Multimodal Approach for Automatic Recognition of Machiavellianism

Zahra Nazari, Gale Lucas, Jonathan Gratch
University of Southern California, Institute for Creative Technologies
Los Angeles, California
Email: {zahra, lucas, gratch}@ict.usc.edu

Abstract—Machiavellianism, by definition, is the tendency to use other people as a tool to achieve one's own goals. Despite the large focus on the Big Five traits of personality, this anti-social trait is relatively unexplored in the computational realm. Automatically recognizing anti-social traits can have important uses across a variety of applications. In this paper, we use negotiation as a setting that provides Machiavellians with the opportunity to reveal their exploitative inclinations. We use textual, visual, acoustic, and behavioral cues to automatically predict High vs. Low Machiavellian personalities. These learned models have good accuracy when compared with other personality-recognition methods, and we provide evidence that the automatically-learned models are consistent with existing literature on this anti-social trait, giving evidence that these results can generalize to other domains.

Keywords—Machiavellianism; Classification; Personality; Dark Triad

I. INTRODUCTION

Affective computing techniques typically explore momentary states such as emotions, but people also have characteristic patterns of emotion, thought and behavior that persist over their entire life. Personality Computing [1] seeks to recognize, perception and synthesize such measurable individual characteristics. Within personality computing, automatic personality recognition attempts to infer the true personality of an individual from behavioral evidence. Such techniques have wide-ranging benefits such as targeting technology or advertising campaigns to the right potential customers, diagnosing chronic diseases such as depression, or identifying social groups that one may wish to engage with or avoid.

In this article, we address the challenge of automatically recognizing an anti-social personality trait known as Machiavellianism from multi-modal behaviors elicited in a competitive negotiation. Contrary to the huge volume of studies on Machiavellianism in clinical and social psychology, the automatic recognition of this exploitative personality has remained largely unexplored. Machiavellianism takes its name from Niccolò Machiavelli (1469-1527), the Italian scholar and advisor to the rulers of Florence. Machiavellians are known to be strategic, manipulative and goal-seeking. They are less affected by emotions but are more willing to exploit others’ emotions. Given anti-social and untruthful nature of this personality, it would be useful to screen for this trait in potential negotiation partners or in deciding whether to join or avoid certain social networks [2].

Previous work in the social sciences has identified a number of characteristics of the Machiavellian personality. From a trait standpoint, Machiavellians share characteristics with narcissism and psychopathy individuals [3]. From an emotion standpoint, they are often described as rational or cold, stay emotionally-detached from situations, and fail to “catch” the excitement of others [4]. Although Machiavellians may not feel emotions as strongly, they are comfortable using emotional displays to gain advantage, such as feigning emotion for strategic gain. For example they are comfortable ingratiating themselves to others and are particularly good at feigning dislike for someone they actually like [5]. From a communication standpoint, Machiavellians are often vague and reluctant to disclose information about themselves and conceal their true intentions [6]. These characteristics often benefit Machiavellians in competitive negotiations and several studies have shown that Machiavellians often earn more money in negotiations or other economic exchanges, at least in the short term [6].

Scant previous work in automatic recognition has primarily focused on textual features. For example, Sumner and colleagues used the Linguistic Inquiry and Word Count (LIWC) software [7] to recognize Machiavellianism from Twitter posts. They found that Machs tended to use more swear words, more negative words and fewer positive words. Cizber et al. used similar text tools in a different context [8]. Following an economic game, they asked participants to explain their decisions in writing. Machs explained their decisions in a more cool and rational manner, using far fewer emotional terms. The discrepancy between these findings likely reflects the differences in how Machs report that they feel, compared with how they behave in competitive face-to-face interactions (in that a Tweet is a social act but the explanation is not). In a negotiation setting, we would expect results similar to [2].

We are unaware of studies using automatic facial expressions techniques to predict Machs, but findings in the social sciences suggest that Machs may be more comfortable feigning emotional expressions to gain competitive advantage. Thus, as in the textual features, we predict that Machiavellians will show more emotional expressivity in face-to-face negotiations.

