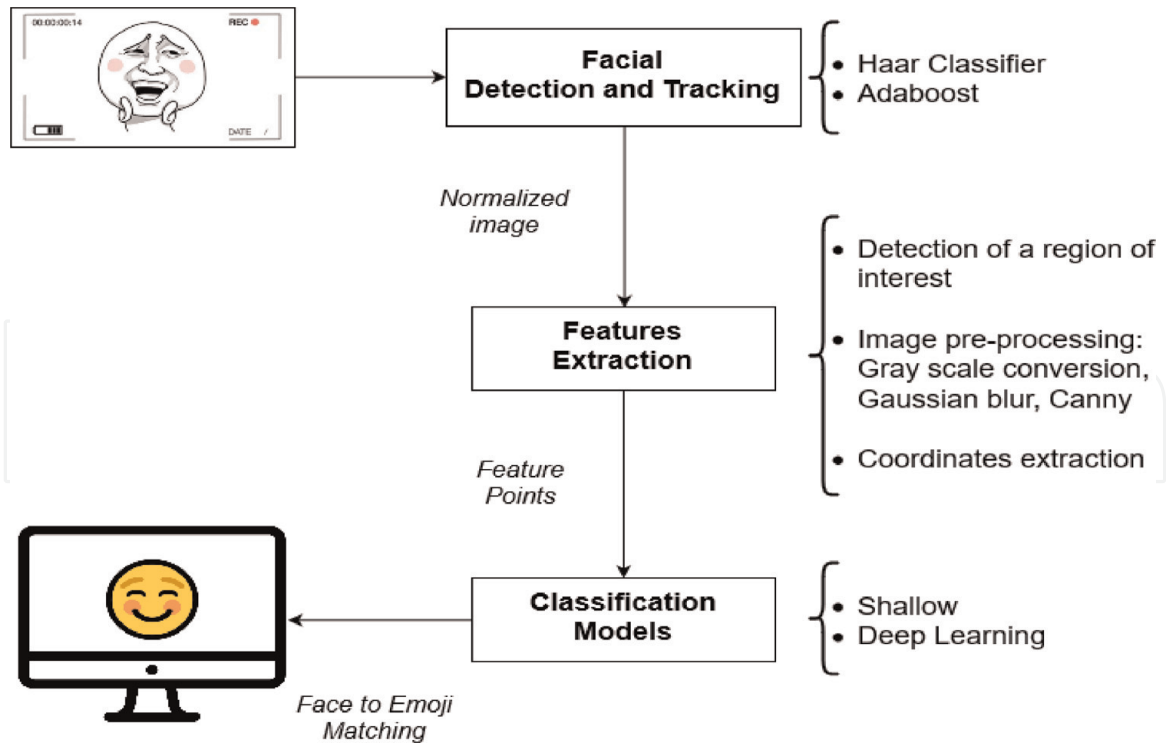


Figure 2.
General process of facial detection and its corresponding classification using emoji.

Figure 2 shows in a general way the operation of what has been explained above.

4. Applications to virtual and embodied robotics

As already mentioned, in this work, our intention is to elaborate the elements that will power an artificial intelligent entity, either virtual or physically embodied, with the capacity to recognize and express (R&E) emotional content using emoji. In this sense, we can collect massive amounts of human-human (and perhaps human-machines too) interactions from multiple Internet sources such as social media or forums, to train ML algorithms, which can R&E emotions, utterances, beat gestures, and even assess personality of the interlocutor. Furthermore, we may even reconstruct text phrases from speech in which emoji are embedded to these to obtain a bigger picture of the semantic meaning. For instance, if we asked the robot “are you sure?” while raising the eyebrows to emphasize our incredulity, we may obtain an equivalent expression such as “are you sure? 😏” Once the models are defined and trained, these will be embedded into the artificial entity, which will be interacting with humans. This conceptual framework is displayed in **Figure 3**. While in a virtual entity such as a chatbot, the inference of emotional states or personality, as well as expressing emotions or beat gestures using emoji, is straightforward, in an embodied entity such as a physical robot that requires a little bit more of elaboration. In the latter, an interlocutor’s emotional or personality first requires the humans’ facial expressions and gestures to be transformed into emoji from video streams or speech similarly as shown in [22]. Then, the same pipeline as the one used for a chatbot may be employed, identifying the corresponding emotional state using pretrained sentiment detection algorithm such as in [20]. Therefore, since both, embodied and virtual artificial entities, can employ the same pipeline, we focus on applications to the former. In particular, we discuss some works, which are delved in this direction, and how the cognitive interaction

