Dear Author

Please use this PDF proof to check the layout of your article. If you would like any changes to be made to the layout, you can leave instructions in the online proofing interface. Making your changes directly in the online proofing interface is the quickest, easiest way to correct and submit your proof. Please note that changes made to the article in the online proofing interface will be added to the article before publication, but are not reflected in this PDF proof.

If you would prefer to submit your corrections by annotating the PDF proof, please download and submit an annotatable PDF proof by clicking here and you'll be redirected to our PDF Proofing system.

Pattern Recognition Letters xxx (xxxx) xxx

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Enhancing sentient embodied conversational agents with machine learning

Q1 Dolça Tellols[a,*], Maite Lopez-Sanchez[b], Inmaculada Rodríguez[b], Pablo Almajano[b], Anna Puig[b]

[a] *School of Computing, Tokyo Institute of Technology, Meguro Ôokayama 2-12-1, Tokyo 152-8550, Japan*
[b] *University of Barcelona, Gran Via de les Corts Catalanes, 585, Barcelona 08007, Spain*

A R T I C L E I N F O

Article history:
Received 5 July 2019
Revised 11 November 2019
Accepted 25 November 2019
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Embodied conversational agents
Machine learning
Virtual tutors

A B S T R A C T

Within the area of intelligent User Interfaces, we propose what we call Sentient Embodied Conversational Agents (SECAs): virtual characters able to engage users in complex conversations and to incorporate sentient capabilities similar to the ones humans have. This paper introduces SECAs together with their architecture and a publicly available software library that facilitates their inclusion in applications –such as educational and elder-care– requiring proactive and sensitive agent behaviours. In fact, we illustrate our proposal with a virtual tutor embedded in an educational application for children. The evaluation was performed in two stages: firstly, we tested a version with basic textual processing capabilities; and secondly, we evaluated a SECA with Machine-Learning-enhanced user understanding capabilities. The results show a significant improvement in users' perception of the agent's understanding capability. Indeed, the Response Error Rate decreased from 22.31% to 11.46% using ML techniques. Moreover, 99.33% of the participants consider the global experience of talking with the virtual tutor with sentient capabilities to be satisfactory.

1. Introduction

The current surge in the popularity of chatbots has led to a proliferation of platforms that facilitate their design and implementation. Chatbots are non-embodied agents designed for communicating with the user by means of simple conversational interactions. However, although chatbots can be useful, they fall short when aiming to engage the user in longer and more diverse and complex conversations.

The embodiment of Virtual agents, such as Embodied Conversational Agents (ECAs) [5], represents an improvement in user-agent interaction, not only because agent personification facilitates verbal communication, but also because it allows for enriched interaction incorporating non-verbal communication. ECAs are therefore useful for training, guiding, and giving support to users in a more natural way through the use of both natural language and body language. However, ECAs are usually created ad hoc for a specific purpose, hindering their subsequent reuse and the evolution of their functional and structural components.

Against this background, we go a step further in the state of the art, and propose what we call *Sentient Embodied Conversational Agents* (SECAs) as proactive agents endowed with human-like sentient qualities and capable of taking part in complex structured conversations. On the one hand, with the aim of increasing agents' believability, our proposal incorporates sentient capabilities — personality, needs, and empathy — similar to those possessed by humans. On the other hand, our proposal facilitates the implementation of agents capable of taking part in complex structured conversations by covering different types of dialogs (i.e., communication patterns) which can be initiated either at the user's request or proactively. Our SECAs therefore enable seamless transitions between different dialog types that guide users to achieving their goals when engaged in conversations. Furthermore, SECAs are equipped with domain-specific knowledge, memory, and Natural Language Processing (NLP) capabilities which make human-agent interactions more effective. Specifically, a memory module prevents SECAs from being repetitive in their utterances, and an extension of the Artificial Intelligence Mark-up Language (AIML) [27] together with Machine Learning algorithms improve their understanding of users' inputs.

In addition to our SECAs proposal, we define a general computational architecture and provide a software library to create and integrate ECAs into different applications and platforms (mobile,

* Corresponding author.
E-mail address: tellols.d.aa@m.titech.ac.jp (D. Tellols).

https://doi.org/10.1016/j.patrec.2019.11.035
0167-8655/© 2019 Published by Elsevier B.V.

desktop, web). We illustrate our contributions by designing and implementing a virtual tutor called "Earth", which is integrated into a digital application for children in the context of e nergy e fficiency and sustainability [23], with the purpose of making the experience more educational. Initially, we provided 30 children from different schools with a first version of our Earth SECA, which applies simple NLP techniques to gathering conversational data. Subsequently, we evaluated (with another group of 15 schoolchildren) a second version of the agent which incorporates Machine Learning algorithms trained with the previously gathered data. Both tests report a positive impact on the children's perception of learning and their overall conversational experience.

