

Linguistic Profiling and Behavioral Drift in Chat Bots

Nawaf Ali

Computer Engineering and Computer
Science Department
J. B. Speed School of Engineering
University of Louisville
Louisville, KY. USA
ntali001@louisville.edu

Derek Schaeffer

Computer Engineering and Computer
Science Department
J. B. Speed School of Engineering
University of Louisville
Louisville, KY. USA
dwscha02@louisville.edu

Roman V. Yampolskiy

Computer Engineering and Computer
Science Department
J. B. Speed School of Engineering
University of Louisville
Louisville, KY. USA
roman.yampolskiy@louisville.edu

Abstract

When trying to identify the author of a book, a paper, or a letter, the object is to detect a style that distinguishes one author from another. With recent developments in artificial intelligence, chat bots sometimes play the role of the text authors. The focus of this study is to investigate the change in chat bot linguistic style over time and its effect on authorship attribution. The study shows that chat bots did show a behavioral drift in their style. Results from this study imply that any non-zero change in lingual style results in difficulty for our chat bot identification process.

I. Introduction

Biometric identification is a way to discover or verify the identity of who we claim to be by using physiological and behavioral traits (Jain, 2000). To serve as an identifier, a biometric should have the following properties: (a) Universality, which means that a characteristic should apply to everybody, (b) uniqueness, the characteristics should be unique to each individual being studied, (c) permanence, the characteristics should not change over time in a way that will obscure the identity of a person, and (d) collectability, the ability to measure such characteristics (Jain, Ross & Nandakumar, 2011).

Biometric identification technologies are not limited to fingerprints. Behavioral traits associated with each human provide a way to identify the person by a biometric profile. Behavioral biometrics provides an advantage over traditional biometrics in that they can be collected unbeknownst to the user under investigation (Yampolskiy & Govindaraju, 2008). Characteristics pertaining to language, composition, and writing style, such as particular syntactic and structural layout traits, vocabulary usage and richness, unusual language usage, and stylistic traits remain relatively constant. Identifying and learning these characteristics is the primary focus of authorship authentication (Orebaugh, 2006).

Authorship identification is a research field interested in finding traits, which can identify the original author of the document. Two main subfields of authorship

identification are: (a) Authorship recognition, when there is more than one author claiming a document, and the task is to identify the correct author based on the study of style and other author-specific features. (b) Authorship verification, where the task is to verify that an author of a document is the correct author based on that author's profile and the study of the document (Ali, Hindi & Yampolskiy, 2011). The twelve Federalist papers claimed by both Alexander Hamilton and James Madison are an example for authorship recognition (Holmes & Forsyth, 1995). Detecting plagiarism is a good example of the second type. Authorship verification is mostly used in forensic investigation.

When examining people, a major challenge is that the writing style of the writer might evolve and develop with time, a concept known as behavioral drift (Malyutov, 2005). Chat bots, which are built algorithmically, have never been analyzed from this perspective. A study on identifying chat bots using Java Graphical Authorship Attribution Program (JGAAP) has shown that it is possible to identify chat bots by analyzing their chat logs for linguistics features (Ali, Hindi & Yampolskiy, 2011).

A. Chat bots

Chat bots are computer programs mainly used in applications such as online help, e-commerce, customer services, call centers, and internet gaming (Webopedia, 2011).

Chat bots are typically perceived as engaging software entities, which humans may communicate with, attempting to fool the human into thinking that he or she is talking to another human. Some chat bots use Natural Language Processing Systems (NLPS) when replying to a statement, while majority of other bots are scanning for keywords within the input and pull a reply with the most matching keywords (Wikipedia, 2011).

B. Motivations

The ongoing threats by criminal individuals have migrated from actual physical threats and violence to another dimension, the Cyber World. Criminals try to steal others information and identity by any means. Researchers are

following up and doing more work trying to prevent any criminal activities, whether it is identity theft or even terrorist threats.

