Using Gunnnig-Fog Index to Assess Instant Messages Readability from ECAs

[1]Ong Sing Goh, [2]Chun Che Fung, [3]Arnold Depickere [4]Kok Wai Wong,
School of Information Technology, Murdoch University, Perth, Western Australia
{[1]os.goh, [2]l.fung, [3]a.depickere, [4]k.wong, } @murdoch.edu.au

Abstract

In this paper, we examine the readability of the human-machine conversation transcripts in the instant messaging environment based on the Gunning-fog Index. The study is based on an embodied conversational agent (ECA) called Artificial Intelligent Natural-language Identity (AINI) which was designed to mimic human conversation. The ECA is also expected to supply answers with a sense of humour. We report on the collection and analysis of a corpus containing over 3,280 utterances in a series of real instant messages exchanged between AINI with 65 online "buddies".

1. Introduction

Instant messaging (IM) has gained popularity as an essential means of communication in the past few years as a supplement to e-mail. IM is particularly attractive to the younger generations as they are able to engage in instantaneous conversation with a list of their online "buddies". Popular chat systems or instant messagers (IMs), such as America Online's Instant Messenger, Microsoft's MSN Messenger, ICQ, and Internet Relay Chat (IRC) have changed the way computer users communicate with their friends, acquaintances, and colleagues. Once limited to only desktops, now, more and more instant messaging systems are finding their way to handheld devices and cell phones. This has enabled users to chat from virtually anywhere. Nowadays, IM can be found on almost all PC connected to the Internet. This technology has introduced an easier communication channel and increasing to become a dominant form of communication for many people. Research by Pew Internet & American Life [1] reveals that 53 million adults trade instant messages and of those, 24% of them exchange IMs more frequently than email.

This popularity has created many interest among the IM proprietary, including Microsoft, to integrate conversation robots (or 'bots') in their MSN Messenger system. Microsoft challenged developers worldwide to create conversational bots for MSN® Messenger and Windows Live™ in the "Invasion of the Robots Contest"[1]. As this technology is getting popular, it important to assess the readability of the IM text. The use of the Gunning-Fog index is the focus in this paper.

2. Related Works

There are a number of reports that have been published on the use of IM as a new media of communication between human users. However, to the knowledge of the authors, none of them have investigated the human-machine conversation transcripts. In particular, no report was on the assessment of the readability of the IM text. There are some recent reports on the monitoring of Internet chat, including IM, by U.S. officials. The objectives of the investigation of such exchanges are mainly due to suspect of planned terrorist attacks [2]. There are also cases where there are concerns for the security of younger users who could become victims of criminals [3, 4]. On the issue of social impact, some papers strongly criticize this new form of communication [5] while others suggest IM is here to stay, and that digital communications technologies evolve and improve constantly and quickly [6],[7]. There are also papers referring to research on the design and usability for the public in general [5],[8]. In addition, IM usages in workplace and corporate contexts have recently soared [9],[10]. With regard to the linguistic aspects of IM usage, research has been undertaken in Spain [11], United Kingdom[5], United State [12], Sweden [13] and Portugal [3]. However, none of these researches have attempted to assess the readability of messages produced by human or bots in IMs.

In this paper, we examined the readability or text complexity of the utterances chatted by AINI conversational robot with 65 online buddies against transcripts produced by human-human dialogue such as TRAINS[2], IRC[3] and Blog[4] and human-machine conversation transcripts from award winner Loebner Prize[5].

[1] https://www.robotinvaders.com
[2] http://www.cs.rochester.edu/research/speech/93dialogs/
[3] http://swhack.com/logs/
[4] http://googleblog.blogspot.com/
[5] http://www.loebner.net/Prizef/loebner-prize.html

Third International Conference on Natural Computation (ICNC 2007)
0-7695-2875-9/07 $25.00 © 2007 IEEE

3. AINI'S Conversation Agent Architecture

This research project involves the establishment of an AINI conversational bot system with the MSN Messenger communication framework. The objective is to use AINI's conversational bots as an online character to simulate a human-machine conversation in IM. The real-time prototype relies on a distributed agent architecture designed specifically for Desktop, Web, Mobile devices and Personal Digital Assistant (PDA)[14]. All software agents, such as the conversation engine, knowledge model and natural language query, communicate with one another via TCP/IP. This is a combination of natural language processing and multimodal communication. A human user can communicate with the developed system using typed natural language conversation.