2. Related work

Chatbots are conversational agents originally designed to hold informal conversations (chats) with users. Chatbots are currently receiving a great deal of attention as they are being integrated into applications with the aim of improving users' experience. Examples of chatbots abound on the web: Irene,[1] a chatbot for a railway company; Rinna,[2] a Microsoft chatbot that uses Artificial Intelligence technology to speak like a Japanese secondary school student; and Amy,[3] a well known chatbot for banking are only a few examples. In fact, the mounting interest in chatbots has been accompanied by the emergence of a number of tools for chatbot development, including DialogFlow from Google, wit.ai from Facebook, Watson Assistant from IBM, and LUIS from Microsoft.[4]

However, chatbots have limitations when embedded in applications requiring more complex conversations and human-like properties, which may enhance the believability of the agent and foster user engagement. To overcome those limitations, ECAs like ALMA [10] incorporates personality and emotions by following the Five Factor Model (FFM) [18] and the Ortony, Clore and Collins (OCC) model [20] respectively. Regarding needs, Max [3] is an ECA which implements the concept of boredom to represent the absence of stimuli from the user. Moreover, Kristina [29] is a multilingual virtual assistant for elderly emigrants that uses ontology-based reasoning techniques to structure and adapt conversations to users' cultural background. It is able to recognize a user's emotions by processing audio and video. However, when it comes to the expression of an agent's emotions, these ECAs are merely based on the semantics of the message they utter, whilst our SECA architecture contemplates a holistic model of agent personality including emotions and moods.

To model conversations, ECAs like eCoach [22] make use of Behaviour Trees (BT) [24]. Specifically, this agent helps patients to understand the benefits and drawbacks of alternative treatments for prostate cancer and it is able to express emotions. Other researchers [4] also use BTs and combine them with a cognitive model to implement personality. Alternatively, we propose the use of Finite State Machines to model conversations and their building blocks —the so-called Dialog Types. Furthermore, our modelling allows for the inclusion of additional information, which results in richer conversations.

Regarding virtual tutors, most of them provide domain-specific knowledge. For example, Duolingo Bots,[5] which are devoted to teaching languages, have different personalities. AutoTutor [11] poses challenging problems to students and provides them with feedback. Other initiatives, like the Emote research project [25], which develops robotic tutors for specific tasks, proves the importance of empathy with the user during the interaction.

Natural Language conversations constitute a key component in any ECA (Embodied Conversational Agent). Artificial lligence Markup Language (AIML) is a widely used keyword matching mechanism used to implement chatbots (e.g., A.L.I.C.E. [27]). Other works, such as [7,13,16], include more advanced NLP Modules that use Machine Learning techniques to interact with users. From these, we hightlight [16], which based on the so-called human-centered ML, proposes a hybrid imitation and reinforcement learning method to improve the performance of ML-based conversational systems. Moreover, the recent research on intelligent conversational agents also focuses on personalization to keep more coherent, interesting and engaging conversations. Wang et al. [28] studied personalized persuasive Dialogues and classified 10 different persuasion strategies for social good, i.e., donating to a charity. Furthermore, they found evidence about the relationship of users' psychological backgrounds and persuasion strategies. Also in the line of personalization, Zhang et al. [30] studies the modelling of dialogue agents who ask personality-related questions, remember the answers, and use them naturally in conversations.

Overall, it is worth highlighting that our proposed SECA architecture and its publicly available SW library are conceived to design and integrate conversational agents into any application, while previous works provided particular solutions for specific purposes and domains.

3. Sentient Embodied Conversational Agents (SECA)

A *Sentient Embodied Conversational Agent* (SECA) is defined as an Embodied Conversational Agent (ECA) capable of engaging the user in structured conversations and having some human-like sentient qualities so that it can perceive and "feel" certain aspects and respond to them [26].

Fig. 1 details our publicly available SECA library.[6] It consists of a controller that orchestrates different modules implementing an agent's features. Its design is inspired by a previous work [2]. In particular, we have: (i) redesigned the Personality, Needs and Conversational Modules and (ii) created new Knowledge, Memory, Empathy, and NLP Modules. In what follows, we will briefly introduce and formalize these modules.

Conversational Module allows SECAs to engage in rich conversations with users by supporting different structured conversations in natural language. These conversations facilitate user interaction in applications –such as educational apps, citizens portals, and apps for the elderly– characterized by rich contents and/or complex tasks. Fig. 1 formalizes this module as a tuple composed of a set of *n Conversations* ($Conv_i$), together with a reference to the current one (*currConv*).

Conversations can be tailored to different application contexts by defining them as a set of specific *Dialog Types* ($DT_1, \ldots, DT_m$) that can be *proactive* –if the agent starts the dialog– or upon *user-demand*. Each DT constitutes an interaction template specified by means of a Finite State Machine (FSM) where states are basic conversational states and transitions are conditions depending on previous utterances (see [26]). Thus, for example, the interaction template is different if the dialog revolves around the user asking a FAQ than if the SECA aims to arouse a user's interest. Then, a conversation is defined as a hierarchical FSM such that some states are Dialog Types, which can be reused in different conversations. This leads to natural conversation flows that consider not only the

[1] Irene: http://consulta.renfe.com.
[2] Rinna from Microsoft: https://www.rinna.jp.
[3] Amy from HSBC: https://www.business.hsbc.com.hk/en-gb/everyday-banking/ways-to-bank/innovative-digital-banking-experience.
[4] http://dialogflow.com, https://wit.ai, https://www.ibm.com/watson/, https://www.luis.ai/home.
[5] Duolingo: http://bots.duolingo.com.
[6] SECA software library source code is publicly available at https://github.com/dtelloas/SECA-Library.