II. Application and Data Collection

Data was downloaded from the Loebner prize website (Loebner, 2012), in which a group of human judges from different disciplines and ages are set to talk with the chat bots, and the chat bots get points depending on the quality of the conversation that the chat bot produces. A study was made on chat bot authorship with data collected in 2011 (Ali, Hindi & Yampolskiy, 2011); the study demonstrated the feasibility of using authorship identification techniques on chat bots. The data in the current study was collected over a period of years. Our data only pertained to chat bots that were under study in (Ali, Hindi & Yampolskiy, 2011), which is why this study does not cover every year of the Loebner contest, which started in 1996. Only the years, that contain the chat bots under study, were used in this research.

III. Data Preparation

The collected data had to be preprocessed by deleting unnecessary labels like the chat bot name, and time-date of conversation (Fig. 1). A Perl script was used to clean the files and split each chat into two text files, one for the chat bot under study, the other for the human judge. The judge part was ignored, and only the chat bot text was analyzed.

Copyright (C) 2003 Cambridge Center for Behavioral Studies. All rights reserved.
 ALICE Silver Edition conversing with Judge 06

```

2003-10-18-15-40-46 PROGRAM: Hello? Is anyone there?
2003-10-18-15-41-00 JUDGE: Hi. Who are you?
2003-10-18-15-41-05 PROGRAM: Hello there. I am Alice. Who are you?
2003-10-18-15-41-22 JUDGE: I'm Marilyn. What are you doing here?
2003-10-18-15-41-27 PROGRAM: Hi there Marilyn. I am chatting with people online.
What are you doing?
2003-10-18-15-41-48 JUDGE: Well at the moment I'm talking to you Alice?
2003-10-18-15-41-53 PROGRAM: Well that's okay. Sounds like fun.
2003-10-18-15-42-11 JUDGE: It could be. What shall we chat about then?
2003-10-18-15-42-16 PROGRAM: Under what circumstances? Quite a lot of things.
2003-10-18-15-42-51 JUDGE: As we have only just met perhaps we should just get to
know each other.
2003-10-18-15-42-56 PROGRAM: Do you mean your name is we have only just met perhaps
we should just get to know each other?
2003-10-18-15-43-15 JUDGE: No. My name is Marilyn
2003-10-18-15-43-20 PROGRAM: What is your real name? I am always glad to make new
friends, Marilyn.
2003-10-18-15-43-43 JUDGE: That is my real name. Are you really called Alice?
2003-10-18-15-43-48 PROGRAM: Who gave you your real name? I don't know whether or
not I am called. I am a Mammal.. Really.
2003-10-18-15-44-19 JUDGE: My parents gave me my name.
2003-10-18-15-44-24 PROGRAM: Such an original name!
2003-10-18-15-44-42 JUDGE: Thank you. Who are you called after?
2003-10-18-15-44-47 PROGRAM: you are quite welcome. Human beings. I have never been
asked that before. How did you hear about Alice?
  
```

Fig. 1. Sample conversation between a chat bot and a judge.

IV. Chat Bots used.

Eleven chat bots were used in the initial experiments: Alice (ALICE, 2011), CleverBot (CleverBot, 2011), Hal (HAL, 2011), Jeoney (Jeoney, 2011), SkyNet (SkyNet, 2011), TalkBot (TalkBot, 2011), Alan (Alan, 2011), MyBot (MyBot, 2011), Jabberwock (Jabberwock, 2011), Jabberwacky (Jabberwacky, 2011), and Suzette (Suzette, 2011). These were our main baseline that we intend to

compare to the chat bots under study, which were: Alice, Jabberwacky, and Jabberwock

V. Experiments

The experiments were conducted using RapidMiner (RapidMiner, 2011). A model was built for authorship identification that will accept the training text and create a word list and a model using the Support Vector Machine (SVM) (Fig 2), and then this word list and model will be implemented on the test text, which is, in our case, data from the Loebner prize site (Loebner, 2012).

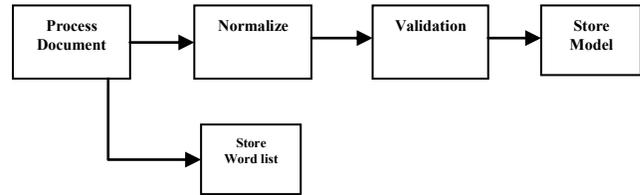


Fig. 2. Training model using Rapid Miner.