We are also unaware of studies examining vocal correlates of Machiavellianism; however social science findings suggest that Machs may be more comfortable using deception to gain advantage. Research on deceptive speech has identified differ-
ences in various acoustic features [9][10]. Thus, we consider several acoustic features to help identify Machs.

Finally, the social science literature suggests that Machs can be distinguished by the task-related behaviors they exhibit in a competitive situation. In the context of negotiation, Machs would be more willing to use negotiation tactics that gain them a competitive advantage. Thus, they would be less likely to reveal their own preferences, more willing to misrepresent their own preferences, and more motivated to learn about what their opponent wants. Further, we would expect Machs to end up with better deals. To capture this, we consider several features related to participants negotiation behavior.

Taking into account all of the shown differences between those High and Low in Machiavellian features in the literature, we train and test classifiers using textual, visual, acoustic and behavioral features. We are not only interested in predicting if a person is High or Low Mach but we also attempt to show these learned models are consistent with the known literature on Machiavellian personality.

The remainder of this paper is organized as follows: Section II describes the corpus, selected features and statistical analyses of the most predictive features, along with the prediction models and the experimental methodology. Results from the trained and tested models are presented in Section III and a discussion of the results is provided in Section IV.

II. EXPERIMENTS

A. Corpus

A total of 113 same-sex dyads (38 female dyads, 75 male dyads) were enrolled in this study. The recruitment of the participants was done through Craigslist and every participant was paid for completing the study. Participants were also given a financial incentive to win the negotiation: they earned tickets for a $100 lottery based on the quality of the deal they obtained. The average age of the 226 participants was 26.83 (SD = 12.77), among which 41.6% were racially identified as African-American, 32.3% as Caucasian, 8.8% as Hispanic, 8.0% as other, 6.2% as Asian, 3.1% as Native American/Hawaiian.

![Illustration of two participants engaged in the negotiation task](image)

**Fig. 1.** Illustration of two participants engaged in the negotiation task

**Design** Each dyad performed a negotiation task in which they role-played characters of negotiating over six antique items (three crates of LP records, two art deco lamps, and one art deco painting). Items had varied levels of values assigned to them. Participants were told that their task is to decide on how to split up these six items between the two of them, and if they fail to reach an agreement, the two of them will receive the same number of lottery tickets as they would have received for only one of their highest value items if they had reached an agreement.

There were two kinds of negotiation tasks defined: *distributive vs. integrative*. Dyads were randomly appointed to perform one of the two types of negotiation. In the two settings, items were either given the same level of value for both partners and different levels of value to each partner (Table 1). Specifically, in the *distributive* task, the LP records are the highest value items, for both participants, and the lamps have moderate value to them, whereas in the *integrative* task, for participant A, the records are assigned the highest value items and the lamps were given moderate value, but for participant B, the lamps have the highest value and the records have moderate value. In both settings, the painting is of little value to participant A, and of no value at all to participant B.

From the original 226 participants, 8 were removed for non-compliance with study procedures (e.g., obvious intoxication, starting the negotiation before reading the instructions), 5 dyads were excluded as they failed to reach an agreement in the negotiation, and an additional 28 participants were excluded from analyses due to failure to accurately report the preferences to which they had been assigned, after completing the study.

**Measure of Machiavellianism** Prior to the negotiation, all participants completed Christie and Geis’ 20-item Mach-IV scale [11]. Respondent’s rate different statements on a scale of 1 (strongly disagree) to 7 (strongly agree). Example questions include items positively associated with Machiavellianism “The best way to handle people is to tell them what they want to hear” and negatively associated with Machiavellianism “Honesty is the best policy.” Negative items are reverse-scored: a score of seven (i.e., strongly agree) is scored as one.

Each participant receives a Mach score that corresponds to the mean of all twenty items (reversing the scores of negative items). Among all participants mean was 2.71 (SD = .41). We used a common procedure of dividing participants into two classes for prediction. A median split was used to divide participants into High Machs and Low Machs. 90 participants fell in each category.

