Mimicry in Online Conversations: An Exploratory Study of Linguistic Analysis Techniques

Tom Carrick, Awais Rashid, Paul J Taylor

Security Lancaster / CREST

Lancaster University

Lancaster, United Kingdom

Email: {t.carrick, a.rashid, p.j.taylor} @lancaster.ac.uk

Abstract—A number of computational techniques have been proposed that aim to detect mimicry in online conversations. In this paper, we investigate how well these reflect the prevailing cognitive science model, i.e. the Interactive Alignment Model. We evaluate Local Linguistic Alignment, word vectors, and Language Style Matching and show that these measures tend to show the features we expect to see in the IAM, but significantly fall short of the work of human classifiers on the same data set. This reflects the need for substantial additional research on computational techniques to detect mimicry in online conversations. We suggest further work needed to measure these techniques and others more accurately.

I. INTRODUCTION

When interacting with others, humans show a tendency to mimic one another's behavior, including speech [1]–[4], physical actions [5] and written text [6], [7]. Existing research has shown that people who mimic each other's language have more successful outcomes in joint tasks [8], and that decreasing mimicry can be an indicator of deception [9]. Hence, detection of mimicry can be a valuable indicator in dyadic conversations as well as group dynamics in collaborative settings [10]. Potential applications include promoting cooperation, detecting malintent, and identifying leaders.

There are a number of measures designed to capture mimicry in text, and some efforts have been made to compare and trial some of these techniques [11]. However, few techniques are grounded in what cognitive science knows about how mimicry works within humans. That is to say, few of the existing techniques are designed to measure the linguistic components of verbal mimicry as specified by contemporary models from cognitive science.

This paper addresses this limitation by contrasting how well existing linguistic analysis techniques work to detect mimicry in relation to a prominent model of human interpersonal mimicry, the Interactive Alignment Model. Further, we consider some existing techniques that have not previously been used to detect mimicry and discuss their viability. The novel contributions of this paper are as follows:

- We compare two existing mimicry detection techniques and how they map to the Interactive Alignment Model, Local Linguistic Alignment [12] and Language Style Matching [13].

- We evaluate and discuss the feasibility of using word vectors [14] to measure semantic mimicry.
- We find how one measure, Language Style Matching [13], that does not appear to fit into the model, may nevertheless have a place in the model.

The rest of this paper is structured as follows. In Section II, we briefly introduce the Interactive Alignment Model. Section III provides a brief introduction to each measure tested in the paper, and defines the layer of the IAM (if any) in which the measure appears to reside. Section IV describes the data set used in our analysis. Section V presents the results of our analysis and highlights how current techniques fit into the IAM in practice. Finally, Section VI concludes the paper and identifies directions for future work.

II. THE INTERACTIVE ALIGNMENT MODEL

In order to achieve conversation, two or more interactants must work together to establish a joint understanding of what they are talking about. The Interactive Alignment Model (IAM) [15], [16], shown in Fig. 1 is an attempt to model such alignment as it occurs over time in conversations. It consists of six layers, which with the exception of the top layer, the situation model, are all representations of different linguistic levels at the same time, as understood by each speaker within an interaction. In basic terms, these levels are:

1) The situation model — a representation of the situation being discussed
2) The semantic representation — the meaning of the utterance in terms of what it stands to convey.
3) The syntactic representation — the sentence structure
4) The lexical representation — the choice of words used
5) The phonological representation — the abstract representation of word sounds
6) The phonetic representation — the sounds of an utterance

The theory states that alignment is automatic and that conversation partners will 'interactively align' their speech towards each other at every level of the model. The observable behavior of such alignment is greater mimicry at each layer.

The IAM also postulates how such alignment emerges over time. It suggests that alignment 'percolates' between layers such that increased alignment in one layer will lead to

increased alignment in other layers. For example, tightening the semantic representation will naturally restrict the number of words and syntactic structures that can be used to convey that meaning. Similarly, particular word choices may shape the kinds of semantic representations that become prominent. In short, alignment at one level enables a listener to fill in the gaps and appropriately mimic at another level.

The phonological and phonetic representations are not discussed in this paper as they relate purely to the sound of speech, representations that do not exist in the written text of online conversations. Further, the situation model is not included as there are currently no measures that can be used for finding alignment at the situational level.

III. ANALYSIS TECHNIQUES

There are many techniques for finding mimicry, but few of them specifically target any layer of the IAM. We chose one measure for each layer, as well as Language Style Matching to have a technique that could fit into the lexical layer, or could be seen to work outside of the IAM.

Techniques were chosen mostly to try to fit into a layer of the IAM. Simplicity of implementation and access to source implementations was also a factor, and modern techniques were favoured over older ones.