An AINI conversation bot can be seen as a 'digital character', capable of occupying and controlling a physical entity such as robot, or an embodied container as the one used in the conversational agent[15]. AINI is a conversation bot designed by the authors that is capable of having a meaningful conversation with human users who interact with AINI.

For the purposes of this research, the application area chosen for designing the conversation bot is primarily focused on the ability to communicate based upon scripts and/or artificial intelligence programming. AINI adopts a hybrid architecture that combines the utility of multidomain knowledge bases, multimodal interfaces and multilevel natural language query software. Given a question, AINI first performs a question analysis by extracting pertinent information to be used in query formulation. The tools used in this phase included Noun Phrases (NPs) and Verb Phrases (VPs) by deploying MINIPAR parser [16] as part of the newly built full parsing Natural Language Understanding and Reasoning (NLUR) system [17]. MINIPAR is a broad-coverage parser for the English language. An evaluation with the SUSANNE corpus shows that MINIPAR achieves about 88% precision, 80% recall with respect to dependency relationships. In our experiment, we have used corpus extracted by the Automated Knowledge Extraction Agent (AKEA)[18] and the MINIPAR parser, It is capable to parses nearly 500 words per second on a Dell Precision 380 Server 2GH with 1GB memory.

AINI employs an Internet three-tier, thin-client architecture that may be configured to work with any web application. It comprises of a data server layer, application layer and client layer. AINI architecture is can be found in [18].

4. Instant Messaging by AINI

The conceptual basis for the development of conversational bots are based on DesktopChat, WebChat and MobileChat on the client layer. This architecture enables AINI to interact with online users through a MSN Messenger protocol as shown in Figure 1. In addition, handheld devices are also becoming an important platform to deliver art and entertainment contents. This is mainly due to the tremendous growth of the number of mobile phone users world-wide. There are also ongoing improvements of technologies for content displaying content, interactivity, conversation, wireless and collaboration among networked users.

To provide a background of sch development, MSN Messenger for Desktop, or DesktopChat, was a free instant messaging client developed and distributed by Microsoft Windows since 1999. MSN Messenger was renamed to Windows Live Messenger in 2006. The WebChat sessions allow the users to interact in real time with the AINI software robot at the website via a browser through MSN Web Messenger. It is possible for virtually any computer with an Internet connection and browser to connect to the Messenger Service by using MSN Web Messenger. On the other hand, MobileChat uses a mobile chatting module, and is implemented in a series of logical phases which includes mobile-to-internet → internet-to-bots → bots-to-mobile chats. Mobile chat is an alternative in which users can chat with AINI using GPRS, WI-FI and 3G services.

5. Experiment Setup

This study is based on a corpus of instant messages produced by AINI with the online buddies using MSN Messenger. This corpus was collected during an Invasion of the Robots Contest.

5.1 Participants

The experiment's portal is open to the public from all over the world who can access this portal and freely participate in the study. The participant gets to know AINI from the advertisement of 8 famous BBS (bulletin board systems), which include blog websites and the AINI portal. The usage data gathered automatically were logins, logouts, joining, as well as chat messages. The portal allows online users to add AINI's contact to their "buddy-list", by allowing them to easily send and receive short textual messages. When a participant opens a message window to a buddy for the first time (and that buddy was online), an alert is sent to the buddy notifying them of their participation in the study.

5.2 . Chatlog System

A Chatlog System used MySQL was developed which stores user messages onto secondary storage. It provides real-time archive that captures chat messages so that they can be searched and indexed. This allows topic-based retrieval of chat sessions. These chat messages are essentially plain text messages that are quite small in comparison with images, video, or even documents.

5.3 The Corpus

Previous research has shown significant differences in IM communication resulting from the frequency of communication [19, 20]. In this study, we use word frequency for our analysis of the corpus collected from two difference sources. For the human-human transcripts, we extracted the conversation text from TRAINS, IRC and Blog transcripts. For the human-machine conversation text, the transcripts extracted from award winner Loebner Prize transcripts 2001 and 2004 (ALICE); 2005 and 2006 (Jabberwacky). These transcripts were separated based on the human transcripts (LPJudges) and the computer program or bots (LBBots) dialogue. Loebner Prize was discussed as a method to evaluate chatbots in terms of fooling people that they are chatting with a real human.

Another corpus is collected from the transcripts between AINI and 65 online buddies. We processed 29,447 words of running text and there are 2,541 unique words, 129,760 characters, and 4,251 sentence counts were recorded. From these data, we collected a total of approximately 63 hours of recorded conversation data, over 3,280 outgoing and incoming instant messages exchanges between AINI and the 65 buddies. Out of these, only 3 of them used MSN Mobile. The average sentence length of an IM transmission was 6.90 words, with approximately 13% of all transmissions being a single word in length. Table 1 provides a summary of data collected.