**Annotations** All dialogues were manually transcribed and annotated using several different dialog acts. For the present paper, three dialog acts are relevant (see the complete annotation scheme described here [12] ). Each *Offer* corresponds to a negotiator’s statement about division of items. For example, the statement “How about if I take two records and you take both lamps?” is a partial offer. This is considered partial in two distinct senses as it (i) fails to mention the painting, and (ii) it requests two records while leaving the third record unspecified.
which are used to define its 80 language categories. LIWC processes. The default LIWC dictionary includes 4,500 words towards automatic analysis of verbal behavior [7]. It recognizes these into two broad classes (I-like-ITEM, and I-don’t-like, etc.) as automatic language recognition is not capable of distinguishing such subtle distinctions. In summary, preference-assertions indicate an issue (e.g., records) and a sentiment (positive or negative) expressed toward that issue.

Generic-Dialog-Acts were used as the second set of the textual features. 45 dialog acts fall into this category and were normalized by the total number of dialog acts. The last textual feature, Offer-Ratio compares the number of offers a participant has relative to their opponent throughout the negotiation. It is calculated by dividing an individual’s offer count to the number of total offers in the negotiation.

Visual Features: Using FACET1, a facial expression recognition and analysis tool, seven expressions of primary emotions were extracted. FACET provides two sets of classification categories: Positive and Negative, and the basic emotions individually as Joy, Anger, Sadness, Surprise, Contempt, Disgust and Neutral. Here, we also propose a new feature by adding all the values from the seven primary emotions. We name this feature as facial expressivity and along with 10 other features we were provided with 11 total visual features.

Acoustic Features: We used COVAREP (v 1.2.0)2, an open source Matlab toolbox that provides an extensive selection of open-source robust and tested speech processing algorithms enabling comparative and cooperative research within the speech community. Following the standard procedure, we extracted the pitch(f0) in addition to six features that are often used to segment voice qualities on breath and tense dimensions, which are considered to be on opposite ends of the voice quality spectrum.

<table>
<thead>
<tr>
<th>Task</th>
<th>Part- ner</th>
<th>Rec- ords</th>
<th>Lamps</th>
<th>Painting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributive</td>
<td>A</td>
<td>High</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>High</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td>Integrative</td>
<td>A</td>
<td>High</td>
<td>Moderate</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Moderate</td>
<td>High</td>
<td>None</td>
</tr>
</tbody>
</table>

Generic-Dialog-Acts were used as the second set of the textual features. 45 dialog acts fall into this category and were normalized by the total number of dialog acts. The last textual feature, Offer-Ratio compares the number of offers a participant has relative to their opponent throughout the negotiation. It is calculated by dividing an individual’s offer count to the number of total offers in the negotiation.

Visual Features: Using FACET1, a facial expression recognition and analysis tool, seven expressions of primary emotions were extracted. FACET provides two sets of classification categories: Positive and Negative, and the basic emotions individually as Joy, Anger, Sadness, Surprise, Contempt, Disgust and Neutral. Here, we also propose a new feature by adding all the values from the seven primary emotions. We name this feature as facial expressivity and along with 10 other features we were provided with 11 total visual features.

**Fundamental frequency (Pitch or f0)** The fundamental frequency, or pitch was tracked by using residual harm on-

**Normalized Amplitude Quotient (NAQ)** The ratio between the maximum of the glottal flow and the minimum of its derivative, after being normalized by the fundamental frequency is given by the NAQ [14].

**Quasi-Open Quotient (QQO)** The QQO is measured by detecting the duration during which the glottal flow is 50% above the minimum flow; it is normalized by the local glottal period [15].

**H1-H2 ratio** The fundamental frequency relative to the second harmonic is given by the H1-H2 parameters. The H1-H2 ratio is best thought of as a descriptor of the open q uotient [16].

**Maxima Dispersion Quotient (MDQ)** The extent of the dispersion of the maxima derived from the wavelet decomposition is represented by the MDQ parameter [17].

1 http://emotient.com/products/FACET
2 http://covarep.github.io/covarep/
Peak Slope (PS) The Peak Slope is found by decomposing the speech signal into octave bands that are then regressed on the maximum amplitudes. Voice breathiness is measured by the slope coefficient [18].

ANN Glottal Open Quotient (OQ) A method that uses spectral features as input to artificial neural networks (ANNs) in order learn the mapping from spectral measurements to the time domain OQ values [19].

