Text-to-Emotion Engine for Real Time Internet Communication

Xu Zhe and Anthony Boucouvalas
Multimedia Communications Research Group, School of Design Engineering and Computing,
Bournemouth University, Fern Barrow, Poole, Dorset, BH12 5BB, UK.
Email: {zxu, tboucouv}@bournemouth.ac.uk

Abstract- In this paper, the latest version of an emotion extraction engine used for real time Internet text communication is presented. The engine can analyse input text from a chat environment, extract the emotion being communicated, and deliver the parameters necessary to invoke an appropriate expressive image on screen to the communicating user display. The parameters include the emotion extracted from the text being typed and the intensity of the emotion. The measurement of duration of the expression displayed is also possible being an agreed function of the emotion intensity. Semantic analysis is used to extract emotional words. Analysing the individual word position, the person the emotion is referred to, and the time the emotion occurred, identification of emotional words, as well as using of a set of grammatical rules allow the engine to perform satisfactorily. Static and dynamic tests were carried out in order to test the engine performance.

1 Introduction

The ability to effectively communicate in real time over distance is important to people both at personal level and in-groups.
Various real time videoconferencing systems have been developed. However, transmission and display of moving images require large bandwidth and are expensive.
Communication of emotions through facial expression has received great attention since Darwin's [3] analysis. An emotion extraction engine, which can analyse text sentences typed by the users, is presented. The sentences are analysed and the detected emotive content and the appropriate expressions are displayed automatically. The intensity and duration of the expression are also calculated and displayed in real time automatically.
In order to test the functionality of the emotion extraction engine, two types of test were defined. The first is a static test, which uses artificially created emotional sentences to test the engine success. The second type is a dynamic test, in which users can input any sentences arbitrarily.
This paper is organised as follows. In section 2, the background knowledge is reviewed. In section 3, the emotion extraction engines are described in detail. In section 4, the test strategy for the engine is given. Section 5 illustrates possible applications using the engine. Finally in section 6, conclusions are given.

2 Background

In human language, sounds and letters make up the significant parts of the communication. Words are groups of letters and may be combined in many ways to make up a sentence [9].
From Chomsky [10] and Saussure [11], sentences can be broken down into two groups of words: function words and content words. Function words include nouns, adjectives, verbs and adverbs, while content words are prepositions, conjunctions and auxiliary verbs. All these groups contain sub classes, which can be represented by tree diagrams. In figure 1 an example tree presentation for nouns is shown.

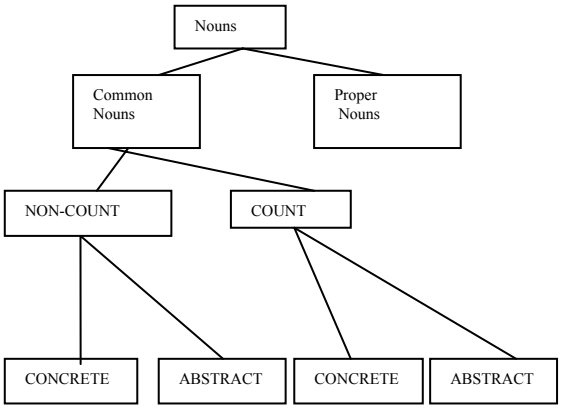

Figure 1 tree presentation for nouns

According to what groups the words belong to and the words preceding and subsequent, the sentences can be decomposed and be analysed.

- ***Sentence Structure and Analysis***

A sentence may only consist of a single word or a group of words. In order to analyse sentences the sentence definition groups have been developed. E.g., one of the sentence definition groups is: Article + noun + verb + preposition
General sentence constructions are identified in [4]. The constructions include:

Declarative statement
Interrogative statement
Finite verb phrases
Negative statement

Declarative statement: This kind of statement usually contains a subject and a verb where the subject precedes the verb. E.g., I am running.

Interrogative statement: A question, e.g., was he angry?

Finite verb phrases: Expresses statement or directives. E.g., Sit down.

Negative statements: This kind of statement expresses anything but what the sentence mean. E.g., I am not happy.

Rewrite rules or so-called tree presentation [10] are technologies to analyse sentence structures. In figure 2, a tree representation of a sentence is showed.