Table 1. Frequency of Word from Conversation Logs

	AINI	Human	Total
Word	18,358	11,089	29,447
Unique Word	1,368	1,173	2541
Character count	79,884	49,876	129,760
Sentence count	2,840	1,411	4,251
Utterance	1,721	1,559	3,280
Average sentence	6.46	7.85	6.90

6. Readability with Gunning-Fog Index

Readability formulas are getting popular nowadays. There are readability formulas for Spanish, French, German, Dutch, Swedish, Russian, Hebrew, Hindi, Chinese, Vietnamese, and Korean [21]. In Analytics of Literature [22], Sherman's proposed that literature is a subject for statistical analysis. He showed the importance of average sentence length and the relationship between spoken and written English. In linguistics, text complexity are related to the readability test and also called as the Gunning-Fog Index [23]. This test designed to measure the readability of a sample of English text. The resulting number is an indication of the number of years of formal education that a person requires in order to easily understand the text on the first reading. That is, if a passage has a fog index of 12, it has the reading level of a United State senior high school. The Gunning-Fog index can be calculated with the following algorithm:

$$0.4*\left(\left(\frac{words}{sentence}\right)+100\left(\frac{complexwords}{words}\right)\right) \quad (1)$$

G = 0.4(S + W), where S is the average sentence length and W is the percentage of words with three or more syllables.

The readability formulae have been around for a long time. Extensive research [24], [25] has shown that the formulae predict the difficulty of a prose passage quite well. Using this scale, most popular novels have Fog Index of 8 to 10, and academic papers are somewhere between 15 to 20.

7. Results and Discussion

Most of the readability studies are looking into the written text and none of them were used to examine conversation text. As IMs become more usable as a communication media, it is important to assess the readability for conversation text. In this study we examine the ease of understanding or comprehension based on style of the transcript generated from human-human dialogue and human-machine dialogue. The aim is to apply this understanding to issues such as text rating and texts complexity to human users.

After using word frequency techniques for analysing the corpus, results are displayed in Figure 2. This gives an easy illustration for comparing the IM and AINI utterances in the corpus. From the analysis of the total versus unique words, average sentence length and average word length, the results shows that it progresses in a linear form. Figure 2 shows that human-human conversation obtained a higher Gunning-Fog index as compared to the human-machine conversation. The official Google Blog (9.02) website gets the higher scored follows by IRC (6.22) and task-oriented dialogues TRAINS (4.66). Blogging is a form of online communications and could be very well related to journalism. The second higher score was Swhack Internet Relay Chat (IRC). Swhack was a cultural forum which continues to manifest itself in the form of a

publicly logged IRC channel. Its participants are predominently computer scientists, system administrators, writers, journalists, and engineers. The topics discussed encompass cryptography, linguistics, internationalization, law, Web technologies etc. TRAINS project based on the human-human Natural Spoken Dialogue based on task-oriented spoken dialogues. TRAINS corpus [26] consists of six and a half hours of speech, about 5900 speaker turns and 55000 transcribed words.

Table 2. Unique word and Lexical Density

	AINI	Human	LBJudges	LBBots
Unique Word	1,368	1,173	873	996
Lexical Density	13.2%	12.3%	35.5%	34.5%
Average Sentence Length (word)	6.46	7.85	5.58	5.01

On the other hand, human-machine conversation using LPJudges and LPBots get a lower score as compared to AINI and IM human users. The Gunning-Fog Index is shown in the Table 2. The AINI (1,368) and IM human (1,173) transcripts have more unique words compared to LPJudges (873) and LPBots (996). In term of the lexical density or different words, Loebner Prize judges (35.5%) and LBBots (34.5%) have higher lexical density. In terms of the average sentence length used, AINI (6.46) and LBBots (5.01) computer programs used shorten sentence in comparison to IM (7.85) and LBJudges (5.58) human users. Some of these aspects serve as indicators of syntactic difficulty. For instance, the longer a sentence is, the heavier the memory and mental load it places on the reader (Bormuth, 1966, cited by DuBay [24]). Thus, a longer sentence tends to be the more difficult than a shorter one. While not obvious, factors such as word frequency and word length are indicative of semantic difficulty. According to Zipf's Law[27], it is easier to understand words that are used frequently in a language. Furthermore, the most frequently used words tend to become shorter. There are a few reasons that why this phenomenon occurs. First, Loebner Prize was discussed as a method to evaluate conversation bots in terms of fooling people that are chatting with real human judges. 10 minutes given for simulation chatting with control situation may not be sufficient to judge naturalness. Secondly, the chats are based on unrestricted Turing Test [28]. Hence, the topics of the discussion coverage are wider and non-specific domain was demonstrated in the conversations. This also illustrates that the popularity of the Loebner prize bots is based on AIML language [29]. It is also observed that a general lack of progress in text understanding and natural language dialogue systems. Therefore, till now, an annual prize of bronze medal is awarded to the most human-like computer which is capable to demonstrate

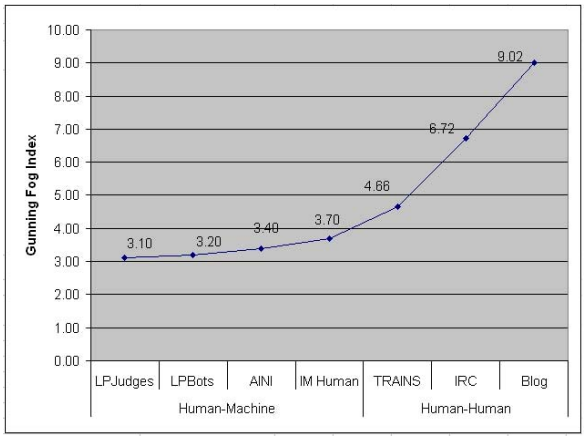

Fig. 1. Gunning-Fog Index

the conversational behavior than real human conversation. However for our conversation bots, AINI used Natural Language Understanding and Reasoning (NLUR) [17] for question-answering generation. AINI also equipped with multidomain knowledge bases which includes domain-specific and open-domain knowledge base to mimic human communication. The AINI transcripts chatted with real-time human conversation in IM showed remarkably consistent in their average sentence lengths (6.46 words).

The chart in Figure 1 shows that, the higher Gunning-Fox Index, the more human involvement in the conversation instead of preset dialogues. The observation being made from the experiment is that the more task-oriented or restriction on the domain topic, the higher Gunning-Fog Index will become. The more unrestricted is the domain, the more lexical density is observed.

8. Conclusion

Based on this experiment, IM conversation between human and machine shows an interesting pattern of behaviour displayed by the natural conversation bots. In this study, we simulated the proxy conversation log that contains clients' requests. It has to be appreciated that new simulations from other traces may have different results referred to in this paper.

Our study suggests that IM human-machine conversations display considerable variation on the text complexity and readability between the machine and IM human users. Evidence also suggests that AINI's "buddies" are interested in chatting with bots just to seek information, to be friends, to express their emotions, or just chat for leisure. To a certain extent, AINI is successful in imitating human conversation. Although the standard and content of the conversation may not be claimed to be exactly of "high quality", the bot's responses are "human" enough to its IM "buddies".

The main contribution in this study is the observation that conversational robots may attempt to

reinforce the impression that there is a real dialogue by using shorter sentences and concrete terms to increase readability. It was also discovered that the readability measures of literacy correlate closely with the measures of intelligence and ability of conversation between human-human or human-machine. In addition, these measures also associate with the breadth of the domain knowledge. In our finding, advanced communicator (human or machines) which has vast amount of knowledge will perform well across a diversity of domain knowledge. Over time, it is expected that bots-language will become more efficient and closer to human spoken languages.

9. References

[1] Eulynn Shiu and Amanda Lenhart, (2004),How Americans Use Instant Messaging, [Online]. Available: http://www.pewinternet.org/pdfs/PIP_Instantmessage_Report.pdf

[2] USAToday, (June 24, 2002),Agents pursue terrorists online, [Online]. Available: http://www.usatoday.com/enws/world/2002/06/21/terrorweb-usat.htm

[3] Silvina Ruth Crenzel and Vera Lúcia Nojima, (2006),Children and instant messaging, [Online]. Available: www.iea.cc/ergonomics4children/pdfs/art0233.pdf

[4] Hoffman Kathryn R., "Messaging Mania in Time for Kids," *Time*, vol. 8, 2003.

[5] S. Livingstone, (2006),UK Children Go Online: Surveying the experiences of young people and their parents, [Online]. Available: www.lse.ac.uk/collections/children-go-online/UKCGO_Final_report.pdf

[6] Rebecca E Grinter and Leysia Palen, "Instant messaging in teen life," presented at 2002 ACM conference on computer supported cooperative work, New Orleans, 2002.

[7] Naomi Baron, "Instant messaging and the future of language," *Communications of the ACM*, vol. 48, pp. 29-31, 2005.

[8] A.F Rovers and H.A Van Essen, "Him: A framework for haptic instant messaging," presented at CHI, Viena, April 2004.

[9] James D. Herbsleb, David L. Atkins, David G. Boyer, Mark Handel, and Thomas A. Finholt, "Introducing Instant Messaging and Chat in the Workplace," presented at CHI'2002, Minneapolis, Minnesota, USA., 2002.

[10] Jacki O'Neill and David Martin, "Text Chat In Action," presented at GROUP'03, Sanibel Island, Florida, USA., 2003.

[11] Ruben C. Forgas and Jaume S. Negre, "The use of new Technologies amongst minors in the Balearic Islands.," presented at IAARE Conference, Melbourne, 2004.

[12] David Craig, (2003),Instant messaging: the language of youth literacy. The Boothe Prize Essays, [Online]. Available: www.stanford.edu/group/pwr/publications/Boothe_0203/PWR%20Boothe-Craig.pdf

[13] Ylva Hård af Segerstad and Sylvana Sofkova Hashemi, "Exploring the Writing of Children and Adolescents in the Information Society EARLI SIG Writing," presented at 9th International Conference of the EARLI - Special Interest Group on Writing, Geneve, Switzerland, 2004.

[14] Ong Sing Goh, C. C Fung, Cemal Ardil, K.W Wong, and A Depickere, "A Crisis Communication Network Based on Embodied Conversational Agents System with Mobile Services," *Journal of Information Technology*, vol. 3, pp. 257-266, 2006.

[15] Ong Sing Goh, Cemal Ardil, Wilson Wong, and C. C Fung, "A Black-box Approach for Response Quality Evaluation Conversational Agent System," *International Journal of Computational Intelligence*, vol. 3, pp. 195-203, 2006.

[16] Dekang Lin, "Dependency-based Evaluation of MINIPAR," presented at Workshop on the Evaluation of Parsing Systems, Granada, Spain, 1998.

[17] Ong Sing Goh, C. C Fung, K.W Wong, and A Depickere, "Multilevel Natural Language Query Approach for Conversational Agent System," *International Journal of Computer Science*, vol. 33, pp. 7-13, 2007.

[18] Ong Sing Goh and Chun Che Fung, "Automated Knowledge Extraction from Internet for a Crisis Communication Portal," in *First International Conference on Natural Computation*. Changsha, China: Lecture Notes in Computer Science (LNCS), 2005, pp. 1226-1235.

[19] E. Isaacs, A. Walendowski, S. Whittaker, D.J. Schiano, and C. Kamm, "The Character, Functions, and Styles of Instant Messaging in the Workplace," presented at CSCW '02, NY, 2002.

[20] Daniel Avrahami and Scott E. Hudson, "Communication Characteristics of Instant Messaging: Effects and Predictions of Interpersonal Relationships," presented at CSCW'06, Banff, Alberta, Canada, 2006.

[21] A. T Rabin, "Determining difficulty levels of text written in languages other than English," in *Readability: Its past, present, and future*, L. Zakaluk and S. J. Samuels, Eds. Newark, DE: International Reading Association, 1988.

[22] A. L Sherman, *Analytics of literature: A manual for the objective study of English prose and poetry.* Boston: Ginn & Co, 1893.

[23] Robert Gunning, (2004),Plain Language At Work Newsletter, [Online]. Available: http://www.impact-information.com/impactinfo/newsletter/plwork08.htm

[24] W.H DuBay, (2004),The Principles of Readability, [Online]. Available: http://www.impact-information.com/impactinfo/readability02.pdf

[25] Jared M. Spool, Tara Scanlon, Will Schroeder, Carolyn Snyder, and Terri DeAngelo, *Web Site Usability: A Designer's Guide*: Morgan Kaufmann Publishers, Inc., San Francisco, 1999.

[26] Teresa Sikorski and ames Allen, "A task-based evaluation of the TRAINS-95 dialogue system," presented at ECAI Workshop on Dialogue Processing in Spoken Language Systems., 1996.

[27] George Kingsley Zipf, *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison- Wesley., 1949.

[28] Hugh Loebner, (2006),In Response, [Online]. Available: http://www.loebner.net/Prizef/In-response.html

[29] Alice, (2005),Artificial Linguistic Internet Computer Entity, [Online]. Available: http://www.alicebot.org