A. Local Linguistic Alignment

Local Linguistic Alignment (LLA) is a simple, dual purpose method that can work at both the lexical and syntactic layers [11], [12]. LLA was chosen specifically as it was the only measure the authors could find where it is explicitly stated to work at the lexical and syntactic layers.

Lexical Indiscriminate Local Linguistic Alignment (LILLA) works at the lexical layer. It gives the probability that a word from the prime document will appear at a particular position in the target document. It is defined as:

$$LILLA(target, prime) = \frac{p(target|prime)}{p(target)}$$

Syntactic Indiscriminate Local Linguistic Alignemnt (SILLA) works at the syntactic layer. It works in exactly the same way as LILLA, but instead of comparing words, each sentence is annotated with a phrase structure tree, and LLA is run on these instead. As such, it measures the probability of a phrase structure subtree in the prime document appearing occurring at a given position in the target document.

B. Word Vectors

Word2vec [14], [17] is a machine learning technique that, given a word, tries to predict the two adjacent words. Each word is then mapped to vector space and can be directly compared. The closer the two word vectors to each other, the more semantically related they are.

While word vectors have not been used to measure mimicry before, we posit that it may give good results for finding mimicry as it is used for finding semantic similarity in text, we would expect it to also find semantic mimicry.

C. Language Style Matching

Language Style Matching (LSM) [13], [18] is a measure that counts the usage of function words by an interlocutor. For each class of function word (e.g., determiners, pronouns, adpositions), the number of words is calculated as a percentage of the total document, which is the score for each class. Scores are calculated between the dyad using a weighted difference, shown below, and these scores averaged to produce a single LSM score for that dyad.

$$LSM_{class} = 1 - [(|class_1 - class_2|)]/(class_1 + class_2 + .0001)$$

IV. METHOD

The data set consists of 103 dyadic conversations taken from Twitter. Tweets were found via Twitter's search API, using English language and being a reply as criteria. We then work up the chain of tweets until we hit the end, or a third interlocutor is found, at which point the conversation is cut off and kept, as long as there are no fewer than 10 tweets in the conversation. 200 tweets were found using this method. Of these, 62 were removed as they were not true conversations (either monologues or collaborative fiction writing/roleplay). 18 conversations were removed as they had significant amounts of non-English text. 17 conversations involved bots and were also removed.

The data was processed as follows:

- Screen names at the start of the message were removed. These are just the names of the interlocutors and only add noise.
- All URLs were replaced with the string '[URL]'.
- HTML entities such as & were converted to plain text (&).
- Any adjacent messages by the same author were merged together such that each message always alternates between the two interlocutors.

Twitter was chosen as it is relatively easy to extract data from, and the short utterance and conversation length make it more reasonable for humans to classify, as with longer texts they may find more difficulty in keeping less recent utterances in mind while evaluating more recent ones.

The data set is small, which is both an advantage as it is possible to classify the data manually and a disadvantage in that the sample size is smaller. We did not actively seek out mimicry by targeting our search as it was important to have conversations that would test negatively for mimicry as well as positively.

The data was classified by three researchers with backgrounds in mimicry independent of the authors of this paper. Each conversation was classified for lexical, syntactic and semantic mimicry on a three point scale where 0 represents no mimicry, 1 represents low levels of mimicry at that layer, and 2 represents high levels of mimicry. Classifications from each author were averaged together to give each conversation three scores between 0 and 2.
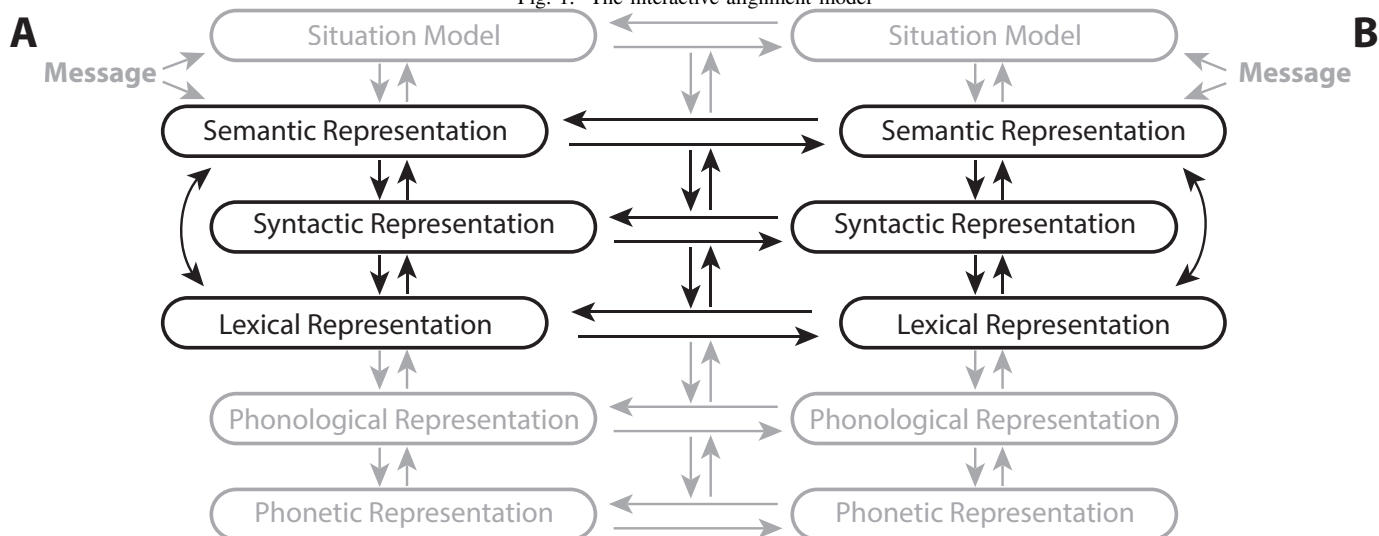
Fig. 1. The interactive alignment model

A **B**

TABLE I
CORRELATIONS OF CLASSIFIERS WITH EACH OTHER

Pair	Lexical		Syntactic		Semantic	
	r	p	r	p	r	p
AB	0.28	0.779	0.36	< 0.001	0.19	0.058
AC	0.32	0.001	0.35	< 0.001	0.12	0.220
BC	-0.02	0.863	0.29	0.003	0.10	0.312

TABLE II
CORRELATIONS OF CLASSIFIER SCORES AND TECHNIQUES

Layer, Technique	r	p
Lexical, LILLA	0.27	0.005
Syntactic, SILLA	0.19	0.057
Semantic, Word2vec	0.13	0.178
Lexical, LSM	0.14	0.158
Syntactic, LSM	0.31	0.002
Semantic, LSM	-0.13	0.202
N/A (mean scores), LSM	0.16	0.111

TABLE III
CORRELATIONS OF TECHNIQUES

Technique	r	p
LILLA, SILLA	0.26	0.007
LILLA, Word2vec	-0.30	0.005
SILLA, Word2vec	-0.82	< 0.001

TABLE IV
CORRELATIONS OF HUMAN CLASSIFICATIONS

Layer	r	p
Lexical, Syntactic	0.57	< 0.001
Lexical, Semantic	0.33	< 0.001
Syntactic, Semantic	0.28	0.005

Each measure was then run against the data such that each message from author A is compared with each message from author B that occurred before the original message. [1]

V. RESULTS AND DISUCSSION

Table I shows the correlations between the scores of each pair of classifiers. In general, the human classifiers agreed with each other, but the agreement is fairly weak. They most strongly agreed on syntactic mimicry. Additionally there seemed to be more agreement on lexical mimicry than semantic, but p values are too high to make a definitive judgment.

Table II shows the correlations between the mean scores from the human classifiers with the measure they should fall into. Additionally means of all human classifier scores were used as a comparison with LSM, and LSM was compared with

all other scores, to see where it best fits within the IAM. We see that scores from human classifiers show positive but weak correlation with the scores from their respective techniques. Additionally, LSM has a much better correlation with our syntactic scores than any other. Despite operating at the word level, LSM appears to be more of a syntactic measure than a lexical measure. This makes some sense when considering that choosing different function words may force a change of syntax.

Table III shows the correlations of the computational methods with each other, to observe the IAM's prediction that increased levels of mimicry on one layer results in a higher probability of mimicry in other adjacent layers. The comparison of LILLA and SILLA is quite promising, however the strong negative correlations when word vectors are involved are both concerning and interesting, as this is the opposite of what we expect.

Table IV shares the same idea, but compares the human classifiers' lexical, syntactic and semantic scores with each other. Here, the results are much closer to what we expect, with moderate positive correlations between all layers. The

[1] Source code available at https://github.com/knyghty/linguistic-analysis.

contrast with the word vector results from Table III suggest that word vectors may not be a good solution for finding semantic mimicry, but doesn't explain the strong negative correlation between SILLA and Word2vec. The strong correlation between human classifiers' lexical and syntactic scores compared with the weaker correlation between LILLA and SILLA could suggest that humans are better suited for this task.

As results involving the semantic layer generally lead to lower levels of correlation, it could be argued that discerning semantic mimicry is a harder task than for the lexical and syntactic layers. However, it should be noted that there is still a weak to moderate positive correlation.

Even defining semantic mimicry was found to be a difficult task. It seems to make little sense to say that semantic mimicry is present where the same subject is discussed, as there are few situations where interlocutors will have a conversation about different things at the same time. It also makes little sense to define it as agreement with the other interaction partner, as agreement could be natural, rather than evidence of mimicry. Instead, we defined semantic mimicry as the people in the conversation communicating using similar concepts. This seems to be difficult for humans to analyse, especially from short twitter conversations, and could help to explain why computational methods also find this task difficult.

VI. CONCLUSION AND FUTURE WORK

We have presented some initial findings from our study into mimicry in online conversations. We have found that current computational methods are not especially good at finding mimicry in accordance with the Interactive Alignment Model, and that human experts currently seem to do a somewhat better job.

Our results further show support for the IAM's theory that mimicry in one layer increases the chance of mimicry in the other layers.

More research is needed, particularly in the methods used to detect mimicry by computational analysis. There are many more measures, such as the Hierarchical Alignment Model [19], Zelig [20] and RepDecay [21], some of which should fit neatly into the lexical and syntactic layers, and some that do not have any basis in the IAM. Our data set is small, and using a larger data set, ideally with a greater number of human classifiers would be useful for a more extensive investigation of the phenomenon highlighted in this paper.

More work is especially required to find if word vectors or related techniques can be used to find semantic mimicry with different methodology than ours, perhaps by parsing the text differently or experimenting with differently trained models. More research in general is suggested in finding ways to capture mimicry at the semantic layer, and so far there is very little work looking at alignment at the situation layer.

ACKNOWLEDGMENT

REFERENCES

[1] W. J. Levelt and S. Kelter, "Surface form and memory in question answering," *Cognitive psychology*, vol. 14, no. 1, pp. 78–106, 1982.

[2] J. K. Bock, "Syntactic persistence in language production," *Cognitive psychology*, vol. 18, no. 3, pp. 355–387, 1986.

[3] K. Bock, "Closed-class immanence in sentence production," *Cognition*, vol. 31, no. 2, pp. 163–186, 1989.

[4] S. Garrod and A. Anderson, "Saying what you mean in dialogue: A study in conceptual and semantic coordination," *Cognition*, vol. 27, no. 2, pp. 181–218, 1987.

[5] F. J. Bernieri, "Coordinated movement and rapport in teacher-student interactions," *Journal of Nonverbal behavior*, vol. 12, no. 2, pp. 120–138, 1988.

[6] C. Danescu-Niculescu-Mizil and L. Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs," *CoRR*, vol. abs/1106.3077, 2011.

[7] L. E. Scissors, A. J. Gill, and D. Gergle, "Linguistic mimicry and trust in text-based cmc," in *CSCW*, 2008.

[8] B. H. Richardson, P. J. Taylor, B. Snook, S. M. Conchie, and C. Bennell, "Language style matching and police interrogation outcomes.," *Law and human behavior*, vol. 38, no. 4, p. 357, 2014.

[9] P. J. Taylor, C. J. Dando, T. C. Ormerod, L. J. Ball, M. C. Jenkins, A. Sandham, and T. Menacere, "Detecting insider threats through language change.," *Law and human behavior*, vol. 37, no. 4, p. 267, 2013.

[10] L. E. Scissors, A. J. Gill, K. Geraghty, and D. Gergle, "In cmc we trust: The role of similarity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2009, pp. 527–536.

[11] Y. Xu and D. Reitter, "An evaluation and comparison of linguistic alignment measures," 2015, pp. 58–67.

[12] Y. Wang, D. Reitter, and J. Yen, "Linguistic adaptation in conversation threads: Analyzing alignment in online health communities," *ACL 2014*, p. 55, 2014.

[13] M. E. Ireland and J. W. Pennebaker, "Language style matching in writing: Synchrony in essays, correspondence, and poetry.," *Journal of personality and social psychology*, vol. 99, no. 3, p. 549, 2010.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *ArXiv preprint arXiv:1301.3781*, 2013.

[15] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and brain sciences*, vol. 27, no. 02, pp. 169–190, 2004.

[16] S. Garrod and M. J. Pickering, "Why is conversation so easy?" *Trends in cognitive sciences*, vol. 8, no. 1, pp. 8–11, 2004.

[17] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases

and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.

[18] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker, "Language style matching predicts relationship initiation and stability," *Psychological Science*, vol. 22, no. 1, pp. 39–44, 2011.

[19] G. Doyle, D. Yurovsky, and M. C. Frank, "A robust framework for estimating linguistic alignment in twitter conversations," in *Proceedings of the 25th International World Wide Web Conference*, 2016.

[20] S. Jones, R. Cotterill, N. Dewdney, K. Muir, and A. Joinson, "Finding zelig in text: A measure for normalising linguistic accommodation," Coling, 2014.

[21] D. Reitter, F. Keller, and J. D. Moore, "Computational modelling of structural priming in dialogue," in *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, Association for Computational Linguistics, 2006, pp. 121–124.