Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)
Journal Track

Predicting Strategic Behavior from Free Text (Extended Abstract)*

Omer Ben-Porat[1†], **Lital Kuchy**[1], **Sharon Hirsch**[1], **Guy Elad**[2], **Roi Reichart**[1] and **Moshe Tennenholtz**[1]

[1]Faculty of Industrial Engineering and Management, Technion- Israel Institute of Technology, Israel
[2]Faculty of Computer Science, Technion- Israel Institute of Technology, Israel
{omerbp, slitalku, sharonhi, sguyelad}@campus.technion.ac.il, {roiri,moshet}@ie.technion.ac.il

Abstract

The connection between messaging and action is fundamental both to web applications, such as web search and sentiment analysis, and to economics. However, while prominent online applications exploit messaging in natural (human) language in order to predict non-strategic action selection, the economics literature focuses on the connection between structured stylized messaging to strategic decisions in games and multi-agent encounters. This paper aims to connect these two strands of research, which we consider highly timely and important due to the vast online textual communication on the web. Particularly, we introduce the following question: can free text expressed in natural language serve for the prediction of action selection in an economic context, modeled as a game? We initiate research on this question by providing preliminary positive results.

1 Introduction

Much online activity is centered around text written by people. People send messages through a wide variety of communication media including email, Whatsapp, SMS, blogs and Facebook pages. In many cases these texts are connected to future actions, e.g. a request for some course of action or activity scheduling, among other alternatives. Moreover, web-search is probably the most successful tool which has emerged in the Web. In this context one aims to predict an action of an individual, i.e. whether he will click on a particular page, from a text this individual provides. Web-search is far from being the only example of a powerful text-based action prediction tool. Another such prominent application is sentiment analysis [Pang *et al.*, 2002; Pang and Lee, 2008], that provides a valuable signal for the prediction of an individual's choice among alternatives (see, e.g., [Ghose and Ipeirotis, 2010; Ravi and Ravi, 2015]). This signal is particularly useful when extracting opinions from textual reviews, that are abundant across the Web and have a strong impact on purchase decisions of wide crowds.

The connection between messaging and intention signaling to action is central also in economics. Indeed, the 2001 Nobel prize in economics was presented to Akerlof, Spence, and Stiglitz, for their pioneering lines of research, showing how signaling of information can alter strategic interactions (see [Spence, 1973]). Of particular interest is the study of cheap talk [Crawford and Sobel, 1982], where messaging may include arbitrary communication about private information which is carried out prior to the action that determines the parties' payoff – that is, the economic outcome is independent of the messages transmitted prior to action selection. An interesting variant of this setting is pre-play communication [Rabin, 1994], where economic agents communicate before playing in a non-cooperative game. Typically, such communication refers to private information where the agents might lie, or to a declaration of an intention to choose an action. Importantly, it is only the game itself and not the pre-game messaging that affect the outcome. Notice that studies of cheap talk and pre-play communication typically use a structured language: participants communicate through announcements on planned behavior or available information, which are elements of the game, rather than through arbitrary free text.

Given the above, it is evident that the connection between messaging and action is fundamental to both central applications in the web as well as to economics. The major differences are obvious as well: While the major online applications focus on the study of the connection between messaging in natural language to (non-strategic) action selection, the economics literature focuses on the connection between structured stylized messaging to strategic decisions in games and multi-agent encounters. It is therefore natural to ask: can free text expressed in natural language serve for the prediction of an action to be selected in an economic context, modeled as a game? Indeed, the main aim of this paper is to initiate such a study. Moreover, text in online media need not necessarily be expressed in the context of a given game or action selection context. Therefore, it may be even more intriguing to understand whether free text provided by individuals who are unaware of the games to be played, may induce a useful signal for the prediction of action selection in a game.

In service of the above, we introduce an experimental setup consisting of two steps. In the first step, individuals were

*This paper is an extended abstract of an article in the Journal of Artificial Intelligence Research [Ben-Porat *et al.*, 2020].

†Contact Author.

	Speed	Stop
Speed	$(0,0)$	$(14,2)$
Stop	$(2,14)$	$(6,6)$

(a)

	Left box	Right box
Left box	$(8,8)$	$(16,12)$
Right box	$(12,16)$	$(6,6)$

(b)

	Door A	Door B	Door C
Door A	$(10,10)$	$(0,0)$	(0,0)
Door B	$(0,0)$	$(10,10)$	(0,0)
Door C	$(0,0)$	$(0,0)$	(8,8)

(c)

Figure 1: Normal form representation of our games: (a) *Chicken*, (b) *Box*, and (c) *Door*

asked to provide free text with some personal story. They were told there would be a second step, but did not know what would be its structure. In the second step, they were matched to play three one-shot games (with no communication). These games were classical one-shot games, namely a pure coordination game, a congestion game, and a classical non-cooperative game (*Door, Box* and *Chicken*, correspondingly, see Figure 1). The aim was to study whether one can predict the individuals' actions in a particular game based on their provided texts.

The approach we have taken to tackle the above challenge is as follows. We created a description of the texts through commonsensical personality attributes. In order to make a sound treatment, we asked a group of students to reach a consensus on a set of attributes, and used crowd-sourcing in order to annotate the texts according to these attributes. Given these, we employed an ML technique known as transductive learning [Gammerman *et al.*, 1998]. The characteristic property of transductive learning is that at training time, the algorithm is exposed to all examples (i.e., all written texts), but to the labels (i.e., actions selected) of only a subset of these examples. The goal of the algorithm is then to infer the unknown labels from the known ones. An elegant property of our particular method is that we point to one structure that has predictive power with respect to different games. This is achieved by clustering over the input features (text-based personality attributes) while ignoring the labels. Notice that as choice prediction based on text is typically employed in settings where we already have the population texts (e.g., people share info in blogs, email, etc.) before we need to predict their actions, transductive learning is a natural setup. It is not that a new individual is born when we are called to make a prediction – we can use the texts he/she wrote in the past.

Our aim was to test whether we can outperform in our predictions a standard majority benchmark, where the action predicted by an individual in the test set is taken as the most popular action of individuals in the entire data set. Our results are encouraging and show that indeed an individual action prediction in strategic situations may be performed based on free text he/she provided. Moreover, in ablation analysis, we demonstrate the contribution of our specific modeling choices: employing a clustering algorithm and representing the text with personality attributes of their authors. Needless to say that this study is mainly a call for action – large scale experiments will be needed to test the significance of our findings.

1.1 Related Work

Machine learning (ML) has been applied in the past to predict human actions, both based on previous actions and from text. This previous work, however, is substantially different from ours. In this section, we provide a short discussion on those differences and crystallize our novel contributions. For an extensive literature review, the reader is referred to the full paper [Ben-Porat *et al.*, 2020].

Machine Learning for action prediction in strategic-form games is becoming increasingly popular. While our setup refers to action prediction in a single one-shot strategic form game, previous work successfully employed ML techniques in service of action prediction in an ensemble of games. Such techniques heavily base their choice prediction on having access to the way the agents chose their actions in other games. Particularly, they fall into one of two settings. Works that address the first setting (see, e.g., [Altman *et al.*, 2006] and the references therein), try to predict the behavior of individuals. They encode every individual by her play in several labeled games, and the predicted variable is the behavior of that individual in a new, unseen game. Works that address the second setting (e.g., [Hartford *et al.*, 2016; Plonsky *et al.*, 2017]), do not care about individual predictions. Every "point" is a choice problem, e.g., a selection between two lotteries that are encoded by probabilities and rewards, and its label is the population statistics. In particular, these works aim to predict the statistics (e.g., mean, variance, etc.) of the predictions people make on this data point, and this is different from predicting the behavior of one individual. In contrast, in the setting we address in this paper, we aim to predict the behavior of each individual in a new game, but we do not learn from previous plays. In order to address this challenging task, we exploit texts written by individuals and attempt to map texts to actions in a given game.

In the realm of Natural Language Processing (NLP), two strands of the literature on text-based prediction are most related to our efforts. Firstly, several works aimed to draw predictions about future actions of the authors of given texts, in cases where the texts are directly related to the actions (see, e.g., [Niculae *et al.*, 2015]). Secondly, some works tried to infer properties of an author's character [Bamman *et al.*, 2013; Gill *et al.*, 2009; Golbeck *et al.*, 2011] or of his/her emotional state [Eichstaedt *et al.*, 2018] from sources of text such as social media posts, blogs and tweets, as well as from descriptive text such as movie plot summaries. In this paper, we take a step forward, aiming to draw text-based predictions about actions that are not discussed or even implied in the text – the author is even unaware of the game he/she is going to play after writing the text. Instead, we ask our participants to write a personal text about a previous meaningful experience they are willing to share, and try to predict their actions in unrelated strategic situations.