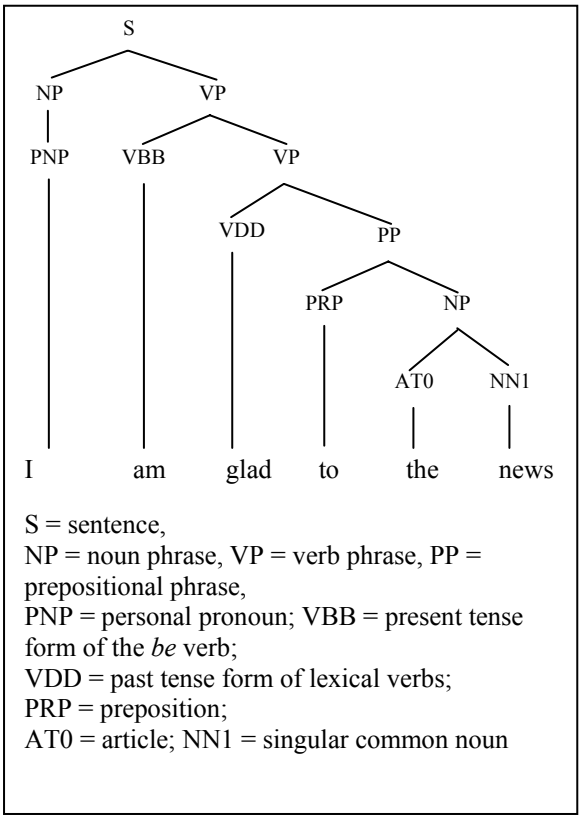

Figure 2 Tree representation of a sentence

- **Universal Expressions**

Research in facial expression has concluded that there are six universal categories of facial expressions that are recognised across cultures. [13] The categories are happy, sad, anger, fear, disgust and surprise. Within each of the categories, a wide range of expression intensity and variation of detail expression exists.

3 Emotion extraction engine

Three models made up the engine: Input analysis function, tagging system and the parser.

The working flow of the engine is showed in Figure 3.

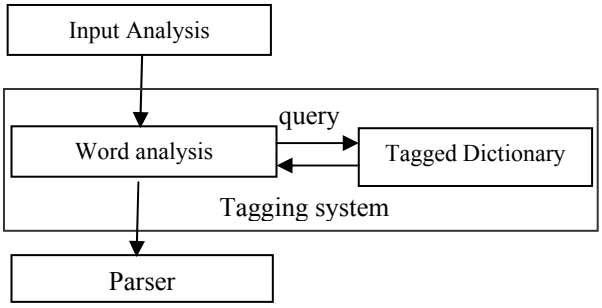

Figure 3 the engine working flow

- **Input analysis function**

The engine only analyses one sentence each time. The input sentences will be analysed individually without context information.

The Punctuation will be replaced with the pre-defined characters. E.g., sentence "I am happy." will be treated as "I am happy FLSTOP".

- **The Tagging system**

Although some tagging systems already exist, these tagging systems do not fulfil the engine's requirement. The engine requires the entire word to be properly tagged in order to keep a minimum response time. For every possible emotional word and its related intensity, the engine also requires particular marks. Typical traditional tagging systems only specify the suffixes or prefixes to differentiate between different word groups; unlike some tagging systems used in some famous corpus, e.g., BNC [7] and Brown corpus [12], a special tagset, which includes 119 categories, was developed for the engine.

In our tagset, A mark "EMO_W" is assigned to each possible emotion word. From section 2, it is can be known that there are six expression categories. Accordingly, the tagset use numbers from 1 to 6 to represent happy, sad, fear, surprise, anger, and disgust. Each emotion word tag also provides the *intensity level information. E.g., tag 1N2 stands for expression category "Happy" and intensity level 2.*

A subset of the tagset is showed in figure 4.

Tag	Representation	Example
NOUN	NOUN	Kindness
1N2 EMO_W	Happy, Noun, intensity 2	Happy
2N2 EMO_W	Sad, Noun, intensity 2	sad
3N2 EMO_W	Fear, Noun, intensity 2	fear
4N2 EMO_W	Surprise, Noun, intensity 2	surprise
5N2 EMO_W	Anger, Noun, intensity 2	angry
6N2 EMO_W	Disgust, Noun, intensity 2	disgust
6N3 EMO_W	Disgust, Noun,	Aversion

	intensity 3,	
PCF	Present continuous	IS
PP	Present prefect	have
past	Past	was
VI	Verb	Abandon
NT	Noun third person	she

Figure 4 the tag subset

Even for each word group a unique tag is specified, there are several words that appear in several groups and therefore have to be tagged to accommodate for all those possibilities. To deal with this, a special ambiguity tag, which specifies all the possible groups, has been developed.