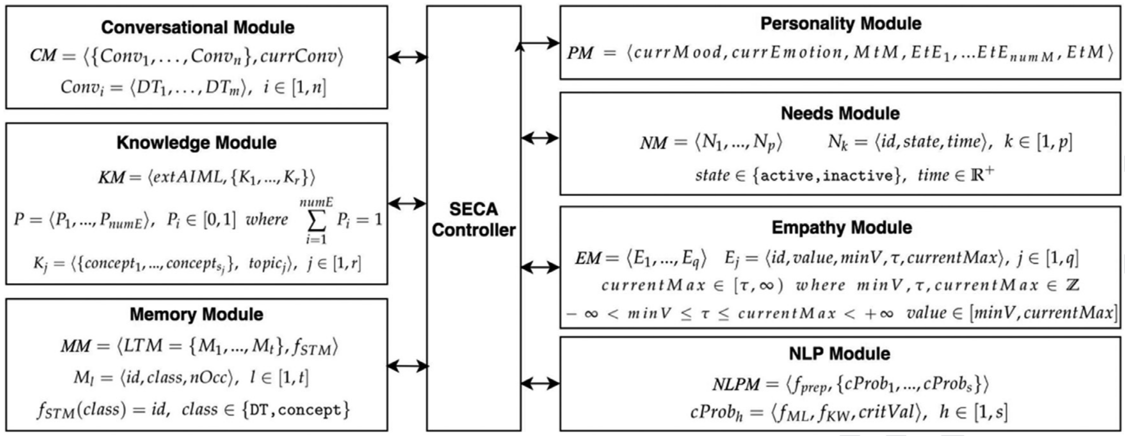

Fig. 1. SECA Library structure and formalization.

user's input, but also information from the Knowledge, Memory, Personality, and Empathy modules.

The **Knowledge Module** manages the static knowledge associated to the targeted conversations. Fig. 1 formalizes it as a tuple of *extAIML* and a set of predefined *Knowledge components* (K_j). We define *extAIML* as an extension of AIML [27] containing the data –utterances and probability vectors (P)– the agent uses to communicate and manage emotions. Each P is associated with an utterance and its dimension equals the number of emotions (*numE*) we define in the Personality Module. A *Knowledge component* K_j is a set of s_j *concepts* related to a *topic*$_j$ (the number s of concepts depends on j because it may be different for each K_j). Thus, for example, the set of concepts {`purple, blue`} are related to *topic* = `color`.

Memory Module is in charge of storing dynamic information to avoid repetitiveness and infuse naturalness to conversations. In particular, it manages Long-Term (*LTM*) and Short-Term (*STM*) memories (see Fig. 1). Memories can be either DTs or knowledge concepts. Each memory M_l has an identifier (*id*), a DT or `concept` *class*, and its number of occurrences (*nOcc*). Hence, *LTM* stores the complete set of t memories, whereas *STM* uses the f_{STM} function to retrieve the most recently used memory of a given *class*.

Personality Module provides the SECA with personality traits. This module follows Kshirsagar et al.'s work [15] in defining personality, moods, and emotions as independent layers. We consider a fixed personality and define different moods (*numM*) and emotions (*numE*), moods being more long-lasting than emotions. For example, a SECA could be in a "neutral" mood while showing different emotions such as "smile" or "cry". Additionally, we also consider that emotions are more tied to user input than moods [14]. Consequently, we first update emotions based on both the current mood (*currMood*) and the user input and we subsequently modify the mood slightly. Thus, if a SECA is in a "happy" mood but the user keeps saying so many sad things that the agent ends up "crying", the mood is more likely to change from "happy" to "sad".

In order to control mood and emotion changes, we propose the use of several transition probability matrices. On the one hand, for each mood k there is an Emotion-to-Emotion transition matrix (EtE_k) where each position defines the probability of one emotion changing to another. To update the agent's current emotion, we linearly combine the values from the matrix of the current mood with the ones from the probabilities vector (P) stored in the *extAIML*. On the other hand, Mood-to-Mood transition matrix (*MtM*) defines the probability of one mood changing to another. There is also a Emotion-to-Mood probability matrix (*EtM*) which stores values that indicate how each emotion slightly affects mood changes. We linearly combine values from both matrices to update the agent's current mood.

The **Needs Module** manages and brings to light the different needs the agents may have. Based on Maslow's Hierarchy of Needs [17] –which assumes self-realization requires the fulfillment of some basic and social needs– it seeks to develop an emotional bond with the user. Thus, for example, a lack of user interaction can activate both the agent's attention need and its proactivity. Moreover, SECAs may have other needs related to specific goals –such as completing certain tasks– of the application embedding the agent. In general, Fig. 1 formalizes the module as a set of p needs, where each N_k is defined in terms of: an identifier (*id*); a *state* signalling whether this need is accomplished (i.e. `inactive`) or not (`active`); and a *time* counter which is reset when a need is accomplished.

The **Empathy Module** tries to guess a user's thoughts and feelings based on their interactions. As for the needs, we include empathy [8] with the aim of establishing stronger bonds with the user. Fig. 1 formalizes SECA's *Empathy* components as a set of q "user state" indicators (E_j) whose (bounded) numerical values (between *minV* and *currentMax*) are monitored along the interaction. We assume E_j's maximum values might be different depending on the user. Indeed, *currentMax* varies in a range defined with a certain τ fixed when initializing each E_j. Variations can occur in real-time depending on the interactions of the user with the SECA. All in all, SECAs can adapt their behaviour depending on empathy values. Thus, for instance, having a `Tiredness` empathy component whose value increases as conversations get longer may urge the SECA to end conversations earlier. And since some users might get tired later than others, the maximum value of the `Tiredness` need could be adjusted accordingly.

The **Natural Language Processing (NLP) Module** contains functions devoted to preprocessing (f_{prep}) and analysing user input through the consideration of different classification problems ($cProb_h$). Since this NLP module is key to our proposal, next section provides further details about the different phases it implements.