Finally, our work has several interesting related lines of

work carried out in the multi-agent systems community, dealing with argumentation [Walton, 2009] and negotiation [Kraus and Arkin, 2001].

2 Task and Data

We now formally describe our choice prediction setup. Let $\mathcal{I}$ be a set of individuals, and let $\mathcal{Y}$ be a set of action choices (which serve as our labels). The set $\mathcal{Y}$ is composed of all possible choices an individual can make in a given situation. While typical day-to-day decisions correspond to dilemmas such as whether to buy a product or not or which academic institute to apply for, in this paper $\mathcal{Y}$ serves as the action space of each individual $z \in \mathcal{I}$ in a normal form, two-player one-shot game that is played between z and another individual, where each one is oblivious to the identity of the other.

To predict the actions of individuals, we represent individuals with their personality attributes, where the attribute space is denoted by $\mathcal{X}$. More concretely, we consider a representation function $\mathcal{R} : \mathcal{I} \rightarrow \mathcal{X}$, such that an individual $z \in \mathcal{I}$ is represented by $\mathcal{R}(z) \in \mathcal{X}$. Since individuals are represented according to $\mathcal{R}$, we shall make the simplifying assumption that each individual z is determined by his/her representation, $\mathcal{R}(z)$. We address two action prediction setups: inductive and a transductive [Gammerman *et al.*, 1998; Joachims, 1999; Joachims, 2003], although our focus is on the latter.

We focus on three (non-cooperative) symmetric two-player games, *Chicken, Box* and *Door* (see Figure 1). Importantly, we focus on these games not because individuals encounter them in their everyday activity, but since the dilemmas they capture correspond to real-life situations. In addition, these games reflect a diverse portfolio of dilemmas (see the full paper for game-theoretic justification).

2.1 Data

We published an invitation to participate in the experiment on the Facebook page of our university's students, a vibrant page that is used extensively. The reward was declared to consist of a compensation for one hour of work plus an additional performance-based bonus.

The experiment was composed of three parts. In the first part, each participant was requested to provide personal information. In the second part, participants were requested to write an English text of at least 1000 characters on a personal topic of their choice. In the third part, participants were presented with the three non-cooperative games described in Figure 1, and were told that they are playing with/against another randomly selected participant. The experiment was online for a time period of one week. We had 280 participants who completed the experiment form. Of these, 9 participants were filtered out, since the text they provided was a copy of a text found on the web. The total number of participants is hence 271. The texts we received are of high-quality and address topics such as dreams, trips, complaints, among others. The game statistics are summarized in Figure 2. [1]

[1]Code and data are available at: https://github.com/omerbp/Predicting-NLPGT.

3 Our Approach

Our approach is based on two-steps: In the first step, we represent each text with the personal attributes of its author. In the second step, we apply a transductive classifier to these examples so that we can predict human choices in games from the choices made by other humans with similar attributes. We now elaborate on the two steps.

In the first step, we asked two independent graduate students to read the texts and construct a set of personality attributes. They then merged the two sets into one final set, merging duplications and similar attributes. This protocol reflects our commonsense approach to personality attribute extraction, and when taking such an approach we could naturally not rely on existing theory-based automatic NLP tools for this extraction task. Due to the relatively small number of samples we could not develop a data-driven attribute classifier that would predict our attributes from the texts, and instead we turned to human judgments through a crowd-sourcing platform. We used Figure Eight (previously known as Crowd Flower) to extract eight estimates for each text-attribute pair, taking into account the necessary differences between human judgments when it comes to personality questions. The vector representation of each text was set to the vector whose coordinates correspond to the attributes, and the value in each entry is the average of the scores collected for this text-attribute pair.

In the second step, we aim to predict the choices made by our participants in a transductive learning setup. We propose a transductive classifier based on a deterministic clustering algorithm. The clustering algorithm is applied to the attribute representation of our participants, and as we are addressing the transductive setup, all the participants are clustered together. Our clustering algorithm does not have access to the choices made by any of our participants, only to their attributes. Hence, we can make predictions with respect to the various games using a single output (clustering) of our algorithm.

We cluster the example set $\mathcal{X}$ with a bottom-up agglomerative clustering algorithm using the Ward's minimum variance criterion for cluster merging. The output of this process is $n = 271$ possible clusterings - from the one at the top of the hierarchical tree, consisting of a single cluster, downward in the tree till the one that consists of $n = 271$ clusters. We refer to our algorithm as *TAC*, which stands for Transductive Attribute Clustering.

4 Experiments

Baselines. To measure the quality of TAC, we compare its performance to baselines. Our main baseline is the strong *Majority Vote Classifier (MVC)* that assigns to each example (participant) the majority label in the data. Notice that while this is a strong baseline in terms of its overall performance, it fails to predict all classes except from the majority class. Hence, a comparison to this baseline requires a careful selection of evaluation measures. We get back to this point below.

To put the results of both TAC and MVC in context, we also compute the expected scores of two stochastic classifiers. The *Expected Random Guess (ERG)* score is the expected

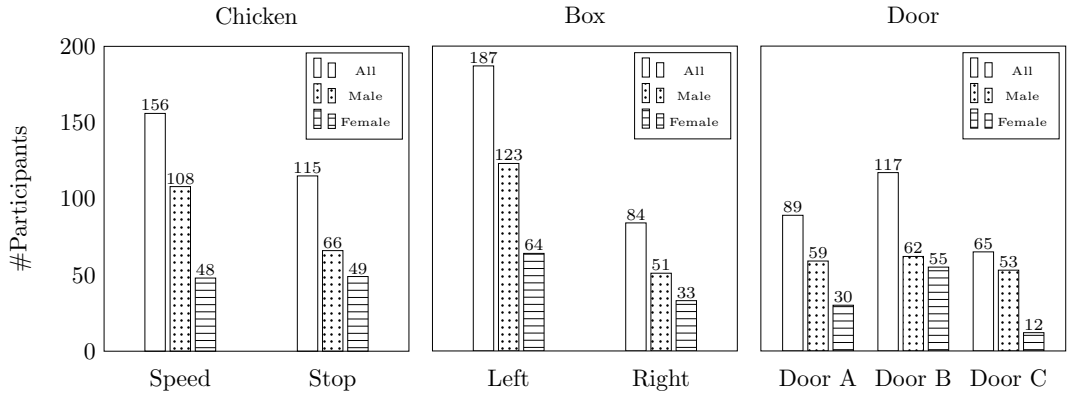

Figure 2: Game play histograms. In *Door*, for instance, 117 participants chose door B, 62 of which are males and 55 are females.

score of a stochastic classifier that assigns every participant with each of the labels in $\mathcal{Y}$ with probability $\frac{1}{|\mathcal{Y}|}$. The *Expected Weighted Guess (EWG)* score is the expected score of a stochastic classifier that assigns labels proportionally to the appearances of that label in the entire example set.

Ablation Analysis. We further perform an extensive ablation analysis to evaluate the importance of each of the components of TAC. The ablation analysis is focused on the importance of the clustering step and the importance of text representation according to the personality attributes of the authors. To evaluate the importance of the clustering algorithm, we compare to two transductive learning algorithms that do not perform clustering: transductive SVM ([Joachims, 1999], *TransSVM*) with a linear kernel, and a K nearest neighbor (*KNN*) classifier. Notice that the transductive SVM is trained for each task separately and in this respect, it is weaker than TAC that can make predictions with respect to multiple prediction tasks based on its single clustering.

To evaluate the importance of our character attribute representation, we compare to several models in which we run a clustering algorithm identical to TAC, but with a different feature representation. We employ two automatic natural language processing tools: The IBM Personality Insights service[2] and Linguistic Inquiry and Word Count (LIWC) [Pennebaker *et al.*, 2001; Tausczik and Pennebaker, 2010]. For each of these tools, we consider two models: One that uses the attributes provided by the tool solely, and another that incorporates both the tool's attributes and our 24 attributes. The last baseline is a tf-idf bag-of-words vector of each text.

Evaluation Measures. Our data demonstrates a class imbalance phenomenon, where in each game one of the actions is observed substantially more frequently than the other(s). We hence consider three complementary evaluation measures, Accuracy, Macro Average F1-score, and Macro Weighted Average F1. These measures can help us tell the full story about our model, the baselines, and the expected values of the random classifiers.

Classifiers Evaluation. To evaluate these algorithms, we perform for each game the following procedure 5000 times: we randomly sample train and test sets such that the training set is comprised of 90% of the data. Then, for each of the unlabeled examples, we predict a label.

[2]https://personality-insights-demo.ng.bluemix.net/.

5 Results

The full paper [Ben-Porat *et al.*, 2020] contains a comprehensive battery of results, along with extensive interpretation from various angles. Overall, these results justify our clustering and representation choices. Particularly, in the vast majority of cases our clustering algorithm is utilized by the winning configuration. Our attribute set is utilized by the winning configurations of two of the three games (*Chicken* and *Box*), and by the second-best configuration of the third game (*Door*). Given the many years and thousands of studies where the IBM and LIWC attributes have been in use, our novel commonsensical approach to personality attribute generation and annotation performs surprisingly well.

6 Conclusions

Our work initiates a study of action prediction in a single one-shot game. In contrast to previous studies, we do not have access to agents' behavior or to population statistics in other games. Instead, we have access to texts provided by the agents (the participants in our experiments), and we exploit these texts and other agents' behavior in the targeted game in service of action prediction. Our work opens a wide spectrum of future work. For example, an important next step would be to generate larger data sets in service of our task and to validate our approach further. On a more conceptual level, an interesting follow-up research would be to try and predict an opponent's action in a game when they play after seeing the agent's text.

Acknowledgments

The authors wish to thank the members of the IE@Technion NLP group for their valuable feedback and advice. The work of O. Ben-Porat is partially funded by a PhD fellowship from JPMorgan Chase & Co. The work of M. Tennenholtz is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 740435).

References

[Altman *et al.*, 2006] Alon Altman, Avivit Bercovici-Boden, and Moshe Tennenholtz. Learning in one-shot strategic form games. In *European Conference on Machine Learning (ECML)*, 2006.

[Bamman *et al.*, 2013] David Bamman, Brendan O'Connor, and Noah A Smith. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2013.

[Ben-Porat *et al.*, 2020] Omer Ben-Porat, Sharon Hirsch, Lital Kuchy, Guy Elad, Roi Reichart, and Moshe Tennenholtz. Predicting human behavior from free text. *Journal of Artificial Intelligence Research*, 2020.

[Crawford and Sobel, 1982] VP. Crawford and J. Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.

[Eichstaedt *et al.*, 2018] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208, 2018.

[Gammerman *et al.*, 1998] Alexander Gammerman, Volodya Vovk, and Vladimir Vapnik. Learning by transduction. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 1998.

[Ghose and Ipeirotis, 2010] Anindya Ghose and Panagiotis G Ipeirotis. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512, 2010.

[Gill *et al.*, 2009] Alastair J Gill, Scott Nowson, and Jon Oberlander. What are they blogging about? personality, topic and motivation in blogs. In *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[Golbeck *et al.*, 2011] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 2011.

[Hartford *et al.*, 2016] Jason S. Hartford, James R. Wright, and Kevin Leyton-Brown. Deep learning for predicting human strategic behavior. In *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[Joachims, 1999] Thorsten Joachims. Transductive inference for text classification using support vector machines. In *Proceedings of the Sixteenth International Conference on Machine Learning (ICML)*, 1999.

[Joachims, 2003] Thorsten Joachims. Transductive learning via spectral graph partitioning. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, 2003.

[Kraus and Arkin, 2001] Sarit Kraus and Ronald C Arkin. *Strategic negotiation in multiagent environments*. MIT press, 2001.

[Niculae *et al.*, 2015] Vlad Niculae, Srijan Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. Linguistic harbingers of betrayal: A case study on an online strategy game. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL) and the 7th International Joint Conference on Natural Language Processing (IJNLP)*, 2015.

[Pang and Lee, 2008] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, January 2008.

[Pang *et al.*, 2002] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.

[Pennebaker *et al.*, 2001] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

[Plonsky *et al.*, 2017] Ori Plonsky, Ido Erev, Tamir Hazan, and Moshe Tennenholtz. Psychological forest: Predicting human behavior. In *Proceedings of the Thirty-First Conference on Artificial Intelligence (AAAI)*, 2017.

[Rabin, 1994] Matthew Rabin. A model of pre-game communication. *Journal of Economic Theory*, 63(2):370–391, 1994.

[Ravi and Ravi, 2015] Kumar Ravi and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 89:14–46, 2015.

[Spence, 1973] A. M. Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

[Tausczik and Pennebaker, 2010] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.

[Walton, 2009] Douglas Walton. Argumentation theory: A very short introduction. In *Argumentation in Artificial Intelligence*, pages 1–22. Springer, 2009.