Tagged dictionary:

Daily communications involve about two thousand words [5]. In order to identify the words, a special designed dictionary should be set up. In this project, a database containing 16400 words was used. The database includes three fields: word field, word category field and emotional tag field. Word field contains all the words and word category contains the corresponding tag. To extract emotional word as quickly as possible, the emotional tag field was added. An example is shown in figure 5.

Word	Word category	Emotional tag
Happy	1N2	EMO_W
Was	Past	

Figure 5 dictionary examples

Word analysis:

When receiving the input sentence, the tagging system will split the sentence into words and check through the tagged dictionary to find each word and the corresponding tag category. If a word is not found in the dictionary, the engine will undergo a suffix and prefix analysis. By examining the suffix and prefix of the word, the assumed tag may be derived. The output of the tagging system includes the words and corresponding categories.

An example working flow of the tagging system is shown in figure 6.

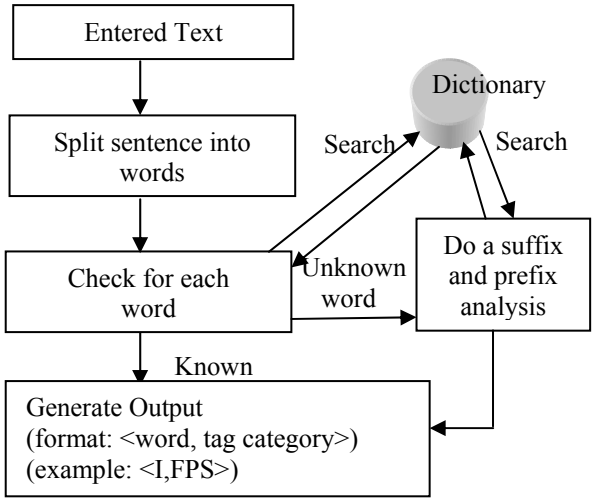

Figure 6 tagging working flow

- **Parser**

The parser's analysis is accomplished through the use of rewrite rules and tree representations [10].

The parser only takes care of the current emotional words, the person to whom the emotional words referred and the intensity of the emotional word.

The parser analyses the *intensity* of the expression by examining the tag of the emotional word and the adjectives.

A set of rules has been set up to specify the parser's work. The following checks are carried out:

Occurrence of emotional words: Unless there is an emotional word in the sentences, all the sentences will be disregarded. If the emotional word refer to the person himself / herself and it is referring to present continuous or present perfect continuous tense then the parser might generate the output.

The parser also analyses conditional emotional sentence i.e., ***'I'm happy when he's there!'*** and interprets this as a non-emotional sentence.

If there is one more emotional word in a sentence and they are connected by a conjunction then the parser will combine these two emotional states. E.g., ***'I am surprised and happy about it'***.

If there is no subject (emotional word / adjective is first word), then it will refer to the person who is in communication i.e., *'**apprehensively** happy' w*ill be treated as *'I am **apprehensively** happy'*.

Emotion intensity: If no adjectives (expressive adjective) in front of the emotional word then the intensity will only depend on the word's tag category.

If there is more than one adjective then the parser will increase the intensity automatically. The parser interprets 'very very sorry' as a high intensity emotional word.

Auxiliary verb: Since the engine only searches for the person's current feeling, the auxiliary verb has to be in the present continuo tense.

Sentences beginning with auxiliary verbs: If beginning contains an auxiliary verb, the sentences are questions and can not provide emotional information for the engine.

Negations: if the emotion is in negative form, then the sentences do not express current feeling and will be disregarded by the engine.

Conditional verbs: If conditional verbs are found, the sentences do not express current feeling and will be disregarded by the engine.

Emotional nouns: some words will grammatically fall into the noun category and are used as such.

Reportive keywords: Reportive form of sentences should be present in a grammatically correct expression.

The parser's output consists of the parameters for generating corresponding expressions, which include the emotional category, the expression intensity and the tense. The output format is: [emotional category] [intensity] [tense]. E.g., output [1][2][present] stands for emotional category happy, intensity level 2 and present tense.

The parser's working procedure is shown in figure 7.