4. Natural language processing module

The first phase of the NLP flow is preprocessing (performed through f_{prep}), which becomes necessary to standardize text and to facilitate further manipulation and analysis of the text entered by the user. Input is cleansed and separated into meaningful elements like words (*tokenization*). Then, for the later classification, an N length real-valued vector is associated to the entry of the user (*embedding*). This vector is obtained from the average of the vec-

Please cite this article as: D. Tellols, M. Lopez-Sanchez and I. Rodríguez et al., Enhancing sentient embodied conversational agents with machine learning, Pattern Recognition Letters, https://doi.org/10.1016/j.patrec.2019.11.035

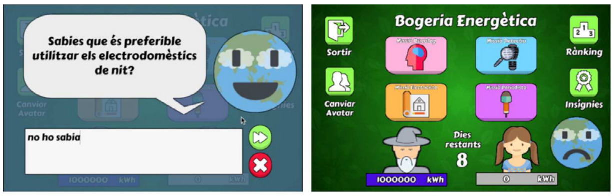

Fig. 2. Left: Earth-User conversation. Right: dashboard.

tors computed for each of the words with *Word2Vec*.[7] If Word2Vec cannot find the embedding of a word, a simple spell checker is applied to it and embedding is retried.

The next phase in the NLP flow is the classification of user's input to better tailor the appropriate response. Different classification problems ($cProb_h$) can be created to discern whether the user's input corresponds, for example, to an affirmative/negative answer, a required explanation, or a question related to a concept. We solve classification problems with a hybrid approach that combines keyword pattern matching and machine learning. These techniques complement each other since keywords may not cover all the words actually uttered but serve well when not enough dialogue data is available. Formally, we define each classification problem as being composed of a *machine learning function* (f_{ML}), a *keyword matching function* (f_{KW}) and certain criteria values (*critVal*) that determine what kind of classification method will be used in each case.

Next Section 5 illustrates previous modules in an educational application and details the specific classification problems that were included. Note that the NLP module in our library provides methods for loading ML models and functions to facilitate the implementation of agents that understand different languages.

5. Educational SECA in the context of energy efficiency

We have integrated our SECA as a virtual tutor in a Cultural Probes[8] application for children [23]. We named this tutor, the Earth.

Since the application focuses on environmental sustainability, we designed the SECA's appearance as an Earth capable of showing moods and emotions through 2D animations, haptic and sound effects. The Earth talks with children (see the left-hand side of Fig. 2) and guides them through the application. The right-hand side of Fig. 2 shows the dashboard screen, which gives access to 4 main tasks for gathering data about the energy consumption habits of their families.[9] Our objectives for creating the Earth agent were twofold: to illustrate the usage of the SECA architecture; and to enhance children's User eXperience by making the application more educative and engaging.

To create the Earth, we used the SECA library shown in Fig. 1. We customized the SECA modules as follows.

Conversational Module makes the Earth proactive in most cases and encourages children to reflect upon topics related to energy. We have designed a total of 6 *Conversations* and 13 *Dialog Types* which, as previously mentioned, are reused in different conversations. More specifically, we designed 5 *proactive* dialog types in which the Earth shares advice related with energy efficiency, reviews energy concepts with the children, asks them if they have any question about energy concepts or the application, and starts a free talk. We also defined 3 *user-demand DTs* so that children can ask the Earth to pose them a question and ask the Earth about energy concepts and the application tasks. The remaining 5 Dialog Types help to complete the tasks in a proactive and engaging way.

The **Knowledge Module** includes a specific application *extAIML* (see Fig. 3 b)) and *Knowledge* components related to both energy, such as $K_j = \langle$ `{hydraulic, wind, nuclear,...}, energy types` $\rangle$, and to the application. Concepts and extAIML content have been selected according to the students' energy-related curricula [9] and to energy efficiency advice provided in some webpages adapted to children and families.[10]

Memory Module stores all necessary *Memories* to remember the `Knowledge` and *DTs* used during the *Conversations*.

The **Personality Module** considers that the Earth has an agreeable personality with 3 moods (happy, neutral, and sad) and 3 emotions (smiling, neutral, and crying). Moreover, as Fig. 3 illustrates, the Earth's Transition Matrices were designed so that the agent focuses more on extreme emotions (happy or sad). By doing so, we intended to make more apparent the impact of the Earth's affective changes on the user experience.

The **Needs Module** manages two different needs. First, we define the need for `Attention` as the necessity in social interaction that aims to increase user engagement. It becomes `active` when students have not interacted with the Earth for more than 8 seconds. Second, `Finish Tasks` corresponds to the necessity of the agent to help the children to accomplish tasks. It becomes `active` after 2 hours of inactivity to remind the kids that they have not yet finished their tasks.

The **Empathy Module** monitors two user state indicators: `Motivation` and `Tiredness`. `Motivation` is computed based on the number of meaningful conversational interactions performed and it is used to adjust how often the Earth appears. Moreover, this module stores children's `Tiredness` to urge the Earth itself to end conversations if they become too long. Initially, conversations are set up to finish if more than 4 user utterances have been received but this value is adjusted in real-time to approximate the length of the last conversation.

The **NLP Module** considers 7 classification problems (see Table 1): 'Y/N', to detect whether the user is making an affirmation, a negation or neither of them; User Explanation ('UE'), to determine whether the user input contains an explanation or not; User-d emand DT ('UDT'), to decide if the user input corresponds to one of the 3 *user-demand DTs* defined in the Earth's Conversational Module; Task ('T'), to find out what task the user is asking about; Energy Concept? ('EC?'), to assess whether the user is asking about an energy concept or not; Energy Concept ('EC'), to determine which energy concept the user is asking about; and Not Energy Concept ('NEC'), to guess the non-energy-related concept the user is asking about. Notice that some of these classification methods are invoked sequentially. Thus, depending of the classification outcome of 'UDT', we may invoke 'T' or 'EC?', and in turn, "EC?" guides the invocation of "EC" or "NEC".

In order to train the ML models, 2263 user messages (written in Catalan) were collected using the first version of the application. Messages were manually annotated and filled in the ML models by balancing the number of training class instances. Thus, "T" got just 78 messages, "EC" 73; and "NEC" 44 (see the last column in Table 1. Initially, we performed a total of 26,640 experiments to find the best model parameters using Grid Search with 10-Fold Cross Validation on 5 different Machine Learning algorithms: Sup-

[7] Word2Vec [19] is a technique which finds the vector representations of very large database words in a short period of time.

[8] Cultural Probes is a user research technique, alternative to interviews, observations and surveys, that gathers data about users by means of tools, artifacts, and tasks that they complete at their own pace.

[9] The 4 tasks were designed as 4 *missions* in a gamified design of the application. Demonstration video available at https://youtu.be/z6otygTaCTo.

[10] https://www.guiainfantil.com/blog/educacion/valores/10-consejos-para-ensenar-a-tus-hijos-a-ahorrar-energia/ https://www.endesaclientes.com/consejos-ahorro/bombillas.html https://www.serpadres.es/familia/tiempo-libre/articulo/como-ahorrar-energia-en-casa.

Please cite this article as: D. Tellols, M. Lopez-Sanchez and I. Rodríguez et al., Enhancing sentient embodied conversational agents with machine learning, Pattern Recognition Letters, https://doi.org/10.1016/j.patrec.2019.11.035

(a)

		newMood		
		happy	*neutral*	*sad*
curr Mood	*happy*	0.8	0.15	0.05
	neutral	0.25	0.5	0.25
	sad	0.05	0.15	0.8

(b)

		newEmotion		
		smile	*neutral*	*cry*
curr Emot.	*smile*	0.8	0.15	0.05
	neutral	0.5	0.4	0.1
	cry	0.1	0.6	0.3

(c)

		Mood Impact		
		happy	*neutral*	*sad*
curr Emot.	*smile*	0.8	0.15	0.05
	neutral	0.25	0.5	0.25
	cry	0.05	0.15	0.8

(d)

```
<category>
<pattern>CONGRATULATE</pattern>
<template>
    [0.6,0.3,0.1]
    Molt bé!%
    Molt bé.%
    Molt bé...%
</template>
</category>
```

Fig. 3. (a) Earth's Mood-to-Mood Transition Matrix (*MtM*), (b) Earth's Emotion-to-Emotion Transition Matrix (*EtE*) corresponding to the happy mood (1 out of 3), (c) Emotion-to-Mood Probability Matrix (*EtM*), (d) extAIML extract containing a probability vector (*P*). For instance, if the current mood is "happy" and the current emotion is "smile", there are more chances of obtaining "smile" as the next emotion when combining the first row of matrix (b) with the probability vector *P* in the extAIML (d). From the "smile" emotion, we would obtain the probabilities of getting a new mood using a combination of the first rows of matrices (a) and (c).

port Vector Machines; Random Forest; Gradient Tree Boosting; Logistic Regression; and Multi-Layer Perceptron (techniques such as Deep Learning have not been considered given the small amount of data available). As Table 1 shows, the algorithms providing better results were Support Vector Machines (SVM) and Multi-Layer Perceptron (MLP), which reached accuracies between 68.00% (for those models having less training data) and 95.40%. Results were also evaluated in terms of precision and recall.

6. Evaluation

In this work, we tested two versions of the application with 10-to-12-year-old volunteer students[11] First, in order to gather user input data, 30 children tested an Earth SECA equipped only with simple keyword-pattern matching techniques (V1). Then, after the ML training, 15 additional children tested the SECA with its enhanced (V2) NLP Module to check whether its performance had improved. Overall, we also aimed to assess user experience.

The evaluation was performed in three stages. First, we presented the project and helped children to install the application on their devices. Subsequently, children used the application at home, at their own pace (i.e., whenever it suited them and for as long they needed) and for a maximum period of 8 days. Finally, we asked them to answer a post-test questionnaire, to elicit their impressions of our Earth SECA.

Fig. 4(a) illustrates how users' perception on the Earth's understanding clearly improved in the second version. Whereas only 20% of V1 users considered the Earth always or almost always understood them, this percentage increased up to 60% for V2 users. Fisher Exact Test (with p-value $= 0.0167 < 0.05$, odds ratio $= 6.0$) confirms the significance of this difference.

[11] A consent form was signed by parents, who were informed about data anonymity and the use of data only for research purposes. We followed evaluation standards and ethical guidelines along the evaluation.

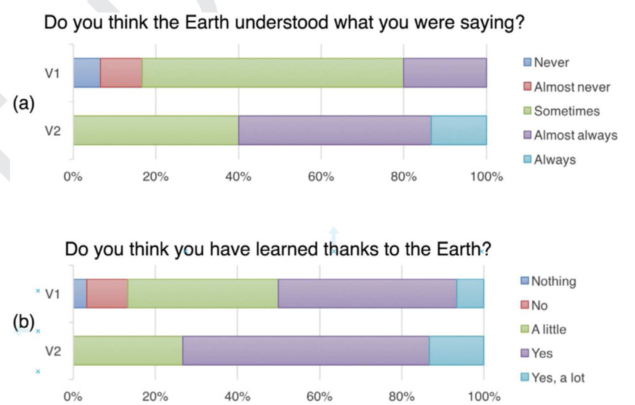

Fig. 4. (a) Perception of the Earth's understanding and (b) perception of the obtained knowledge in both versions of the application: V1, with just keyword-pattern matching; and V2, machine learning enhanced.

Furthermore, Fig. 4(b) confirms that children perceive that they have learned something by talking to the Earth (i.e., they replied that they thought they had learned a little or more) in both versions (86.67% in V1, and 100% in V2). Nevertheless, the difference in this case is not significant (p-value $= 0.2847 > 0.05$), possibly because we did not include additional educational content.

Table 2 provides overall data on user-SECA interactions, which resulted in a total of 7996 utterances. From these, the high number of user utterances is remarkable, although the large standard deviation of its average shows kids engaged unevenly in conversations.

Table 1
Chosen ML model (together with its achieved scores and parameters) for each Classification Problem.

Classification problem	Chosen model	Parameters								Achieved scores			Training data
		kernel	C	gamma	hidden_layer_sizes	max_iter	activation	solver	alpha	Accuracy	Precision	Recall	
Y/N	SVM	rbf	10	0.01	–	–	–	–	–	0.9527	0.95	0.95	2263
UE	SVM	rbf	100	0.01	–	–	–	–	–	0.9308	0.93	0.93	2263
UDT	MLP	–	–	–	250	1000	relu	adam	0.001	0.9540	0.95	0.95	2263
T	MLP	–	–	–	100	1000	tanh	adam	0.001	0.8625	0.86	0.86	78
EC?	SVM	rbf	10	0.001	–	–	–	–	–	0.8705	0.87	0.87	117
EC	MLP	–	–	–	100	1000	identity	lbfgs	0.0001	0.6857	0.63	0.68	73
NEC	MLP	–	–	–	250	1000	relu	sgd	0.01	0.6800	0.65	0.68	44

Table 2
Children-Earth interaction related data.

Data	V1-30 participants	V2-15 participants
Total number of messages	6010	1986
Number of user messages	2263 (37.65%)	698 (35.15%)
Avg. number of messages (SD)	75.43 ($\sigma = 40.70$)	46.53 ($\sigma = 20.02$)
Number of Earth messages	3747 (62.35%)	1288 (64.85%)
Number of *Conversations*	1484	590
Number of *DTs*	165	744
Times users called the Earth	591	180
Transitions to *user-demand DTs*	58	13

Table 3
Analysis of the Earth's incorrect answers with respect to user messages.

Type of mistake	Number of messages (Proportion w.r.t. Total)	
	VI.	V2
Under-performing	245 (10.82%)	37 (5.30%)
Out-of-domain	260 (11.49%)	43 (6.16%)
Total Mistakes (TM)	505	80
RER (TM/Total user messages)	22.31%	11.46%

[Earth] Do you want to know the meaning of an energy related concept? [User] Yes
[Earth] Of what concept do you want to know its meaning? [User] Sustainable
[Earth] Sustainability consist on using machines responsibly, so that energy sources can be preserved. Do you want to talk or ask me something? [User] Ok
[Earth] So tell me! [User] What is efficiency?
[Earth] Energy efficiency is achieved when something is done to reduce energy consumption.

Fig. 5. English translation of an Earth-User interaction example.

Indeed, post-test questionnaires show that, mainly in V1, children who uttered the fewest messages considered the Earth scarcely understood them.

As for the number of conversations (see 3rd row in Table 2), they are quite similar to the number of dialogs (*DT*s), which means that conversations were rather short. The reason is fourfold: conversations were rather short by design; there were failures in the detection of transitions to *user-demand* DTs; Earth's empathetic behaviour ended conversations early to mitigate user tiredness; some kids were not engaged enough in the conversational experience. However, the two last rows of Table 2 indicate that children required the attention of the Earth numerous times. Indeed, each child voluntarily called the agent an average of $(591 + 180)/(30 + 15) = 17.13$ times.

Table 3 analyses the Earth's incorrect answers and groups them in two categories. The first row corresponds to *Under-performing* mistakes that are due missclassification in the NLP module. For instance, a child once asked for the definition of "solar energy" and the Earth provided the definition of "energy" instead. The second row accounts for *Out-of-domain* mistakes due to several reasons, such a: the user's input being unintelligible; the Knowledge module's lacks of related information (e.g., a child asked about Earth's favourite colour); or no dialog types were designed to establish the kind of conversation initiated by the user (e.g., comparative questions like "What uses more energy, the fridge or the computer?"). Overall mistakes are evenly distributed into these two categories for both versions (48.51% of under-performance in V1 and 46.25% in V2). However, the total number of errors has significantly decreased (we performed a Fisher Exact Test [1] –p-value < 0.00001, odds ratio $= 2.220$– to confirm the significance of the difference), causing the Response Error Rate (RER) to change from 22.31% in V1 to 11.45% in V2. Notice that an odds ratio value of 2.220 reveals that the odds of a decrease of the response error rate in V2 are 2.220 times larger than the estimated odds for V1 [12].

The correctly processed misspelled words illustrate the RER decrease in the ML version. As an example, when a child wrote the utterance "betrgia wolar", which is misspelled, asking for the definition of "energia solar" (solar energy), the system provided the correct answer. This would have been impossible using only the keywords approach. Furthermore, quantitative results are aligned with qualitative results gathered from the questionnaires, where some children expressed that "except for a few occasions, the Earth mostly understood me". Indeed, even a student went further in her comment and reasoned that this was mainly due to the fact that "she expressed herself quite well".

As for the different classification problems, Table 4 illustrates the results of testing the enhanced NLP module in V2, which combines machine learning and keyword matching functions by means of certain criteria values (*critVal*). Overall, we can observe that algorithms with a large amount of training data (i.e., 'Y/N', "UE" and "UDT" in Table 1) present high accuracy (above 93.93%). Moreover, 'EC?' and 'EC' prove how, when ML models do not work as expected, it is good to take keyword-pattern matching into consideration by adjusting the criteria values. Indeed, 'T' and 'NEC' problems only reach an accuracy of 44.68% and 65.22% respectively, which may have increased if Keyword pattern matching had been chosen more often. Fig. 5 illustrates a number of interactions that occurred between a child and the Earth SECA. Thus, for example, when the user utters "What is efficiency?" in the fifth line, first, 'UDT' detects that the user is actually asking about a concept and then, 'EC?' determines that it is related to energy. Finally, 'EC' identifies "energy efficiency" as the specific energy concept the child is asking about.

Finally, we evaluate user engagement by asking children if they had enjoyed the experience with our ML-based version of the SECA. As a result, 93.33% of participants' answers score 3 (out of 5) or more. When considering to compare this performance with other applications, we should take into account that this comparison can only be made in very general bases, since interaction dynamics are application specific and post-test questionnaires may vary. Thus, for example, [21] used ECAs in the context of secure access to remote home automation control and [6] report the results for an ECA devoted to travel assistance. In both cases, users where asked about their expected future use of the system whereas we asked if they had enjoyed the experience. Therefore, our question may be interpreted as being more restrictive than the ones asked in these other works since a user may still be open to interact with

Table 4
NLP Module with Keyword matching and ML integration results.

Classification problem	Analysed messages	Total accuracy	Predominant method (% used times)
Y/N	282	96.81%	ML (91.13%)
UE	494	93.93%	ML (71.46%)
UDT	712	97.75%	ML (97.47%)
T	47	44.68%	ML (68.09%)
EC?	59	100%	Keywords (96.61%)
EC	4	100%	Keywords (100%)
NEC	46	65.22%	ML (65.22%)

an ECA even if he or she has not enjoyed the experience. The results reported in these works were, respectively, 4.2 and 2.56 in a Likert-type, 5-point format. Thus, considering that the average scores gathered in our test correspond to a 3.93 out of 5, they seem to be aligned with (or even improve) those of related literature.

Additionally, almost all children noticed that the Earth showed emotions and preferred to talk when it was in a certain mood. Indeed, most children preferred to talk to the Earth when it was 'Happy' because, as reflected in the questionnaire answers, they felt that the Earth's happiness meant it liked their answers or simply caused them to "feel better". However, some children preferred to talk to the Earth when it was sad, though they also explained that the reason was to "make the Earth happy".

7. Conclusions and future work

In this paper we have introduced Sentient Embodied Conversational Agents as virtual characters able to engage users in complex conversations and incorporate sentient qualities similar to those possessed by humans. Our proposal includes the formalization of a SECA library that facilitates their inclusion in applications requiring proactive and sensitive agent behaviours.

We illustrate our proposal by embedding a virtual tutor (the Earth) in an educational application in the context of energy efficiency. First, we evaluated a version of the agent (V1) which only used keyword pattern matching techniques to analyze user input. We also used V1 to collect data to train Machine Learning (ML) algorithms for the classification problems in the NLP Module of the agent embedded in the second version of the application (V2).

The evaluation results of the agent enhanced with Machine Learning models show that most of the participants consider that they had learned while interacting with the Earth (86.65% in V1 and 100% in V2). Additionally there was a significant increase (p-value $= 0.0167 < 0.05$) in the perception of the Earth's understanding (the number of users considering that the Earth understood them always or almost always increased from 20% in V1 to 60% in V2). Overall, participants were satisfied (93.33% of users enjoying the experience) and these results also corroborate our vision that endowing the agents with human-like features such as personality, needs, and empathy increases user bonding, with children calling the Earth an average of 17.13 times.

As future work, modules such as the Conversational one could be enriched through the addition of new conversations and dialog types. Moreover, the Personality module could consider agents with different personalities depending on the user. Regarding the NLP module, though the new system has reduced the Response Error Rate from 22.31% to 11.46%, there is still room for improvement by using other techniques, such as deep learning, which require larger amounts of data.

Declaration of Competing Interest

None.

Acknowledgments

This research is partly funded by research Projects 2017-SGR-341 and MISMIS-Language (PGC2018-096212-B-C33). We also want to thank Rubí-Brilla, schools Montessori, 25 de Setembre, and Escola del Mar, as well as the children and their families.

References

[1] A. Agresti, M. Kateri, Categorical Data Analysis, Springer, 2011.
[2] P. Almajano, D. Tellols, I. Rodriguez, M. Lopez-Sanchez, Meto: a motivated and emotional task-oriented 3d agent, in: Recent Advances in Artificial Intelligence Research and Development, 300, IOS Press, 2017, pp. 263–268.
[3] C. Becker, S. Kopp, I. Wachsmuth, Simulating the Emotion Dynamics of a Multimodal Conversational Agent, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 154–165. 10.1007/978-3-540-24842-2_15
[4] T. Bosse, S. Provoost, Integrating conversation trees and cognitive models within an eca for aggression de-escalation training, in: PRIMA: Principles and Practice of MAS, 2015, pp. 650–659.
[5] J. Cassell, Embodied conversational agents: representation and intelligence in user interfaces, AI Mag. 22 (4) (2001) 67.
[6] R. Catrambone, J. Stasko, J. Xiao, Eca as user interface paradigm: experimental findings within a framework for research, From brows to trust: Evaluating embodied conversational agents (2005) 239–267.
[7] A. Celikyilmaz, L. Deng, D. Hakkani-Tür, Deep learning in spoken and text-based dialog systems, in: Deep Learning in Natural Language Processing, Springer, 2018, pp. 49–78.
[8] N.D. Feshbach, S. Feshbach, Empathy and education, Social Neurosci. Empathy 85 (2009) 98.
[9] A. García-Carmona, A.M. Criado, Enseñanza de la energía en la etapa 6–12 años: un planteamiento desde el ámbito curricular de las máquinas, Enseñanza de las Ciencias 31 (3) (2013) 0087–102.
[10] P. Gebhard, Alma: A layered model of affect, in: Autonomous Agents and Multiagent Systems conference, in: AAMAS '05, ACM, New York, USA, 2005, pp. 29–36.
[11] A.C. Graesser, S. Lu, G.T. Jackson, H.H. Mitchell, M. Ventura, A. Olney, M. Louwerse, Autotutor: a tutor with dialogue in natural language, Beha. Res. Methods Instrum. Comput. 36 (2) (2004) 180–192.
[12] C.K. Haddock, D. Rindskopf, W.R. Shadish, Using odds ratios as effect sizes for meta-analysis of dichotomous data: a primer on methods and issues., Psychol. Methods 3 (3) (1998) 339.
[13] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi, H. Sugiyama, T. Hirano, T. Makino, Y. Matsuo, Towards an open-domain conversational system fully based on natural language processing, in: Conference on Computational Linguistics, 2014, pp. 928–939.
[14] D. Hume, Emotions and moods, Organ. Behav. (2012) 258–297.
[15] S. Kshirsagar, A multilayer personality model, in: Proceedings of the 2nd International Symposium on Smart Graphics, ACM, 2002, pp. 107–115.
[16] B. Liu, G. Tur, D. Hakkani-Tur, P. Shah, L. Heck, Dialogue learning with human teaching and feedback in end-to-end trainable task-oriented dialogue systems, arXiv:1804.06512, (2018).
[17] Maslow, A theory of human motivation, Psychol. Rev. 50 (4) (1943) 370.
[18] R.R. McCrae, O.P. John, An introduction to the five-factor model and its applications, J. Pers. 60 (2) (1992) 175–215.
[19] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv:1301.3781, (2013).
[20] A. Ortony, G.L. Clore, A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1990.
[21] D. Pardo, B.L. Mencia, Á. H. Trapote, L. Hernández, Non-verbal communication strategies to improve robustness in dialogue systems: a comparative study, J. Multimodal User Interfaces 3 (4) (2009) 285–297.
[22] S. Robertson, R. Solomon, M. Riedl, T.G. et al., The visual design and implementation of an embodied conversational agent in a shared decision-making context (ecoach), in: Learning and Collaboration Technologies, 2015, pp. 427–437.
[23] K. Samso, I. Rodríguez, A. Puig, D. Tellols, F. Escribano, S. Alloza, From cultural probes tasks to gamified virtual energy missions, in: Proc. British Computer Society HCI Conf, 2017, p. 79.
[24] Y.A. Sekhavat, Behavior trees for computer games, Int. J. Artif. Intell. Tools 26 (02) (2017) 1730001.
[25] S. Serholt, W. Barendregt, T. Ribeiro, G. Castellano, A. Paiva, A. Kappas, R. Aylett, F. Nabais, Emote: Embodied-perceptive tutors for empathy-based learning in a game environment, in: European Conference on Games Based Learning, 2013, p. 790.
[26] D. Tellols, M. López-Sanchez, I. Rodríguez, P. Almajano, Sentient embodied conversational agents: architecture and evaluation, Artif. Intell. Res. Dev. 308 (2018) 312.
[27] R. Wallace, The Elements of AIML Style, Alice AI Foundation, 2003.
[28] X. Wang, W. Shi, R. Kim, Y. Oh, S. Yang, J. Zhang, Z. Yu, Persuasion for good: towards a personalized persuasive dialogue system for social good, arXiv:1906.06725, (2019).
[29] L. Wanner, E. André, J. Blat, S.D. et al., Kristina: a knowledge-based virtual conversation agent, in: Advances in Practical Applications of Cyber-Physical MAS, 2017, pp. 284–295.
[30] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, J. Weston, Personalizing dialogue agents: I have a dog, do you have pets too?, arXiv:1801.07243, (2018).

Please cite this article as: D. Tellols, M. Lopez-Sanchez and I. Rodríguez et al., Enhancing sentient embodied conversational agents with machine learning, Pattern Recognition Letters, https://doi.org/10.1016/j.patrec.2019.11.035