In Fig. 3 we use the saved word list and model as input for the testing stage, and the output will give us the percentage prediction of the tested files.

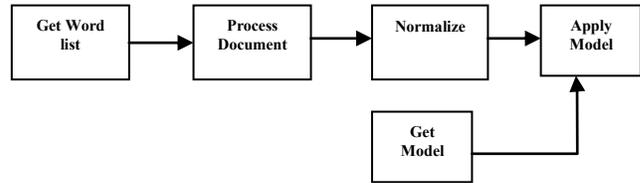


Fig. 3. Testing stage using Rapid Miner.

The data was tested using two different saved models, one with a complete set of chat bots (eleven bots) in the training stage, and the second model was built with training using only the three chat bots under study.

When performing the experiments, the model output is confidence values, in which, values reflecting how confident we are that this chat bot is identified correctly. Chat bot with highest confidence value (printed in boldface in all tables) is the predicted bot according to the model. Table 1 shows how much confidence we have in our tested data for Alice's text files in different years, when using eleven chat bots for training.

Table 1. Confidence level of Alice's files when tested with all eleven chat bots used in training

	2001	2003				2004	2005				
Alice	23.9	13	11	14.6	20	16.5	11	9.2	7	7.8	12
Hal	12.8	11	15	12.4	14	11.6	12	12	13	13	14
Jabberwacky	7.9	8	7.2	11.3	8.6	9.4	10	14	8.4	9	9.1
Alan	7.8	12	10	7.9	9.3	8.6	9	8.6	8.9	11	9.4
Suzette	6.2	7.5	7.5	6.8	7.1	8.3	10	10	11	10	10
SkyNet	5.9	9.4	6	5.1	4.5	6.6	6.9	6.4	7.1	6	5.9
MyBot	6.4	8.2	8.7	7.1	7.3	6.8	7.5	7.8	11	8.8	8.4
Cleverbot	11.8	12	11	11.9	10	12.8	14	12	12	12	12
Talkbot	6	7.1	7.9	8.2	6.8	7	6.6	6.1	7.5	8.6	6.1
Jeoney	6.1	6.6	9.4	8.6	6.2	7.7	6.9	6.2	8.2	6.7	5.7
Jabberwock	5.2	4.9	6.8	6.2	5.3	5	6.7	7.5	7.4	7.4	7.2

Table 2 shows the confidence level of Alice's files when using only the three chat bots under study.

Table 2. Confidence level of Alice's files when tested with only three chat bots used in training.

	2001	2003	2004				2005				
Alice	57.9	46.9	43	46	52	53	32	33	31	34	36
Jabberwac	23.1	31.7	31	29	28	28	36	36	33	36	33
Jabberwock	19	21.4	26	25	20	19	33	32	36	30	31

Fig. 4 shows the results of testing the three chat bots over different years when training our model using all eleven chat bots.

The results in Fig. 5 comes from the experiments that uses a training set based on the three chat bots under study, Alice, Jabberwacky, and Jabberwock. Jabberwock did not take part in the 2005 contest.

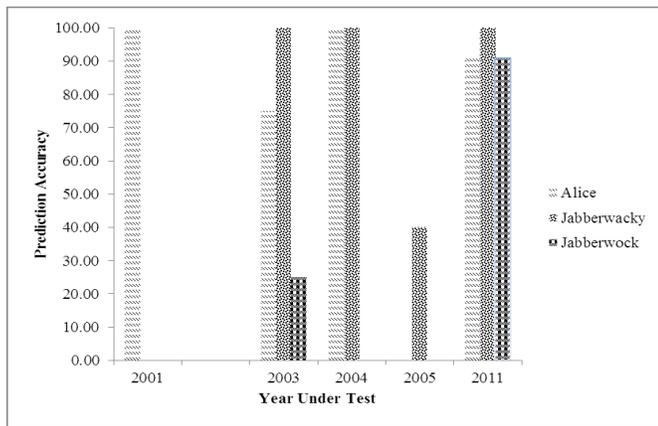


Fig. 4. Identification percentage over different years using all eleven chat bots for training.