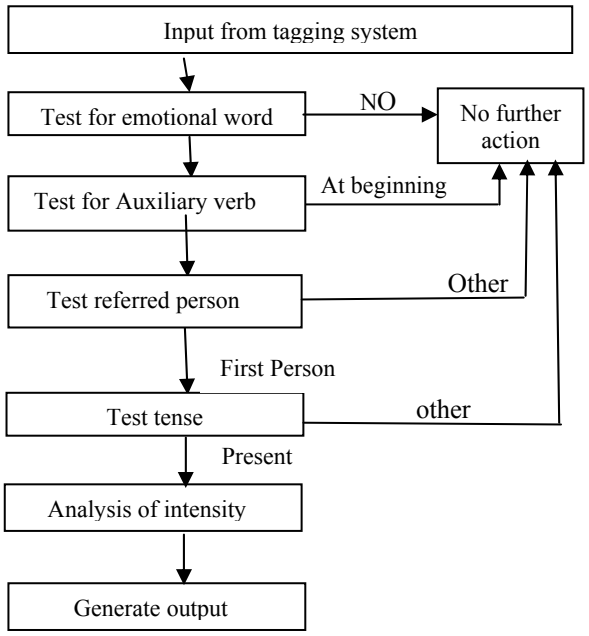

Figure 7 parser's working procedure

- **Generate expression**

The output from the parser will be sent out through the network to related users. After getting the output, corresponding expression images will be generated. The procedure for expression generation is shown in figure 8.

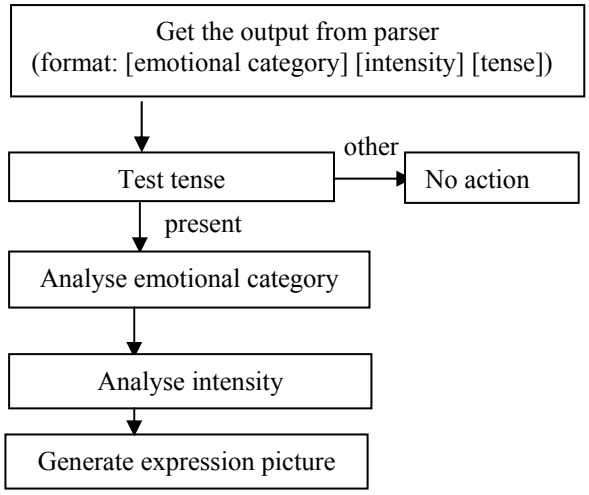

Figure 8 procedure for expression generation

Instead of transmitting pictures through the network, The engine only transmits the parameters needed for generating the corresponding photos. So the bandwidth requirement is extremely low.

4 Test strategy

To test the correctness and effectiveness of the emotion extraction engine, Two types of test are used. The first is the static test, in which emotional sentences have been constructed artificially. The second type is the dynamic tests, in which users can input any sentences freely.

4.1 Static test

The static test uses a number of paragraphs, which are read into the system, and all sentences interpreted as emotional are highlighted. Subsequently the paragraphs are entered into the emotion extraction engine to examine how well its automated estimation correlated to the manual extraction.

17 chapters from nine novels including "A Romance of Exmoor", "All for Love" were used to test the emotion extraction engine. Although the engine would only be looking for emotions expressed in first person and present tense, in order to establish a more complete picture of the systems overall performance, the test was made to account for second and third person as well as future and past tense. The results of the tests are showed in figure 9.

The advantage of this kind of test is that the sentence structure is a lot more complex than what can be expected to be used in oral or chatting context.

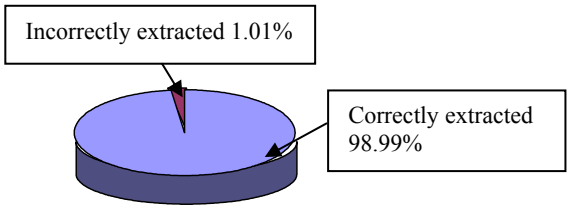

Figure 9 static test results

4.2 Dynamic tests

4.2.1 Feedback test

To avoid bringing together a number of people to take the test at a certain place and time, a questionnaire paper were developed to be completed by a population of students. This also avoids asking people to manually enter sentences themselves with emotional words into a PC.

All users were presented with the same scenarios and the same questions were asked. Users were asked to write down their emotional responses to a set of prescribed situations. The responses were subsequently entered manually into the system.

The questionnaire was presented to 50 people and 450 sentences in total were returned. These replied sentences were then inputted into the engine and results are shown in figure 10.

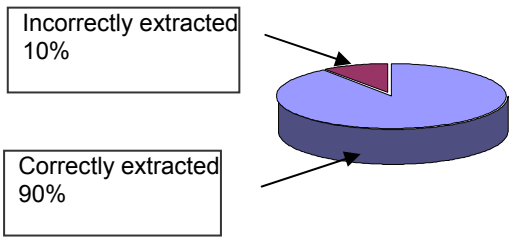

Figure 10 dynamic test results

4.2.2 Prototype application test

With the help of the prototype application, users can test the engine in a chatting context. From the log file collected on the server, the engine can interpret most emotional sentences. Since the chatting environment is a fast paced kind of communication, spelling mistakes gives a lower hit rate.

4.3 Test reflections

The results of the tests show that the engine produced better results in formal writing. The spelling mistakes and slang have big negative influence to the engine. However, the engine has not been designed to deal with this type of errors.

5 Applications

The emotion extraction engine can be used in numerous applications. For example real time communication systems, real time games. The prototype software, 'Expressive Real Time Communications Interfaces', is an example application abased on the emotion extraction engine.
'Expressive Real Time Communication Interfaces' is the visual interface for effective real time collaboration over the Internet or among a group of people. It is capable of invoking expressions conveying feelings without making use of video. The application utilises discrete images of the participants in order to keep the bandwidth requirements to a minimum yet still accommodate for an as an elaborate communication tool as possible. The interfaces allow the viewing at a glance of participants in the system, and those pairs engaged in conversation, as well as the expressive image of the user engaged in the conversation. Figure 11 shows a typical screen of the application and in figure 12, the client and server architecture for the application is showed.

Figure 11 screen view of Expressive Interfaces

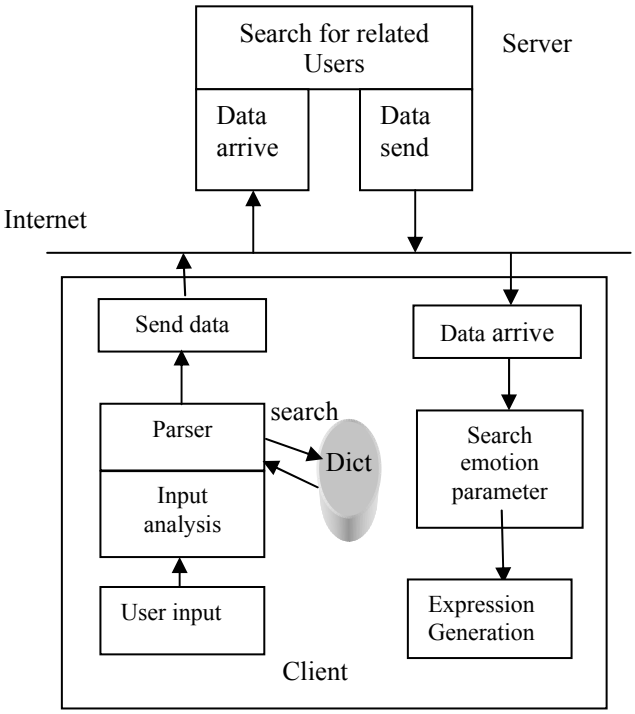

Figure 12 Application architecture

6 Conclusion

An emotion extraction engine, which based on word tagging and analysis of sentences, has been developed successfully.
Since image transmissions are avoided, the bandwidth requirements for real time communications can be extremely low with the emotion extraction engine.
To test the engine performance, some tests have been carried out. Also, prototype software 'Expressive Real Time Communications Interfaces' has been developed.

References

[1]. H Chowdhury and A Boucouvalas, 'expressive Real Time Communications', European Workshop on Distributed Imagine, Nov. 1999.
[2]. E.Reid,'Electropolis: Communication and community on Internet relay chat', thesis department of History, Melbourne Unv, 1991.
[3]. John N.Bassili:'Emotion Recognition', journal of personality and social psychology, 1979.
[4]. David Crystal:' Rediscover Grammar', Longman, 1996
[5]. Scott A. Kuehn, 'Communication innovation on BBS', computing and technology, April 1993
[6]. Stuart Russell, Peter Novig: 'Artificial Intelligence: A Modern Approach' ISBN 0-13-103805-21995
[7]. http://info.ox.ac.uk/bnc/what/gramtag.html
[8]. Hwang, Shin Ja J:' Text analysis in Translation', Christiane Nord 1994.
[9]. Graddol, Cheshire and Swann:' Describing Language'.
[10]. Ian Robinson: 'The New Grammarians' Funeral', Cambridge University Press
[11]. Terrence Gordon: '*Saussure for Beginners*', 1996
[12]. http://www.hit.uib.no/icame/brown/bcm.html
[13]. Frederic I.Parke, Keith Waters "Computer Facial Animation"