Behavioral Features: In addition to the more general features introduced above, negotiation features were considered to detect Machiavellian’s task-related behavior in a competitive situation. Before describing these features, we provide a formal representation of the negotiation setting used in the corpus.

Each of the two parties has a preference over the set of issues. \( w_i \) is the weight a participant holds for issue \( i \). Each issue has a set of discrete levels. A negotiation comes to an end if both participants agree on their assigned level for each issue (\( l_i \)). The resultant assigned weights gives each participant a utility \( u(\omega) \) in range of \([0,1]\), which is calculated as a weighted summation over the assigned levels of issues:

\[
u(\omega) = \sum_{i=1}^{n} w_i \cdot l_i \quad (1)
\]

Since the preferences (weights each individual carries for the issues) are privately known for each participant, some models are proposed to estimate these weights from a negotiator’s pattern of offers and/or preference-assertions throughout the negotiation [20].

In order to evaluate how accurate the predicted weights are, a standard measure is used for the accuracy of a profile estimate, the rank distance of the deals [21]. This metric compares the utility of all possible deals in the outcome space \( (\Omega) \), given the estimated \( u(\omega') \) and the actual weights \( u(\omega) \), and calculates the average number of conflicts in how deals are ranked using the estimated vs. actual utility function:

\[
d_r(u, u') = \frac{1}{|\Omega|} \sum_{\omega \epsilon \Omega \land \omega' \epsilon \Omega} c_{<\omega, \omega'>}(w, w') \quad (2)
\]

The function \( c \), the conflict indicator compares the rank of any pair of deals \( w \) and \( w' \) when calculated by the actual vs predicted weights, returns 0 if they have the same ranks, and returns 1 otherwise.

Misrepresentation Using the technique proposed by [20], we calculated a metric that predicts if an individual is expressing untruthful information about their intentions. Offer/sentiment divergence takes two steps to measure trustworthiness of a negotiator. First, it uses pattern of offers and preference-assertions as two channels to build a preference model based on each. Then it calculates the diversion between the two learned models using rank distance of the deals. The first model estimates a set of weights for the issues using positive and negative assertions stated by a participant and the second model learns the issue weights using an individual’s pattern of offers. In other words, offer/sentiment divergence uses the divergence between what people utter they like and what they show they like in their pattern offers as a measure of trustworthiness. An advantage of this technique is that it does not require the ground truth information for a participant’s real preferences to evaluate their trustworthiness.

Gained Utility Each participant’s outcome utility after agreement is used as the second behavioral feature. It is calculated as the weighted summation over an individual’s assigned level of issues. (Formula 1)

C. Statistical Analyses

An independent-samples t-test was performed to explore the most predictive features. Results are presented in the following.

Among the LIWC categories, High Machs talked less about money (\( M = 0.54, SD = 0.42 \)) compared to Low Machs (\( M = 0.70, SD = 0.6 \)), \( t(90) = -1.973, p = 0.05 \). High Machs used less future tense (will, gonna) (\( M = 0.81, SD = 0.42 \)) compared to Low Machs (\( M = 0.98, SD = 0.52 \)), \( t(90) = -2.387, p = 0.02 \). High Machs used somewhat less third person plural (they, their, they’d) (\( M = 0.15, SD = 0.24 \)) than Low Machs (\( M = 0.23, SD = 0.32 \)), \( t(90) = -1.848, p = 0.66 \). High Machs also used less offers compared to their partner (\( M = 0.44, SD = 0.28 \)) vs Low Machs (\( M = 0.55, SD = 0.28 \)), \( t(90) = -2.38, p = 0.018 \).
Examining the dialog acts, High Machs asked significantly more questions (M = 0.08, SD = 0.06) than Low Machs (M = 0.07, SD = 0.05), t(90) = 1.94, p = 0.053.

For visual features, High Machs showed a higher facial expressivity (M = 1.65, SD = 0.35) than Low Machs (M = 1.53, SD = 0.34), t(90) = 2.20, p = 0.029. Fear intensity was also larger for High Machs (M = 0.017, SD = 0.03) compared to Low Machs (M = 0.00, SD = 0.01), t(90) = 1.95, p = 0.052. One acoustic feature that was marginally significant between High and Low Machs was pitch (f0)'s standard deviation. High Machs: (M = 24.87, SD = 10.89), Low Machs: (M = 22.20, SD = 9.58), t(90) = 1.74, p = 0.08).

Among behavioral features, misrepresentation was higher for High Machs (M = 0.15, SD = 0.09) compared to Low Machs (Mean = 0.12, SD = 0.08), t(90) = 2.27, p = 0.024 and High Machs gained a higher outcome utility (M = 62.63, SD = 12.90) than Low Machs (M = 57.16, SD = 15.73), t(90) = 2.55, p = 0.012). Six of the statistically significant features are shown in Figure 2.

D. Prediction Models

The 6 classification models that are used for automatic prediction of High vs. Low Machiavellians are described in this subsection.

**Chance Baseline** Since median split of all Mach scores was used to divide the population into High vs. Low Machs, assigning everyone to one group resulted in 50% accuracy in classification.

**Unimodal classifiers** (Textual, Visual, Acoustic, or Behavioral) We used a Support Vector machine (SVM) classifier for each of the modalities to evaluate their performance individually.

**Textual + Visual + Acoustic + Behavioral** To fuse information from different modalities, we used an early fusion scheme by stacking the features from all modalities as an input to a new SVM classifier.

E. Methodology

We used one-person-leave-out test scheme for all of our classifiers to assure generalization across participants. Using LibSVM [22] package we trained and validated each classifier by performing 5 fold cross validation to find the optimal C for the third order polynomial kernel SVM on the training set.

Due to the large number of textual features, we used an automatic feature selection to help with interpretation and performance. Our original set of textual features were 126 features, and using t-test we selected the features with a p-value threshold of 0.1, where the two populations correspond to High vs Low Machs. This feature selection resulted in 20 textual features along with the 10 visual, 14 acoustic, and 2 behavioral features providing us with a total of 46 features for the multimodal classifier.

The classification results are shown in Figure 3. Accuracy presents the percent of individuals correctly classified as High vs. Low Machs performed by each classifier. Chance baseline shows 50% accuracy, and textual, visual, acoustic, and behavioral unimodal classifiers resulted in 56%, 56%, 54%, and 56% accuracy respectively. The model stacking all the features from the four modalities, resulted in the highest accuracy (71%) and was significantly better than any of the modalities alone (p < 0.001).

Fig. 3. Accuracies for chance, unimodal and multimodal classifiers
used significantly less often. While previous researchers interpreted decreased use of the word "we" as evidence that Machiavellians less often referenced others, we believe that less frequent use of the word "they" could be an even better indicator for Machiavellians not directly referring to others. In terms of visual features, facial expressivity was higher for Machiavellians. We interpret this as consistent with the literature that Machiavellians use emotion displays to gain advantage, such as feigning emotion for strategic gain. Such emotional displays may allow the machine annotation to capture more facial expressivity.

Lastly, in terms of the negotiation outcomes, we observed High Machs winning more lottery values. This again is consistent with prior studies reporting Machiavellians outperforming their opponents in most short term interactions [6].

In this study, we attempted to automatically recognize Machiavellianism, an anti-social trait that is often neglected by the personality computing research community. We used negotiations as a competitive situation that can provide Machiavellians with the opportunity to behaviorally diverge from their opponents. As the first multi-modal attempt toward predicting Machiavellianism, we used textual, visual, acoustic and behavioral features and reached 71% accuracy, which we believe is a good accuracy when compared to the other personality recognition models. We then provided several pieces of evidence that our most predictive features are consistent with the known Machiavellian related traits in social psychology. Thus, we believe the learned models could be generalized to other domains.

The focus of this work was on the recognition of Machiavellian behavior. In our future work, we would like to extend our study to the perception and synthesis of this personality trait. Specifically, we are interested in studying how Machiavellians are perceived by other people, and compare it to the results from this study.

ACKNOWLEDGMENT

This work was supported by the Air Force Office of Scientific Research under grant FA9550-06-1-0206 and the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES