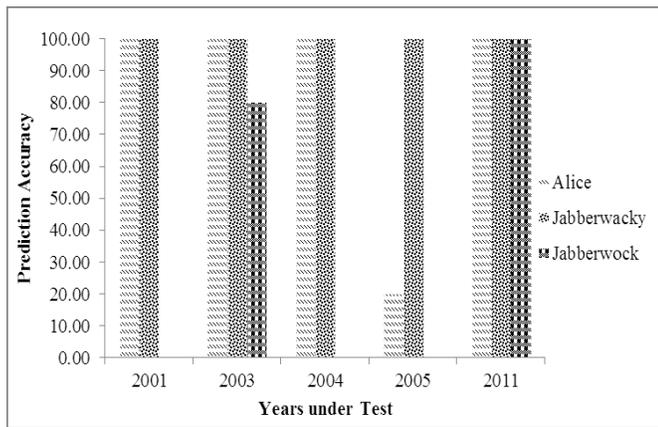


Fig 5. Identification percentage over different years using only the three chat bots under study for training.

Table 3 shows the confidence level of Jabberwacky's files values when tested with the complete set of eleven chat bots.

Table 3. Confidence level of Jabberwacky's files when tested with all 11 chat bots used in training.

	2002	2003		2004		2005				
Hal	13	8.5	11	8.1	9.7	13	11	12	7.2	9.5
Jabberwacky	9	18	17	21	16	9	10	14	18	21
Alan	10	8.1	7.8	7.1	8.2	11	10	11	8	6.3
Suzette	10	8.4	9.5	6.5	9.8	10	11	7.4	9	9.9
Skynet	6	5.6	5.2	5.2	5.8	6	6.9	5.3	6.2	5.2
Mybot	8.8	9	10	6.9	11	8.8	12	11	14	7.2
Cleverbot	12	16	15	17	15	12	15	14	15	17
Alice	7.8	6.3	4.6	7.8	6	7.8	6.3	5.3	5.7	6
Talkbot	8.6	6.3	6.7	7.9	8.2	8.6	5.9	8	6.9	8.7
Jeeney	6.7	5.8	5.2	7.1	5.1	6.7	4.8	6	5.1	4.9
Jabberwock	7.4	7	8	6.1	5.8	7.4	7.8	6	5.5	5

Table 4 shows the confidence level of Jabberwock's files when all the chat bots are used for training.

Table 4. Confidence level of Jabberwock's files when tested with all eleven chat bots used in training.

	2002	2003				
Hal	8	9.8	9.8	8.9	10	11
Jabberwacky	16	11	8.3	12	9.2	10
Alan	8.5	11	9.2	10	8.6	8.5
Suzette	7.4	7.5	11	8.8	6.3	7.5
Skynet	6.5	6.2	5	5.2	5.7	6.6
Mybot	7.3	8.9	7.8	6.7	9.1	8.1
Cleverbot	14	13	10	11	12	10
Alice	10	8.6	5.9	7.1	10	8.1
Talkbot	7	7.6	7.8	7.2	7.8	9.4
Jeeney	8.8	9.1	6	12	10	11
Jabberwock	7.1	7.8	18	10	11	9.5

VI. Conclusions and Future Work

The initial experiments conducted on the collected data did show a variation between chat bots, which is expected. It is not expected that all chat bots will act the same way, since they have different creators and different algorithms.

Some chat bots are more intelligent than others; the Loebner contest aims to contrast such differences. Alice bot showed some consistency over the years under study, but in 2005 Alice's style was not as recognizable as in other years. While Jabberwacky performed well for all years when training with just three bots and was not identified in 2001 when the training set contained all eleven chat bots for training, Jabberwacky gave us a 40% correct prediction in 2005. Jabberwock, the third chat bot under study here, was the least consistent compared to all other bots, and gave 0% correct prediction in 2001 and 2004, and 91% for 2011, which may indicate that Jabberwock's vocabulary did improve in a way that gave him his own style.

With three chat bot training models, Jabberwacky was identified 100% correctly over all years. Alice did well for all years except for 2005, and Jabberwock was not identified at all in 2001 and 2004.

With these initial experiments, we can state that some chat bots do change their style, most probably depending on the intelligent algorithms used in initializing conversations. Other chat bots do have a steady style and do not change over time.

More data is required to get reliable results; we only managed to obtain data from the Loebner prize competition, which in some cases was just one 4KB text file. With sufficient data, results should be more representative and accurate.

Additional research on these chat bots will be conducted, and more work on trying to find specific features to identify the chat bots will be continued. This is a burgeoning research area and still much work need to be done.

References

- Alan. (2011). AI Research. Retrieved June 10, 2011, from http://www.a-i.com/show_tree.asp?id=59&level=2&root=115
- Ali, N., Hindi, M., & Yampolskiy, R. V. (2011). *Evaluation of authorship attribution software on a Chat bot corpus*. XXIII International Symposium on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosnia and Herzegovina, 1-6.
- ALICE. (2011). ALICE. Retrieved June 12, 2011, from <http://alicebot.blogspot.com/>
- CleverBot. (2011). CleverBot Retrieved July 5, 2011, from <http://cleverbot.com/>
- HAL. (2011). AI Research. Retrieved June 16, 2011, from http://www.a-i.com/show_tree.asp?id=97&level=2&root=115
- Holmes, D. I., & Forsyth, R. S. (1995). The Federalist Revisited: New Directions in Authorship Attribution. *Literary and Linguistic Computing*, 10(2), 111-127.
- Jabberwacky. (2011). Jabberwacky-live chat bot-AI Artificial Intelligence chatbot. Retrieved June 10, 2011, from <http://www.jabberwacky.com/>
- Jabberwock. (2011). Jabberwock Chat. Retrieved June 12, 2011, from <http://www.abenteuermedien.de/jabberwock/>
- Jain, A. (2000). Biometric Identification. *Communications of the ACM*, 43(2), 91-98.
- Jain, A., Ross, A. A., & Nandakumar, K. (2011). *Introduction to Biometrics*: Springer-Verlag New York, LLC.
- Jeeneey. (2011). Artificial Intelligence Online. Retrieved March 11, 2011, from <http://www.jeeneey.com/>
- Loebner, H. G. (2012). Home Page of The Loebner Prize. Retrieved Jan 3, 2012, from <http://loebner.net/Prize/loebner-prize.html>
- Malyutov, M. B. (2005). Authorship attribution of texts: a review. *Electronic Notes in Discrete Mathematics*, 21, 353-357.
- MyBot. (2011). Chatbot Mybot, Artificial Intelligence. Retrieved Jan 8, 2011, from <http://www.chatbots.org/chatbot/mybot/>
- Orebaugh, A. (2006). *An Instant Messaging Intrusion Detection System Framework: Using character frequency analysis for authorship identification and validation*. 40th Annual IEEE International Carnahan Conference Security Technology, Lexington, KY.
- RapidMiner. (2011). Rapid- I. Retrieved Dec 20, 2011, from <http://rapid-i.com/>
- SkyNet. (2011). SkyNet - AI. Retrieved April 20, 2011, from <http://home.comcast.net/~chatterbot/bots/AI/SkyNet/>
- Suzette. (2011). SourceForge ChatScript Project. Retrieved Feb 7, 2011, from <http://chatscript.sourceforge.net/>
- TalkBot. (2011). TalkBot- A simple talk bot. Retrieved April 14, 2011, from <http://code.google.com/p/talkbot/>
- Webopedia. (2011). What is chat bot? A Word Definition from the Webpedia Computer Dictionary. Retrieved June 20, 2011, from www.webopedia.com/TERM/C/chat_bot.html
- Wikipedia. (2011). Chatterbot- Wikipedia, the free encyclopedia. Retrieved June 22, 2011, from www.en.wikipedia.org/wiki/Chatterbot
- Yampolskiy, R. V., & Govindaraju, V. (2008). Behavioral Biometrics: a Survey and Classification. *International Journal of Biometrics (IJBM)*. 1(1), 81-113.