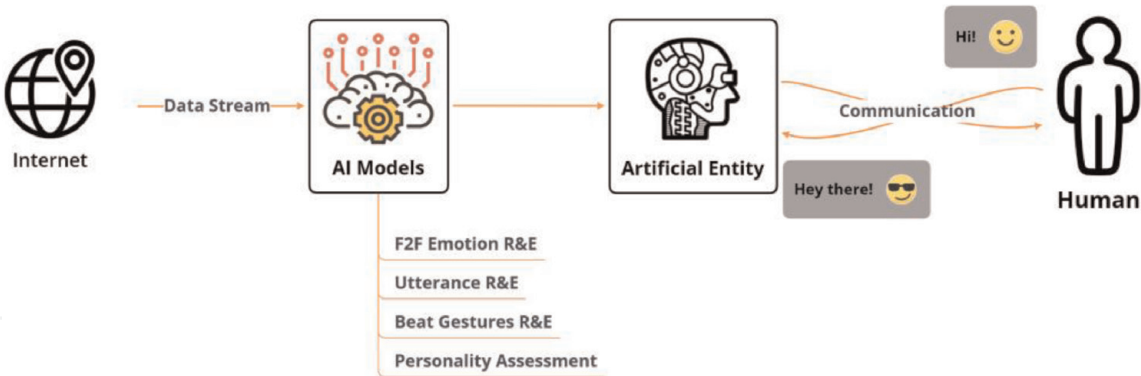


Figure 3.
Emoji emotional communication conceptual framework.

between humans and artificial entities may be improved by modeling the emotional exchange as shaped by emoji usage.

4.1 Embodied service robots study cases

Service robots are a type of embodied artificial intelligent entities (EAIE), which are meant to enhance and support human social activities such as health and elder care, education, domestic chores, among others [32–34]. A very important element for EAIE is improving the *naturalness* of human-robot interactions (HRI), which can provide EAIE with the capacity to R&E emotions to/from their human interlocutors [32, 33].

Regarding the emotional mechanisms of an embodied robot *per se*, a relevant example is the work by [33], which consists in an architecture for imbuing an EAIE with emotions that are displayed in an LED screen using emoticons. Such architecture establishes that a robot’s emotions are in terms of long-medium-short affective states suchlike its personality (i.e., social and mood changes), the surrounding ambient (i.e., temperature, brightness, and sound levels), and human interaction (i.e., hit, pat, and stroke sensors), respectively. All of these sensory inputs were employed to determine EAIE emotional state using ad hoc rules, which are coded into a fuzzy logic algorithm, which is then displayed in an LED face. Facial gestures corresponding to Ekman’s basic emotions expressions are shown in the form of emoticons.

An important application of embodied service robots is the support of elder’s daily activities to promote a healthy life style and providing them with an enriching companion. In such case, a more advanced interaction models for EAIE based on an emotional model, gestures, facial expressions, and R&E utterances are proposed [32, 35–37]. The authors of these works put forward several cost-efficient EIAE based on mobile device technologies namely *iPhonoid*, *iPhonoid-C*, and *iPadrone*. These are robotic companions based on an architecture, which among other features is built upon the informationally structured spaces (ISS) concept. The latter allows to gather, store, and transform multimodal data from the surrounding ambience into a unified framework for perception, reasoning, and decision making. This is a very interesting concept since, not only EAIE behavior may be improved by its own perceptions and HRI but also from remote users’ information such as elder’s activities from Internet or grocery shopping. Likewise, all these multimodal information can be exploited by any family member to improve the quality of his/her relation with the elder ones [36]. Regarding the emotional model, the perception and action modules are the most relevant. Among the perceptions considered in these frameworks stand the number of people in the room, gestures, utterances,

colors, etc. In the same fashion as [33], these EAIE implements an emotional time-varying framework, which considers emotion, feeling, and mood (from shorter to longer emotional duration states, respectively). First, perceptions are transformed into emotions using expert-defined parameters, then emotions and long-term traits (i.e., mood) serve as the input of *feelings* whose activation follows a spiking neural network model [32, 35]. Particularly, mood and feelings are within a feedback loop, which emphasize the emotional time-varying approach. Once perceptions are turned into its corresponding emotional state, the latter is sent to the action module to determine the robot behavior (i.e., conversation content, gestures, and facial expression). As mentioned earlier, EAIE also R&E utterances, which provide feedback to the robot's emotional state. Another interesting feature of the architecture of these EAIE is its conversational framework. In this sense, the usage of certain utterances, gestures, or facial expressions depends on conversation modes, which in turn depends on NLP processing for syntactic and semantic analyses [32, 37]. Nevertheless, with regard to facial and gesture expressions, these works take them for granted and barely discuss both. In particular, how facial expressions are designed and expressed can only be guessed from figures of these EAIE papers, which closely resemble emoji-like facial expressions.

Embodied service robots are also beneficial in the pedagogical area as educational agents [38, 39]. Under this setting, robots are employed in a *learning-by-teaching* approach where students (ranging from kindergarten to preadolescence) read and prepare educational material beforehand, which is then taught to the robotic peer. This has shown to improve students understanding and knowledge retention about the studied subject, increasing their motivation and concentration [38, 40]. Likewise, robots may enhance its classroom presence and the elaboration of affective strategies by means of recognizing and expressing emotional content. For instance, one may desire to elicit an affective state that engages students in an activity or identify boredom in students. Then, robot's reaction has to be an optimized combination of gestures, intonation, and other nonverbal cues, which maximize learning gains while minimizing distraction [41]. Humanoid robots are preferred in education due to their anthropomorphic emotional expression, which is readily available through body and head posture, arms, speech intonation, and so on. Among the most popular humanoid robotic frameworks stand the *Nao*[®] and *Pepper*[®] robots [38–40]. In particular, Pepper is a small humanoid robot, which is provided with microphones, 3D sensors, touch sensors, gyroscope, RGB camera, and touch screen placed on the chest of the robot, among other sensors. Through the *ALMood Module*, Pepper is able to process perceptions from sensors (e.g., interlocutors' gaze, voice intonation, or linguistic semantics of speech) to provide an estimation of the instantaneous emotional state of the speaker, surrounding people, and ambiance mood [42, 43]. However, Pepper communication and its emotional expression is mainly carried out through speech consequence of limitations such as a static face, unrefined gestures, and other nonverbal cues, which are not as flexible as human standards [44], for instance while we consider **Figure 4**, which is a picture displaying a sad Pepper. Only by looking the picture, it is unclear if the robot is sad, looking at its wheels, or simply turned off.

4.2 Study cases through the emoji communication lens

In summary, in the above revised EAIE cases (emoticon-based expression, iPadrone/iPhonoid, and Pepper), emotions are generated through an ad hoc architecture, which considers emotions and moods that are determined by multimodal data. A cartoon of these works is presented in **Figure 5**, displaying on (a) the work



Figure 4.
Is Pepper sad or just shutdown?

of [33] on (b) the work of [32, 35–37], and on (c) Pepper the robot as described in [42–44].

In these cases, we can integrate emoji-based models to enhance the emotional communication with humans, for some tasks more directly than for others. Take for instance, the facial expressions by itself, in the case of (a) and (b), the replacement of emoticon-based emotional expression by its emoji counterpart is straightforward. This will not only improve visually the robot's facial expression but also allowing more complex facial expressions to be displayed such as sarcasm (😏) or co-speech gestures as 🗣️ after making a joke. Another important feature of replacing emoticon-based faces by emoji is that the latter are used mostly to convey positive emotions even when criticizing or giving negative feedback [2]. Therefore, this feature could be really useful for maintaining a perpetual friendly tone of an elder robotic partner (b) or as an educational agent (c).

Regarding the emotional expression of the discussed EAIE, this is contingent to the emotional model, which in the case of (a) and (b) are expert-design knowledge coded into fuzzy logic behavior rules and more complex neural networks, respectively. In both cases, this not only will bias the EAIE into specific emotional states but also will require vast human effort to maintain it. In contrast, Pepper's framework is robuster, includes a developer kit, which allows modifying robot's behaviors and the integration of third party chatbots, performing semantic and utterance analysis, and is maintained and improved by a robotics enterprise. Yet, Pepper's

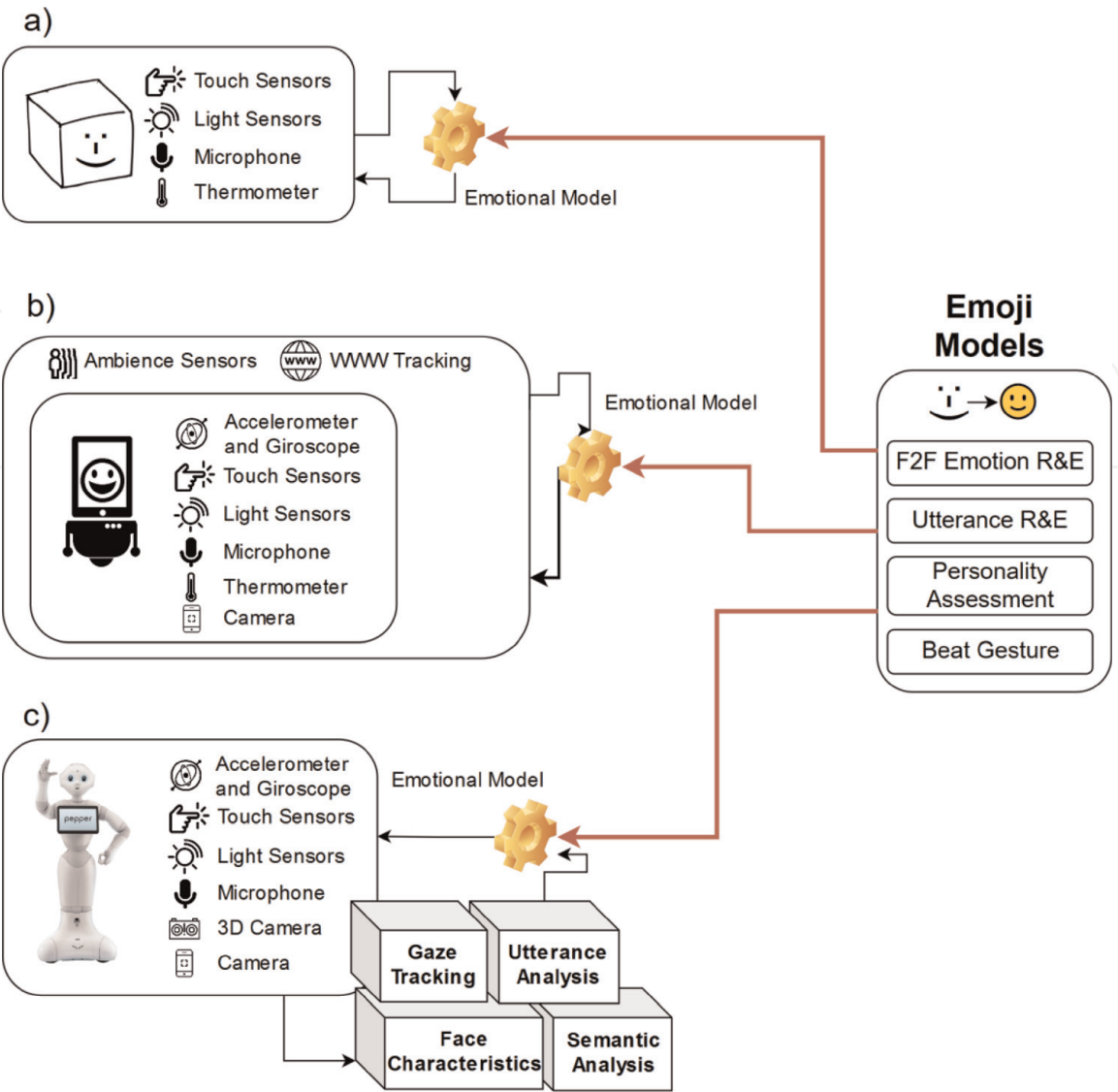


Figure 5.
Case studies using emoji-based modules to improve its emotional R&E models.

emotional communication is constrained by a static face, while it can express emotions by changing the color of its led eyes and adopt the corresponding body posture; its emotional communication is mainly done through verbal expressions. Nevertheless, in a pragmatic sense, do we really need to emulate emotions for a robot to have an emotional communication or is enough to R&E emotions so that a human interlocutor cannot distinguish between man and machine? In this sense, NLP and ML can be used to leverage the emotional communication of a robot by first mapping multimodal data into a discourse-like text where emoji are embedded, and then, using emoji-based models to recognize sentiments, utterances, and gestures so the decision-making module can determine the corresponding message along with its corresponding emoji. In the case of (a), the microphone and in the case of (b), the microphone, camera, and ambient sensors will be responsible for capturing speech and facial expressions that will be converted into a discourse-like text. Once the emotional content of the message is identified, the corresponding emoji shall be displayed. In the case of Pepper, F2F communication can be improved directly by displaying emoji in its front tablet. For instance, when Pepper starts waving to a potential speaker, a friendly emoji such as a waving hand 🙌 or a greeting smile 😊 shall be portrayed in the tablet. Likewise, emoji usage as utterances and beat gestures can be employed by Pepper to avoid silences in a goofy manner (😜), to indicate a lack of knowledge about a particular topic (🤔), or to emphasize politeness when asking an interlocutor for an action (🙏🙏🙏🙏).

5. Discussion

Emotional communication is a key piece for enhancing HRI, after all it will be very useful if our smart phones, personal computers, cars or busses, and other devices could exploit our emotional information to improve our experience and communication. While nowadays, several proposals for robotic emotional communication are undergoing, emoji as a framework for the latter present a novel approach with high applicability and big usage opportunities. Some of the works presented here discussed the linguistic aspects of emoji, as well as the technical aspects in terms of ML and NLP to R&E emotions, utterances, gestures in texts, which contain emoji. Furthermore, we also presented some related works in the area of HRI, which can easily adopt emoji for imbuing an embodied artificial intelligent entity with the capacity for expressing and recognizing emotional aspects of the communication. On the whole, ML models support these issues, but we do not exclude the important task that involves the processing and transformation of data to reach a suitable input representation for training an appropriate model.

On the other hand, there are several open questions regarding the usage of emoji for emotional communication. For instance, are emoji suitable for the communication of every robotic entity? Emoji are mostly employed in a friendly manner and for maintaining a positive communication. If the objective is to model a virtual human, emoji usage will clearly restrain the spectrum of emotions, which may be detected and expressed due to its knowledge base. An important example to consider is the humanoid robot designed by Hiroshi Ishiguro, *the man who made a copy of himself* [45]. Ishiguro's proposal is that in order to understand and model emotions, we must first understand ourselves. Hence, this humanoid robot, namely Geminoid HI-1, is capable of displaying ultrarealistic human-like behaviors. However, do we really want to interact with service robots, which may have bad personality traits such as been unsociable and fickle, or whose mood can be affected by heat and noise like a human does? Do we really want to interact with service robots, which can be *rude* as a real elderly caretaker could? In this sense, emoji usage for the emotional communication may be best suited when the task at hand (e.g., robotic retail store cashier or an educational agent) requires keeping a friendly tone with the human interlocutor. Another question is, should the entire emoji lexicon be used or be restricted only to the core lexicon, which refers to facial expressions? In an ultrarealistic anthropomorphic robot such as Geminoid HI-1, all hand gestures might be carried out by robot's hands itself, thus it should be unnecessary to even fit a screen for displaying a waving emoji (👋) while greeting. On the contrary, more constrained entities such as a Roomba² or Pepper[®] may clearly be benefited from both core and peripheral emoji lexicons for improving its emotional communication with humans. Also, since most of the emoji knowledge is based on short text messages, multimodal data first need to be converted into their corresponding discourse text message, which is, by itself, an open research question.

Acknowledgements

Author GSB thanks the Cátedra CONACYT program for supporting this research. Author OGTL thanks GSB for his excellent collaboration.

² <https://www.irobot.com/roomba>

IntechOpen

Author details

Guillermo Santamaría-Bonfil^{1*} and Orlando Grabiél Toledano López²

1 CONACYT-INEEL, National Institute of Electricity and Clean Energies,
Cuernavaca, Morelos, Mexico

2 University of Informatics Sciences, La Habana, Cuba

*Address all correspondence to: guillermo.santamaria@ineel.mx

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Hurlburt G. Emoji: Lingua franca or passing fancy? *IT Professional*. 2018; **20**(5):14-19
- [2] Danesi M. *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. Bloomsbury Academic: UK; 2016
- [3] Skiba D. Face with tears of joy is word of the year: Are emoji a sign of things to come in health care? *Nursing Education Perspectives*. 2016;**37**(1): 56-57. Available from: <http://insights.ovid.com/crossref?an=00024776-201601000-00015>
- [4] Wiseman S, Gould S. Repurposing emoji for personalised communication. In: *2018 CHI Conference on Human Factors in Computing Systems*. Montréal, QC: ACM; 2018. pp. 1-10
- [5] Barbieri F, Kruszewski G, Ronzano F, Saggion H. How cosmopolitan are emojis?: Exploring emojis usage and meaning over different languages with distributional semantics. In: *Proceedings of the 2016 ACM Multimedia Conference*. 2016. pp. 531-535
- [6] Alshenqeeti H. Are emojis creating a new or old visual language for new generations? A socio-semiotic study. *Advances in Language and Literary Studies*. 2016;**7**(6):56-69
- [7] Ai W, Lu X, Liu X, Wang N, Huang G, Mei Q. Untangling emoji popularity through semantic embeddings. In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media—ICWSM '17* [Internet]. 2017. pp. 2-11. Available from: <https://aiwei.me/files/icwsm2017-ai.pdf>
- [8] Kerslake L, Wegerif R. The semiotics of emoji: The rise of visual language in the age of the internet. *Media and Communication*. 2017;**5**(4):75
- [9] McCulloch G, Gawne L. Emoji grammar as beat gestures. *CEUR Workshop Proceedings*. 2018;**2130**:3-6
- [10] Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of emojis. *PLoS One*. 2015;**10**(12): e0144296. DOI: 10.1371/journal.pone.0144296
- [11] Chomsky N. *Aspects of the theory of syntax*. The Philosophical Quarterly. MIT press; 2014;**11**:1-8
- [12] Barbieri F, Ballesteros M, Saggion H. Are emojis predictable? In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* [Internet]. Stroudsburg, PA, USA: Association for Computational Linguistics; 2017. pp. 105-111. Available from: <http://arxiv.org/abs/1702.07285>
- [13] Hussien W, Al-Ayyoub M, Tashtoush Y, Al-Kabi M. On the Use of Emojis to Train Emotion Classifiers. 2019. Available from: <http://arxiv.org/abs/1902.08906>
- [14] Doiron JAG. Emojis: Visual communication in higher education. *PUPIL: International Journal of Teaching, Education and Learning*. 2018;**2**(2):1-11
- [15] Betz N, Hoemann K, Barrett LF. Words are a context for mental inference. *Emotion*. 2019:1-15. DOI: 10.1037/emo0000510
- [16] Guibon G, Ochs M, Bellot P, Guibon G, Ochs M, Bellot P, et al. From emoji usage to categorical emoji prediction. In: *19th International Conference on Computational Linguistics and Intelligent Text*

- Processing (CICLING 2018). Vietnam: Hanoi. p. 2018
- [17] Querengässer J, Schindler S. Sad but true?—How induced emotional states differentially bias self-rated Big Five personality traits. *BMC Psychology*. 2014;2(1):1-8
- [18] Marengo D, Giannotta F, Settanni M. Assessing personality using emoji: An exploratory study. *Personality and Individual Differences*. 2017;112: 74-78. DOI: 10.1016/j.paid.2017.02.037
- [19] Hanafy M, Khalil MI, Abbas HM. Combining Classical and Deep Learning Methods for Twitter Sentiment Analysis. Switzerland: Springer Nat Switz; 2018. pp. 281-292
- [20] Karthik V. Opinion mining on emoji using deep learning techniques. *Procedia Computer Science*. 2018;132: 167-173
- [21] Li X, Yan R, Zhang M. Joint emoji classification and embedding. *Learning*. 2017;1:48-63
- [22] Chembula AB, Pradesh A. Generating Emoticons Based on an Image of Face. Vol. 2. US Patent; 21 February 2017
- [23] Zhang AX, Igo M, Karger D, Facciotti M. Using Student Annotated Hashtags and Emojis to Collect Nuanced Affective States. London: ACM; 2017
- [24] Liu M, Wong A, Pudipeddi R, Hou B, Wang D, Hsieh G. ReactionBot: Exploring the effects of expression-triggered emoji in text messages. *Proceedings of the ACM on Human Computer Interaction*. 2018;2:1-16
- [25] Deng L, Yu D. Deep learning methods and applications. *Foundations and Trends in Signal Processing*. 2014;7: 197-387
- [26] Bishop CM. *Pattern Recognition and Machine Learning*. Cambridge, UK: Springer Science+Business Media, LLC; 2006. 749 p
- [27] Pentreath N. *Machine Learning with Spark*. Birmingham, UK: Packt Publishing; 2015. 338 p
- [28] Chembula AB, Pradesh A. Generating emoticons based on an image of face. Vol. 2. USA. US 9,576,175 B2. 2017
- [29] Baggio DL. *OpenCV 3.0 Computer Vision with Java*. Birmingham, UK: Packt Publishing; 2015. 174 p
- [30] Larsen L. *Learning Microsoft Cognitive Services*. Birmingham, UK: Packt Publishing; 2017. 484 p
- [31] Pelillo M. *Advances in Computer Vision and Pattern Recognition*. LLC: Springer Science+Business Media; 2013. 293 p
- [32] Tang D, Yusuf B, Botzheim J, Kubota N, Chan CS. A novel multimodal communication framework using robot partner for aging population. *Expert Systems with Applications*. 2015;42(9): 4540-4555. Available from. DOI: 10.1016/j.eswa.2015.01.016
- [33] Daosodsai N, Maneewarn T. Fuzzy based emotion generation mechanism for an emoticon robot. 13th International Conference on Control, Automation and Systems (ICCAS 2013). Gwangju; 2013:1073-1078. DOI: 10.1109/ICCAS.2013.6704075. Available from: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6704075&isnumber=6703852>
- [34] Clabaugh C, Mataric M. Robots for the people, by the people: Personalizing human-machine interaction. *Science robotics*. 2015;3(21):1-2
- [35] Yorita A, Botzheim J, Kubota N. Emotional models for multi-modal communication of robot partners. *IEEE International Symposium on Industrial Electronics*. 2013:1-6

- [36] Obo T, Kakudi HA, Yoshihara Y, Loo CK, Kubota N. Lifelog visualization for elderly health care in informationally structured space. In: 2015 4th Int Conf Informatics, Electron Vision, ICIEV 2015. 2015;(March 2017)
- [37] Woo J, Botzheim J, Kubota N. A socially interactive robot partner using content-based conversation system for information support. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2018;22(6):989-997
- [38] Tanaka F, Isshiki K, Takahashi F, Uekusa M, Sei R, Hayashi K. Pepper learns together with children: Development of an educational application. In: 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids). 2015. pp. 270-275
- [39] Lehmann H, Rossi G. Social robots in educational contexts: Developing an application in enactive didactics. *Journal of e-Learning and knowledge Society*. 2019;15:27-41
- [40] Jamet F, Masson O, Jacquet B, Stilgenbauer J-L, Baratgin J. Learning by teaching with humanoid robot: A new powerful experimental tool to improve children's learning ability. *Journal of Robotics*. 2018;2018:1-11
- [41] Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F. Social robots for education: A review. *Journal of Robotics*. 2018;3(21): 1-9. Available from: <https://robotics.sciencemag.org/content/3/21/eaat5954>
- [42] Europe SR. {ALMood} {API}. 2019
- [43] Val-Calvo M, Grima-Murcia MD, Sorinas J, Álvarez-Sánchez JR, de la Paz Lopez F, Ferrández-Vicente JM, et al. Exploring the physiological basis of emotional HRI using a BCI Interface. In: Ferrández Vicente JM, Álvarez-Sánchez JR, de la Paz López F, Toledo Moreo J, Adeli H, editors. *Natural and Artificial Computation for Biomedicine and Neuroscience*. Cham: Springer International Publishing; 2017. pp. 274-285
- [44] Europe SR. How to Create a Great Experience with Pepper. September 2017. Available from: <http://doc.aldebaran.com/> [Last accessed: 17/09/2019]
- [45] Guizzo E. Hiroshi Ishiguro: The Man Who Made a Copy of Himself [Internet]. *IEEE Spectrum*. 2010. Available from: <https://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself>