DARTMOUTH

Institute for Security Technology and Society at Dartmouth

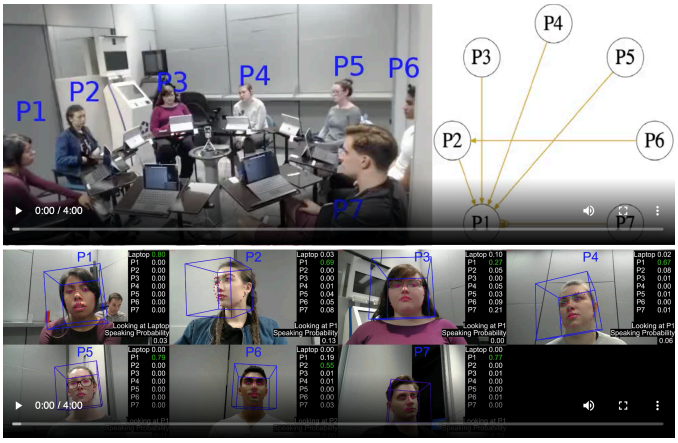
Human Behavior Prediction using Non-Verbal Communication

Nowadays, a large amount of videos such as meetings, discussions, public speech, family activities are recorded and available online. In such videos, the non-verbal communication among people reflects human behaviors and relationships such as dominance, deception, like/dislike, and deception. On one hand, social scientists show great interest in discovering the factors (e.g. visual and vocal) that are related to such social behaviors. On the other hand, automated analysis and prediction are becoming more effective with the development of machine learning (ML) and deep learning (DL) technologies. We take advantages of the insights from social scientists to capture both individual (e.g. voice, expression) and interaction cues (e.g. look at) from videos, and build end-to-end systems to predict deception, dominance, nervousness etc. Promising prediction accuracies are achieved for various tasks.

Human behaviors such as deception and dominance are important social factors. We incorporate multiple modalities (visual, audio, linguistic and group communication) for automatic prediction.

Building Interaction Networks

We propose a novel method for building dynamic face-to-face interaction networks in videos of groups of people. The networks identify various means of non-verbal communications that play a significant role in predicting social behaviors such as deception and dominance. The technique is called ICAF (Iterative Collective Attention Focus), a collective classification model to jointly learn the visual focus of attention groups of people (i.e. who looks at whom). ICAF models the people collectively—the predictions of all other people's classifiers are used as inputs to each person's classifier. This explicitly incorporates interdependencies between all people's behaviors. It also takes the history of visual focus of attentions into account. ICAF outperforms the strongest baseline by 1%–5% accuracy across two datasets. We further propose a light supervised ICAF to create who-looks-at-whom, who-listens-to-whom, and who-speaks-to-whom networks from unseen group interaction videos.



Demo videos showing the predicted probabilities of people looking at each other (bottom) and the dynamic social interaction networks built upon the predictions (upper right).

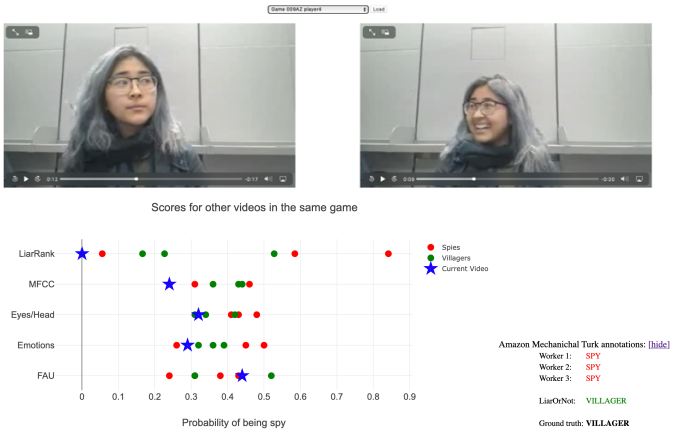
Deception Detection

Three methods are proposed for deception detection. The first focuses on individual short-term deception, utilizing multimodal features (video IDT features and micro-expressions, audio MFCC features, Glove word embeddings). Machine learning models are built for each kind of features and an ensemble of these achieves an AUC score of 0.92 on a 1-minute level courtroom video dataset.

The second method exploits deception detection in long-term group-interaction videos (~30 minutes). We aim to detect the liars in videos of the game Resistance (similar to Mafia). Built upon any kind of feature f (e.g. expression, facial action unit), we propose a class of meta-features called LiarRank, which ranks the value of f of the person of interest against each group in training data. LiarRank

combined with other multimodal features achieves an AUC score of 0.7 in a fully automated setting.

The Liar-Or-Not deception detection system achieves over 0.7 AUC in group-interaction videos while annotators only score a 0.59 AUC. We find that deceivers have a lower entropy and reciprocity of looking than non-deceivers, they interact more with non-deceivers and ignore



Demo showing our deception detection system with predictions made by five kinds of features and the comparison with predictions by humans.

The third mehod explores the interaction network effects to analyze deception behaviors in a data-driven way. Our analysis reveals that less engaged deceivers are identified early. Moreover, pairs of deceivers tend to avoid mutual interaction and focus their attention on non-deceivers. In contrast, non-deceivers interact with everyone equally. We propose the notion of Negative Interaction Networks (NINs) and create a belief propagation neural net algorithm called BPNN based on dynamic FFINs and NINs to detect deceivers using only 1-minute videos. Our method outperforms recent state-of-the-art computer vision, graph embedding, and ensemble methods by at least 20% AUC in identifying deception from videos.

Dominance Prediction

For dominance prediction, we consider the problems of predicting (i) the most dominant person in a group of people, and (ii) the more dominant of a pair of people. We introduce a novel family of features called Dominance Rank from the interaction networks, indicating the relative dominance of people in a group. These features align with social science studies (e.g. visual dominance ratio). We combine features not previously used for dominance prediction (e.g., facial action units, emotions), with a novel ensemble-based approach to solve these two problems. We test our models against four competing algorithms in the literature on two datasets. Our models achieve a 2.4%--16.7% improvement in AUC on one dataset, and a 0.6%--8.8% increase in accuracy on the other.

Additional Information

References

1. Z. Wu, B. Singh, L. S. Davis, V. S. Subrahmanian. Deception Detection in Videos. *AAAI Conference on Artificial Intelligence (AAAI 2018)*
2. C. Bai, M. Bolonkin, J. Burgoon, C. Chen, N. Dunbar, B. Singh, Z. Wu. Automatic Long-Term Deception Detection in Group Interaction Videos. *IEEE International Conference on Multimedia and Expo (ICME 2019)*
3. C. Bai, M. Bolonkin, S. Kumar, J. Leskovec, J. K. Burgoon, N. E. Dunbar, V. S. Subrahmanian. Predicting dominance in multi-person videos. *Intl. Joint Conference on Artificial Intelligence (IJCAI 2019)*
4. C. Bai, S. Kumar, J. Leskovec, M. Metzger, J. F. Nunamaker Jr., V. S. Subrahmanian. Predicting the Visual Focus of Attention in Multi-Person Discussion Videos. *Intl. Joint Conference on Artificial Intelligence (IJCAI 2019)*
5. S. Kumar, C. Bai, V.S. Subrahmanian, J. Leskovec. Deception Detection in Group Video Conversations using Dynamic Interaction Networks. *International Conference on Web and Social Media (AAAI ICWSM 2021)*

Demos

https://cs.dartmouth.edu/~mbolonkin/liar-or-not/demo/
https://cs.dartmouth.edu/~cy/icaf/
https://cs.dartmouth.edu/~cy/dom/

Presentation

https://www.cs.dartmouth.edu/~mbolonkin/scan/webinars/slides/july13webinar.pdf

PARTICIPANTS

Lead: V.S. Subrahmanian

Chongyang Bai, Maksim Bolonkin, Chao Chen, Larry Davis, Srijan Kumar, Jure Leskovec, Bharat Singh, Zhe Wu.

DARTMOUTH