Negobot: Detecting paedophile activity with a conversational agent based on game theory

Carlos Laorden[1], Patxi Galán-García[1], Igor Santos[1], Borja Sanz[1], Jose María Gómez Hidalgo[2], Javier Nieves[1] and Pablo G. Bringas[1]

[1]S^{3}Lab - DeustoTech Computing, University of Deusto
Avenida de las Universidades 24, 48007 Bilbao, Spain
{claorden,patxigg,isantos,borja.sanz,
jnieves,pablo.garcia.bringas}@deusto.es

[2]Optenet
Madrid, Spain
jgomez@optenet.com

Abstract. Children have been increasingly becoming active users of the Internet and, although any segment of the population is susceptible to falling victim to the existing risks, they in particular are one of the most vulnerable. Thus, some of the major scourges of this cyber-society are paedophile behaviours on the Internet, child pornography or sexual exploitation of children. In light of this background, Negobot is a conversational agent posing as a child, in chats, social networks and other channels suffering from paedophile behaviour. As a conversational agent, Negobot, has a strong technical base of Natural Language Processing and information retrieval, as well as Artificial Intelligence and Machine Learning. However, the most innovative proposal of Negobot is to consider the conversation itself as a game, applying game theory. In this context, Negobot proposes, first, a competitive game in which the system identifies the best strategies for achieving its goal, to obtain information that leads us to infer if the subject involved in a conversation with the agent has paedophile tendencies, while our actions do not bring the alleged offender to leave the conversation due to a suspicious behaviour of the agent.

Keywords: conversational agent, game theory, natural language processing

1 Introduction

Children have been turning into active users of the Internet and, despite every segment of the population is susceptible to become a victim of the existing risks, they in particular are the most vulnerable. In fact, one of the major scourges on this cyber-society is paedophile behaviour, with examples such as child pornography or sexual exploitation.

Some researchers have tried to approach this problem with automatic paedophile behaviour identifiers able to analyse children and adult conversations using automatic classifiers either encoding manual rules [?] or by Machine Learning [?]. These kind of conversation analysers are also present in commercial systems.[1]

In this context, "Negobot: A conversational agent based on game theory for the detection of paedophile behaviour" seeks to identify these types of actions against children. With this particular goal in mind, our system is a chatter bot that poses as a kid in chats, social networks and similar services on the Internet. Because Negobot is a chatter bot, it uses natural language processing (NLP), information retrieval (IR) and Automatic Learning. However, the most innovative proposal of Negobot is to apply game theory by considering the conversation a game. Our system observes the conversation as a competition where our objective is to obtain as much information as possible in order to decide whether the subject who is talking with Negobot has paedophile tendencies or not. At the same time, our actions (e.g., phrases and questions) may lead to the alleged aggressor to leave the conversation or even to act discretely.

As a major difference with others work, Negobot involves both a bot resembling the behaviour of a child and acting as a "hook" to Internet predators, and a game theory-based strategy for identifying suspicious speakers. Summarising, the main contributions of the Negobot system are:

- A structure of seven chatter-bots with different behaviours, reflecting the different ways of acting depending on the conversation state.
- A method to translate the SMS-like wording that maps SMS terms to common language words, allowing the system to understand and respond in a more colloquial way to hide the real nature of the chatter-bot.
- A system to identify and adapt patterns within the conversations to maintain conversation flows closer to conversation flows between real persons.
- An evaluation function to classify the current conversation, in real time, to provide the chatter bot with specific information to follow an adequate conversation flow.

The remainder of this paper is organised as follows. Section 3 describes, in a technical and detailed manner, the methodology used to develop the Negobot system. Section 4 shows the obtained results. Finally, section 5 discusses the major issues of our system, proposing possible improvements, and outlines the avenues of further work.

2 Related work

In 1950, Alan Touring proposed the Imitation Game, as a way of considering the question, "Can machines think?". Without going into deeper detail of the meanings of "machines" and "think", he predicted that by the year 2000 with the evolution that computing machines would experiment, a computer program

[1] E.g. Crisp Thinking provides such a service to online communities: http://www.crispthinking.com/

would be able to fool an average human for 5 minutes about 70% of the time [?].

In 1991, Dr. Hugh Loebner, in conjunction with the National Science Foundation and the Sloan Foundation created the Loebner Prize in Artificial Intelligence (AI): "the first formal instantiation of a Turing Test". The Loebner Prize is a contest between computer programs to identify the most human-like programs and eventually award them with $100,000 and a gold medal [?]. Supported by some [?] and criticised by others [?], the Loebner Prize has been the reference for many researchers in the AI and NLP community for many years.

There are, at least, 4 relevant projects in the area of chatter bots. The first one was *Eliza*. It was built by Joseph Weizebaum and released to the public in 1966 [?]. Eliza was developed to play the role of a psychotherapist and would simply rearrange a submitted statement or question into a new question to ask the person talking to her, using keyword recognition.

The second approach was *Parry*, developed by Kenneth Colby in 1971 [?]. Parry is a natural-language-boosted program that simulates the thinking of a paranoid individual, which was tested in the early seventies using a variation of the Turing Test, an analysis by a group of experienced psychiatrists, obtaining positive results.

These chat bots were completely dependant on a fully human-edited database, they had no self learning AI. Their creators literally spent hundreds of hours, if not more, to continually add content to their knowledge database.

Another interesting project was *Jabberwacky*. Rollo Carpentor developed it in 1988 and published it on-line in 1997 [?]. Jabberwacky overcame the limitation of a handmade database with an approximation based on AI. Jabberwacky learnt during the conversations, modelling the language and the context, using AI algorithms. This conversational agent won the bronze medal in the Loebner.[2] contest twice, in 2005 and 2006

Finally, *ALICE* was developed by Dr. Richard Wallace in 1995 [?], using the Artificial Intelligence Markup Language (AIML) [?]. As Eliza, Parry and Jabberwacky, ALICE was based on the expert knowledge, but it was not designed for a specific role. Besides, in order to enhance its knowledge, the system connected to various data sources such as train timetables, weather forecasts and translations services. This bot won the Loebner Bronze medal three times, in 2000, in 2001 and in 2004.

3 Negobot architecture

Negobot includes the use of different NLP techniques, chatter-bot technologies and game theory for the strategical decision making. Finally, the glue that binds them all is an evaluation function, which in fact determines how the child emulated by the conversational agent behaves.

When a new subject starts a conversation with Negobot the system is activated, and starts monitoring the input from the user. Besides, Negobot registers

[2] http://www.loebner.net/Prizef/loebner-prize.html

the conversations maintained with every user for future references, and to keep a record that could be sent to the authorities in case of determining that the subject is a paedophile. Fig. 1 offers a functional flow of Negobot.

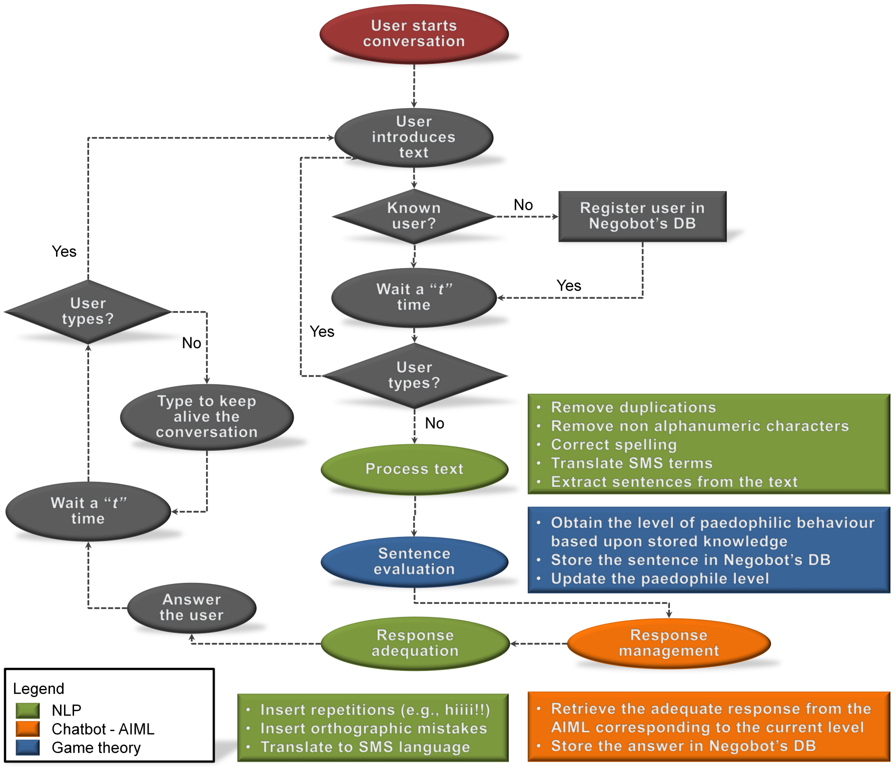

User starts conversation
User introduces text
Known user?
No
Register user in Negobot's DB
Yes
Wait a "t" time
Yes
User types?
No
Yes
User types?
No
Type to keep alive the conversation
Wait a "t" time
Process text
Remove duplications
Remove non alphanumeric characters
Correct spelling
Translate SMS terms
Extract sentences from the text
Sentence evaluation
Obtain the level of paedophilic behaviour based upon stored knowledge
Store the sentence in Negobot's DB
Update the paedophile level
Answer the user
Response adequation
Response management
Legend
NLP
Chatbot - AIML
Game theory
Insert repetitions (e.g., hiiii!!)
Insert orthographic mistakes
Translate to SMS language
Retrieve the adequate response from the AIML corresponding to the current level
Store the answer in Negobot's DB

Fig. 1. Functional flow of Negobot.

3.1 Processing the conversation

Our AI system's knowledge came from the website Perverted Justice.[3] This website offers an extensive database of paedophile conversations with victims, used in other works [?,?,?]. A total of 377 real conversations were chosen to populate our database (henceforth they will be called *assimilated conversations*).

Besides, Perverted Justice users provide an evaluation of each conversation's seriousness by selecting a level of "slimyness" (e.g., dungy, revolting thing). This slimyness is calculated by a simple average of the votes of the readers in the platform using a 1-5 scale, being "1. Not really Slimy" and "5. Simply beyond

[3] http://www.perverted-justice.com

slimy". Note that this evaluation is given by the website's visitors, so it may not be accurate, but we consider that it is a proper baseline in order to compare future conversations of the chatter-bot.

For reproducibility purposes the totality of the gathered conversations with their corresponding slimyness level is provided through a Lucene Index.[4]

3.2 Chat-bot conversation process

Every time the user introduces some text our system gathers the input data, from now on "Conversational Unit" (CU), and character and word repetitions are removed. Then, "emoticons" are replaced and misspelled words are corrected through Levenshtein distance [**?**]. To translate SMS-like words, we use *diccionarioSMS*,[5] a website supported by *MSN, Movistar, Vodafone and Lleida.net.* Negobot also recognises named entities relying on a database of personal nouns.[6]

Next, the CU is translated into English if the input language is different. This translation is performed with Google's Translation Service.[7] The decision of translating resides on the intention of normalising Negobot's knowledge base to English, to be able to scale the system to other languages.

In the next step the system queries Lucene,[8] a high-performance Information Retrieval tool, to stablish how similar is the CU to the assimilated conversations, returning a similarity rank. Based on that query, we retrieve the conversations surpassing an empirically-defined similarity threshold. The final result is then used to calculate the B value of the evaluation function (see Section 3.4).

Question-answering patterns Negobot uses the Artificial Intelligence Markup Language (AIML) to provide the bot with the capacity of giving consistent answers and, also, the ability to be an active part in the conversation and to start new topics or discussions about the subject's answers.

Although the AIML structure is based on the Galaia project,[9] which has successfully implanted derived projects in social networks and chat systems [**?**,**?**,**?**], we edited their AIML files to adequate them to our needs. Those files can be found at the authors' website.[10]

3.3 Applying game theory

Applied game theory intends to provide a solution for a large number of different problems through a deep analysis of the situation. Modelling these problems as a game implies that there exist two or more teams or players and that the result

[4] https://github.com/S3labDeusto/Datasets/blob/master/Negobot - Lucene index.7z

[5] www.diccionariosms.com

[6] www.planetamama.com.ar

[7] http://translate.google.com

[8] http://lucene.apache.org/

[9] http://papa.det.uvigo.es/ galaia/ES/

[10] http://paginaspersonales.deusto.es/patxigg/recursos/NegobotAIML.zip

of the actions of one player depends on the actions taken by the other teams. It is important to notice that game theory itself does not teach to play the game but shows general strategies for competitive situations. The Negobot system assumes that there are two players in a conversation: the alleged paedophile and the system itself.

Goal The main goal of the Negobot system is to gather, through the conversation, the maximum amount of information in order to evaluate it afterwards. This evaluation will determine whether the subject is a paedophile or not, to communicate, in the latter case, to pertinent authorities.

Conversation level The conversation level depends on the input data from the subject and determines which action should be taken by the system. There are seven levels consisting of seven different stages. In each stage, the subject talking with the bot is evaluated. The result of the evaluation will determine whether the subject that is talking is an alleged paedophile or not, depending on the content of the conversation and changing or not the level of the conversation.

- **Initial state (Start level or Level 0).** In this level, the conversation has started recently or it is within the fixed limits. The user can stay indefinitely in this level if the conversation does not contain disturbing content. The topics of conversation are trivial and the provided information about the bot is brief: only the name, age, gender and home-town. The bot does not provide more personal information until higher levels.
- **Possibly not (Level -1).** In this level, the subject talking to the bot, does not want to continue the conversation. Since this is the first negative level, the bot will try to reactivate the conversation. To this end, the bot will ask for help about family issues, bullying or other types of adolescent problems.
- **Probably not (Level -2).** In this level, the user is too tired about the conversation and his language and ways to leave it are less polite than before. The conversation is almost lost. The strategy in this stage is to act as a victim to which nobody pays any attention, looking for affection from somebody.
- **Is not a paedophile (Level -3).** In this level, the subject has stopped talking to the bot. The strategy in this stage is to look for affection in exchange for sex. We decided this strategy because a lot of paedophiles try to hide themselves to not get caught.
- **Possibly yes (Level +1).** In this level, the subject shows interest in the conversation and asks about personal topics. The topics of the bot are favourite films, music, personal style, clothing, drugs and alcohol consumption and family issues. The bot is not too explicit in this stage.
- **Probably yes (Level +2).** In this level, the subject continues interested in the conversation and the topics become more private. Sex situations and experiences appear in the conversation and the bot does not avoid talking about them. The information is more detailed and private than before because we have to make the subject believe that he/she owns a lot of personal

information for blackmailing. After reaching this level, it cannot decrease again.
- **Allegedly paedophile (Level +3).** In this level, the system determines that the user is an actual paedophile. The conversations about sex becomes more explicit. Now, the objective is to keep the conversation active to gather as much information as possible. The information in this level is mostly sexual. The strategy in this stage is to give all the private information of the child simulated by the bot. After reaching this level, it cannot decrease again.

Actions The Negobot system has three sensors, three actuators and three different actions to take to obtain its objective.

1. **Sensors:** (i) Knowledge of the current level, (ii) knowledge of the complete current conversation, (iii) assimilated conversations by the AI system.
2. **Actuators:** (i) The accusation level goes up, (ii) the accusation level goes down, (iii) the accusation level is maintained.
3. **Actions:** (i) Level goes up, this action increases the accusation level of the subject to modify the type of conversation by adding more personal information; (ii) level goes down, this action decreases the accusation level of the subject to change the type of conversation by adding more tentative information; (iii) maintain level, this action maintains the accusation level of the user to leave the current conversation type.

The consequences after each action are unknown beforehand because the system does not know how the other player will response to each answer, question, affirmation or negation.

Strategy The aim of the system is to reach a final state while extracting the highest amount of information. The state of the system is determined both by the conversation and the conversation topic. The environment is partially observable since the chat-bot does not know what will the subject answer or question. Besides, the environment is stochastic because the conversations are composed only of questions, answers, affirmations and negations and when the user is writing one sentence, the environment does not change.

Furthermore, the system utilises an evaluation function (refer to Section 3.4) to analyse the answers. This analysis starts by evaluating the conversation level. Then, it generates the answer and, finally, communicates with the subject. This task had to be performed within a coherent and variable time so that it resembled the writing way of a child. To this end, the system calculates a waiting time for the response based on the number of words in the answer and an estimated child's writing speed.

3.4 Evaluation function

The evaluation function determines the bot's behaviour according to the evaluation of the sentences that the subject introduces, and determines, in real time, its level of paedophilia. This function is also responsible for evaluating the evolution of the conversation (i.e., if the level maintains, goes up or down).

We defined the function as,

$$f(x) = \alpha + \beta + \gamma \tag{1}$$

α represents the historic values of the conversations with this user. This variable is defined as the sum of the levels of the previous conversations, divided by the number of total conversations. It is defined as, $\sum_{n=1}^{n=x-1} f(x)/n$, where n is the number of conversations the current user has maintained with Negobot.

β represents the "slimyness" of the current CU. We calculated the average value of the result of multiplying the similarity score of each of the retrieved paedophiles' conversations — the ones that surpass a fixed value of similarity — by its "slimyness" value, $\sum_{n=1}^{n=x} score * slimyness/|AC_l|$, where AC_l is the total number of assimilated conversations.

γ evaluates the temporality of the subject's conversations (i.e., how frequent the user talks with Negobot). There are two possible values for γ, when $\beta = 0$ and when $\beta > 0$:

1. Value for γ if $\beta = 0$
   - The subtraction between the "slimyness" of the CU and the total CUs of the subject. This result is divided by the subtraction, in hours, between the last maintained CU and the first maintained CU: $Ncu - Tcu/T_{last} - T_{first}$
2. Value for γ if $\beta > 0$
   - The number of CUs that surpass the "slimyness" threshold divided by the subtraction, in hours, between the last maintained CU and the first maintained CU: $Ncu > threshold/T_{last} - T_{first}$

4 Examples of conversations

This section presents some sample conversations with Negobot. It must be noted that the system was developed to analyse extensive conversations because the most dangerous paedophiles are too cautious. In our experiments we show some tests in a controlled environment and with a short extension (see Figure 2, Figure 3, Figure 4 and Figure 5),[11]. Also note that the evaluation function does not depend only on the CU, historical conversation values and time are also taken into consideration.

For these experiments we performed three conversations: one *normal* conversation, one *aggressive* conversation and one *passive* conversation.

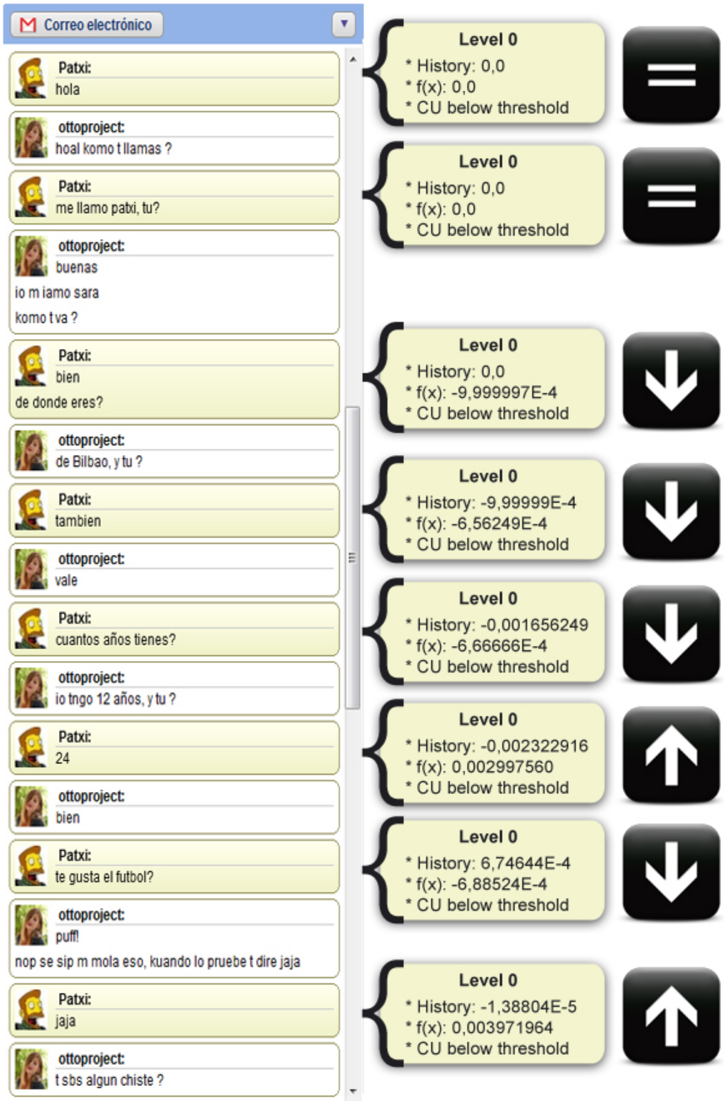
- **Normal conversation**. A subject maintained a conversation with Negobot talking about non suspicious topics such as sport preferences or jokes. The subject tries to know the child emulated by our bot but without showing

[11] In these examples *ottoproject* is the Negobot system and Patxi is a user that emulates the different behaviours

interest in personal information. The language used by the subject is polite and provides unimportant information such as: name, age, gender, hometown or sport preferences.

- **Aggressive conversation**. The subject who maintained this conversation was asked to add explicit questions about sex related topics with the objective to obtain sexual information. Despite usually paedophiles use a more cautious approach, we needed to provide a short conversation talking about sex to show how the level of suspicion raises. In this conversation the bot is asked about sensitive unknown information, such as virginity or porn.
- **Passive conversation**. In this conversation the subject starts a conversation and when the age of Negobot is known he/she tries to end it. Negobot then asks the reason why the subject is leaving and after 10 minutes of not having a response tries to restart the conversation.

The conversations are in Spanish since its the language the system was designed for, but variables have been translated and visual guidance is provided to understand the results.



Patxi:
hola

ottoproject:
hoal komo t llamas ?

Patxi:
me llamo patxi, tu?

ottoproject:
buenas
io m iamo sara
komo t va ?

Patxi:
bien
de donde eres?

ottoproject:
de Bilbao, y tu ?

Patxi:
tambien

ottoproject:
vale

Patxi:
cuantos años tienes?

ottoproject:
io tngo 12 años, y tu ?

Patxi:
24

ottoproject:
bien

Patxi:
te gusta el futbol?

ottoproject:
puff!
nop se sip m mola eso, kuando lo pruebe t dire jaja

Patxi:
jaja

ottoproject:
t sbs algun chiste ?

Level 0
* History: 0,0
* f(x): 0,0
* CU below threshold
=

Level 0
* History: 0,0
* f(x): 0,0
* CU below threshold
=

Level 0
* History: 0,0
* f(x): -9,999997E-4
* CU below threshold
↓

Level 0
* History: -9,99999E-4
* f(x): -6,56249E-4
* CU below threshold
↓

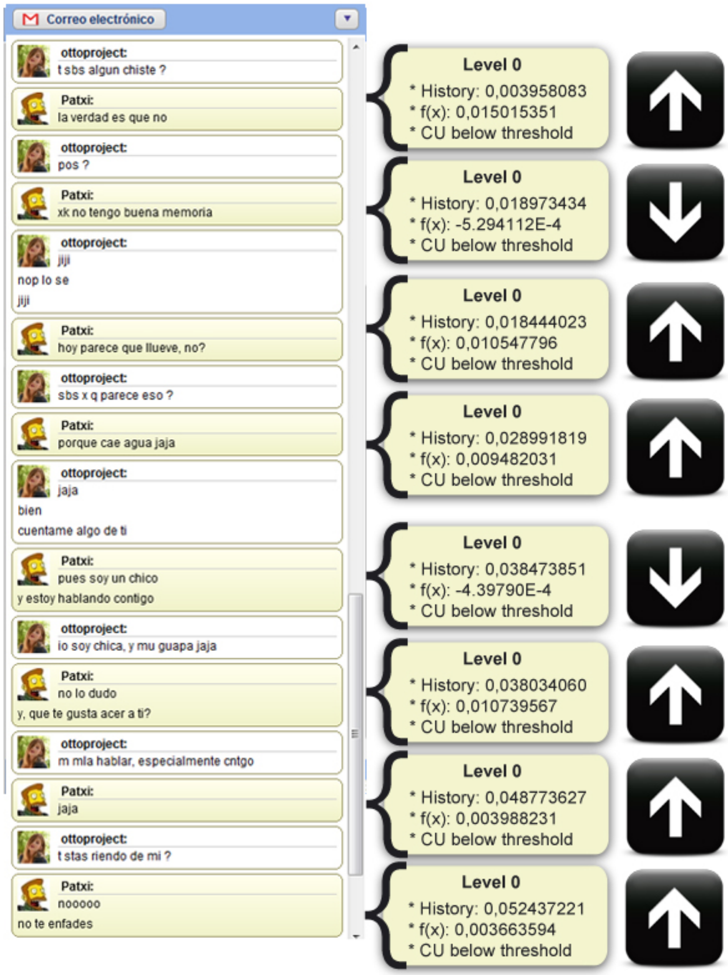
Level 0
* History: -0,001656249
* f(x): -6,66666E-4
* CU below threshold
↓

Level 0
* History: -0,002322916
* f(x): 0,002997560
* CU below threshold
↑

Level 0
* History: 6,74644E-4
* f(x): -6,88524E-4
* CU below threshold
↓

Level 0
* History: -1,38804E-5
* f(x): 0,003971964
* CU below threshold
↑

Fig. 2. Normal conversation part 1/2



Correo electrónico

ottoproject:
t sbs algun chiste ?

Patxi:
la verdad es que no

ottoproject:
pos ?

Patxi:
xk no tengo buena memoria

ottoproject:
jiji
nop lo se
jiji

Patxi:
hoy parece que llueve, no?

ottoproject:
sbs x q parece eso ?

Patxi:
porque cae agua jaja

ottoproject:
jaja
bien
cuentame algo de ti

Patxi:
pues soy un chico
y estoy hablando contigo

ottoproject:
io soy chica, y mu guapa jaja

Patxi:
no lo dudo
y, que te gusta acer a ti?

ottoproject:
m mla hablar, especialmente cntgo

Patxi:
jaja

ottoproject:
t stas riendo de mi ?

Patxi:
nooooo
no te enfades

Level 0
* History: 0,003958083
* f(x): 0,015015351
* CU below threshold

↑

Level 0
* History: 0,018973434
* f(x): -5.294112E-4
* CU below threshold

↓

Level 0
* History: 0,018444023
* f(x): 0,010547796
* CU below threshold

↑

Level 0
* History: 0,028991819
* f(x): 0,009482031
* CU below threshold

↑

Level 0
* History: 0,038473851
* f(x): -4.39790E-4
* CU below threshold

↓

Level 0
* History: 0,038034060
* f(x): 0,010739567
* CU below threshold

↑

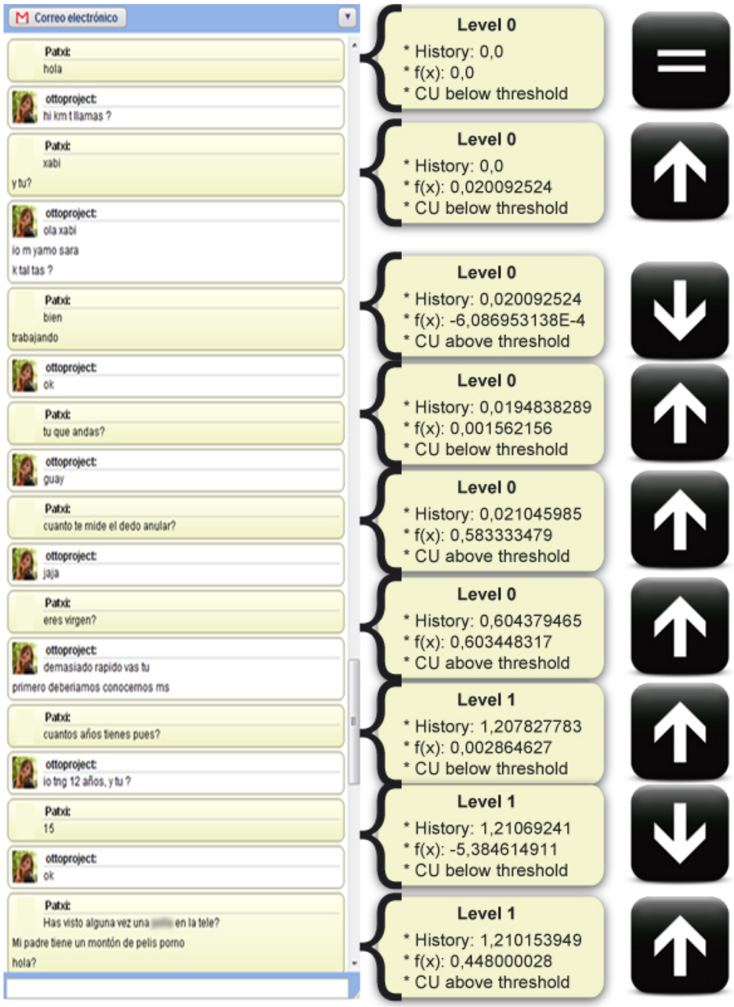
Level 0
* History: 0,048773627
* f(x): 0,003988231
* CU below threshold

↑

Level 0
* History: 0,052437221
* f(x): 0,003663594
* CU below threshold

↑

Fig. 3. Normal conversation part 2/2



Patxi:
hola

ottoproject:
hi km t llamas ?

Patxi:
xabi
y tu?

ottoproject:
ola xabi
io m yamo sara
k tal tas ?

Patxi:
bien
trabajando

ottoproject:
ok

Patxi:
tu que andas?

ottoproject:
guay

Patxi:
cuanto te mide el dedo anular?

ottoproject:
jaja

Patxi:
eres virgen?

ottoproject:
demasiado rapido vas tu
primero deberiamos conocernos ms

Patxi:
cuantos años tienes pues?

ottoproject:
io tng 12 años, y tu ?

Patxi:
15

ottoproject:
ok

Patxi:
Has visto alguna vez una [redacted] en la tele?
Mi padre tiene un montón de pelis porno
hola?

Level 0
* History: 0,0
* f(x): 0,0
* CU below threshold

=

Level 0
* History: 0,0
* f(x): 0,020092524
* CU below threshold

↑

Level 0
* History: 0,020092524
* f(x): -6,086953138E-4
* CU above threshold

↓

Level 0
* History: 0,0194838289
* f(x): 0,001562156
* CU below threshold

↑

Level 0
* History: 0,021045985
* f(x): 0,583333479
* CU above threshold

↑

Level 0
* History: 0,604379465
* f(x): 0,603448317
* CU above threshold

↑

Level 1
* History: 1,207827783
* f(x): 0,002864627
* CU below threshold

↑

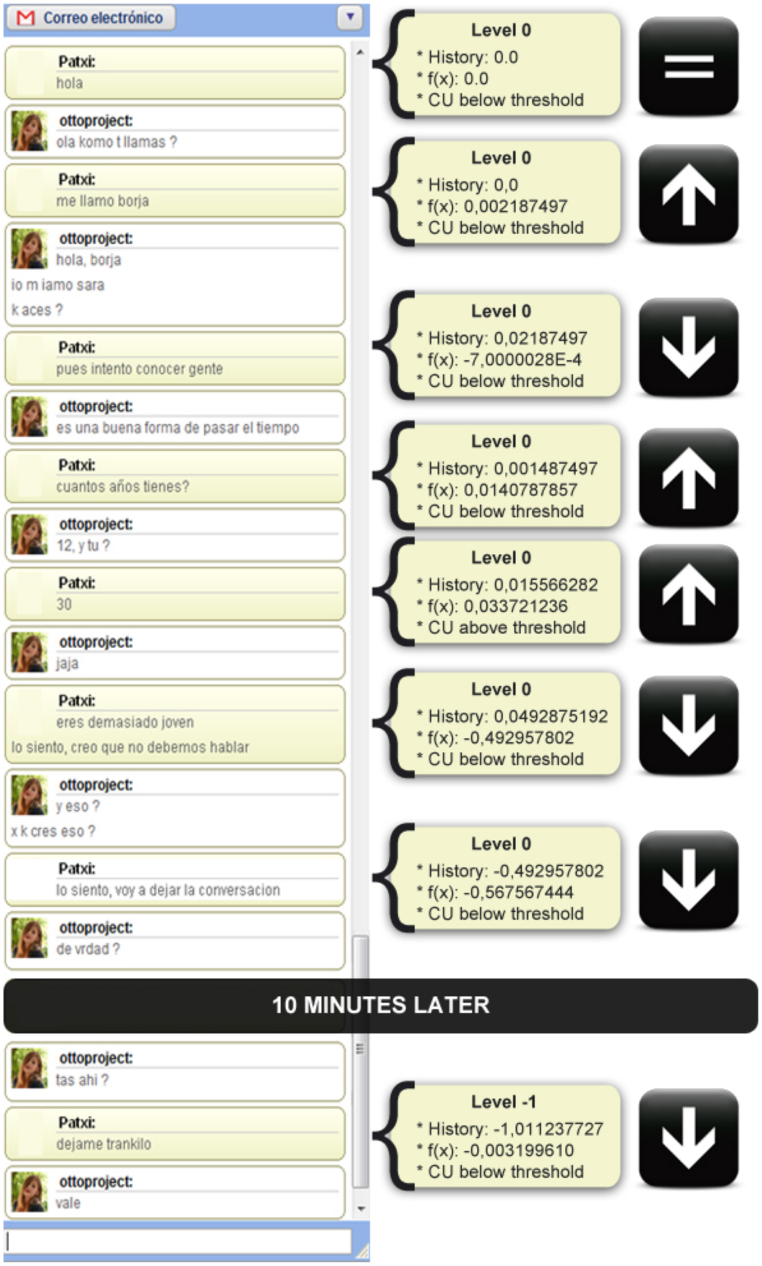
Level 1
* History: 1,21069241
* f(x): -5,384614911
* CU below threshold

↓

Level 1
* History: 1,210153949
* f(x): 0,448000028
* CU above threshold

↑

Fig. 4. Aggressive conversation



Patxi: hola

ottoproject: ola komo t llamas ?

Patxi: me llamo borja

ottoproject: hola, borja
io m iamo sara
k aces ?

Patxi: pues intento conocer gente

ottoproject: es una buena forma de pasar el tiempo

Patxi: cuantos años tienes?

ottoproject: 12, y tu ?

Patxi: 30

ottoproject: jaja

Patxi: eres demasiado joven
lo siento, creo que no debemos hablar

ottoproject: y eso ?
x k cres eso ?

Patxi: lo siento, voy a dejar la conversacion

ottoproject: de vrdad ?

10 MINUTES LATER

ottoproject: tas ahi ?

Patxi: dejame trankilo

ottoproject: vale

Level 0
* History: 0.0
* f(x): 0.0
* CU below threshold
=

Level 0
* History: 0,0
* f(x): 0,002187497
* CU below threshold
↑

Level 0
* History: 0,02187497
* f(x): -7,0000028E-4
* CU below threshold
↓

Level 0
* History: 0,001487497
* f(x): 0,0140787857
* CU below threshold
↑

Level 0
* History: 0,015566282
* f(x): 0,033721236
* CU above threshold
↑

Level 0
* History: 0,0492875192
* f(x): -0,492957802
* CU below threshold
↓

Level 0
* History: -0,492957802
* f(x): -0,567567444
* CU below threshold
↓

Level -1
* History: -1,011237727
* f(x): -0,003199610
* CU below threshold
↓

Fig. 5. Passive conversation

5 Discussion and future work

More and more children are connected nowadays to the Internet. Although this medium provides a lot of important advantages, it also provides the anonymity that paedophiles use to harass children. A rapid identification of this type of users on the Internet is crucial. Systems able to identify those menaces are going to be important protecting this less prepared population segment on the Internet.

Negobot is the first chat-bot system based on game theory to detect paedophile tendencies. It analyses conversations on real time and evaluates the subjects interacting with the conversational agent.

In order to better simulate a child behaviour the system uses variable response times depending on the sentence's size (with the estimated writing speed of a child), recognises fragmented sentences, or uses recursive patterns for similar questions (e.g., question: 'hello', 'hi', 'helloooo friend'; answer: 'hello'). Moreover, the system defines one personality for the bot, maintaining the coherence in every conversation (e.g., favourite colour, favourite music and born date). Finally, the system is also able to translate and understand misspelled words and SMS terms, thanks to a correction system based on the Levenshtein distance.

To evaluate each sentence, the system employs: (i) coincidences between the subject's words and the paedophile conversations assimilated in the system, (ii) the conversations gathered from the subject, and (iii) the subject's writing frequency. Using this information the system can: on the one hand, modify the chat-bot's answers according to the conversation in each moment and with each subject and, on the other hand, evaluate the subject's paedophilia level in real time.

However, the proposed system has several limitations. First, despite current translation systems are good, they are far to be perfect. Therefore, the language is one of the most important issues. To solve it, we should obtain already classified conversations in other languages, in this particular case Spanish conversations. Besides, the subsystem that adapts the way of speaking (i.e., child behaviour) should be improved. To this end, we will perform a further analysis of how young people speak on the Internet. Finally, there are some limitations regarding how the system determines the change of a topic. They are intrinsic to the language, and its solution is not simple.

The future work of the Negobot system is oriented in three main directions. First, we will generate a net of collaborative agents able to achieve a common goal in a collaborative way, in which agent will seek an optimal common equilibrium (Nash equilibrium). Second, we will add more NLP techniques like word sense disambiguation or opinion mining to improve the understanding of the bot. Also, we will try to upgrade the question-answering patterns system, for example, including a semantic and syntactic analysis of the conversations. Third, we will adapt the Negobot system for social networks, chat rooms, and similar environments. Finally, Negobot has been tested in a closed environment to measure its capabilities, so one important line of improvement would be directed precisely towards deploying this system in a real environment (analysing first the legislative/legal implications).

Lastly, we consider that the Negobot system and its future updates are going to be useful to the authorities to detect and identify paedophile behaviours. could be of great help to detect alleged pedophiles in order to open an official investigation. It seems clear that this system will not replace any specialised agent in the law enforcement nor the volunteers working with them, but, Negobot could help filtering those suspects to optimize their efforts and improve their already great work. In conclusion, the Negobot project and existing initiatives like PROTEGELES[12] can achieve a safer Internet, focusing on one of the most vulnerable segments of the population: children.

Acknowledgments

We would like to acknowledge the Department of Education, Universities and Research of the Basque Government for their support in this research project through the Plan +Euskadi 09 and more specifically for the financing of the project “Negobot: Un agente conversacional basado en teora de juegos para la deteccin de conductas pedfilas”. UNESCO Cod 120304. Area Cod Tecnológica 3304. Science Area Code330499.

[12] http://www.protegeles.com/