High-level Features for Multimodal Deception Detection in Videos

Rodrigo Rill-García[1] Hugo Jair Escalante[1] Luis Villaseñor-Pineda[1]
Verónica Reyes-Meza[2]

[1]Instituto Nacional de Astrofísica, Óptica y Electrónica; Santa María Tonanzintla, Pue, 72840, Mexico; {rodrigo.rill, villasen, hugojair}@inaoep.mx
[2]Centro Tlaxcala de Biología de la Conducta; Universidad Autónoma de Tlaxcala, Mexico; veronica.reyesm@uatx.mx

Abstract

Deception (the action of deliberately cause someone to believe something that is not true) can have many different repercussions in daily life. However, deception detection is an inherently complex task for humans. Due to this, not only there is uncertainty on which features should be used as cues for automatic deception detection, but labeled data is scarce. In this paper, we explore typical features that can be extracted from videos for affective computing and study their performance for deception detection in videos. Additionally, we perform a study of different multimodal fusion methods meant to improve the results obtained by using the different sets of extracted features separately, including a novel set of methods based on boosting. For this study, high level features are extracted with open automatic tools for the visual, acoustical and textual modalities, respectively. Experiments are conducted using a real-life court trial dataset for deception detection, as well as a novel Mexican deception detection dataset using Spanish as the spoken language.

1 Introduction

Decision making is a process that requires the analysis of available data. However, an "optimal" decision can be harmful if such data is inaccurate -not to say strictly wrong. Spreading inaccurate or wrong information purposely is a way to mislead people's decisions for our own convenience. According to the Oxford dictionary, that is the action of deceiving someone: "deliberately cause (someone) to believe something that is not true, especially for personal gain". Job interviews, court trials, police investigations... there are many cases where believing in someone who is actually lying can imply severe consequences.

Although deception detection is a hard task for ordinary people, previous research [3, 4, 6, 5] supports a well-known assumption that a difference exists in the way liars communicate in contrast with truth tellers. Of particular interest, evidence suggests that such difference can be pointed out using machine learning.

Furthermore, there are many available sources of cues for deception: eye movements, facial expressions, voice, speech, etc. Recent research [2, 1, 7] points out that multimodal analysis of videos is useful to achieve better results in the deception detection task, rather than using different modalities independently such as visual cues, thermal images, voice analysis or text analysis.

Summarizing, this work aims to: 1) present a study on high-level (interpretable by humans) feature sets that can be automatically extracted from videos for the deception detection task (Fig. 1); 2) analyze the complementarity between such features to provide evidence of the benefits that could be obtained from fusing them (Fig. 2); 3) present a study on first attempts to perform such fusion by using methods inspired in classifier ensembles (Fig. 3); 4) perform a comparison for both single feature sets and fusions on two datasets with different language and contexts, including a novel Mexican database.

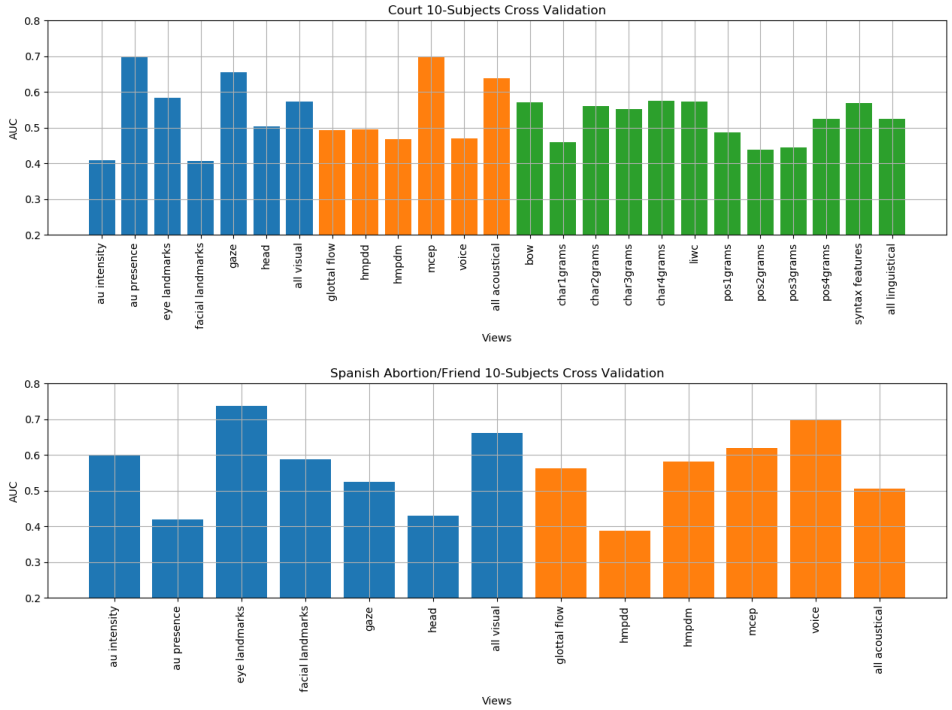

Figure 1: AUC achieved by the different views in the court-trial (top) and Spanish (bottom) datasets, including their concatenation (rightmost column).

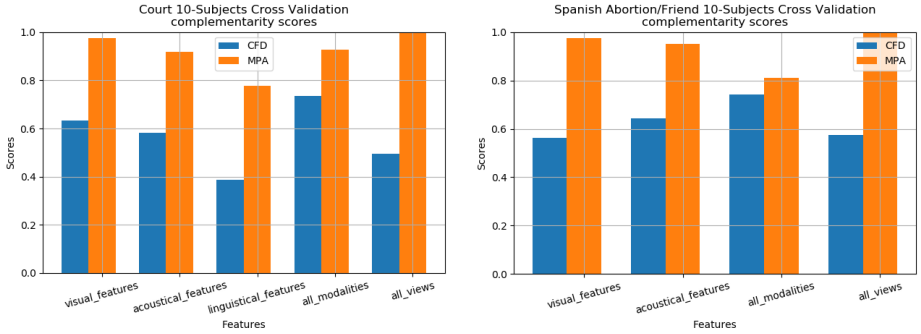

Figure 2: CFD and MPA between views and modalities from the court (left) and Spanish (right) datasets.

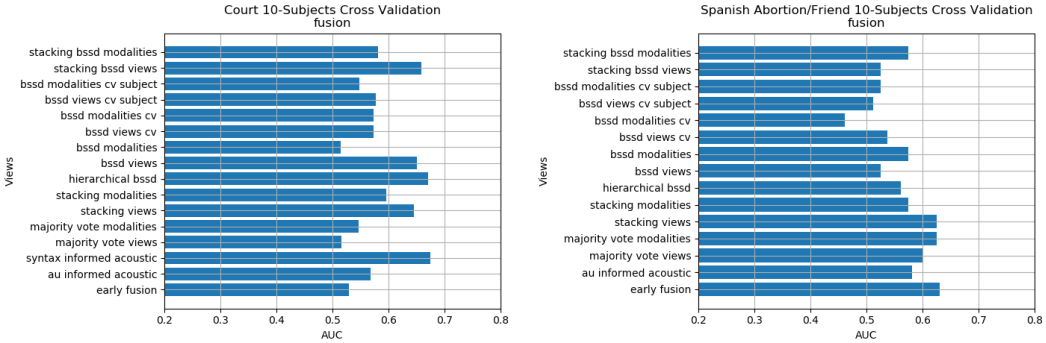

Figure 3: AUC achieved by different fusion methods in the court (left) and Spanish (right) datasets.

References: `https://www.dropbox.com/s/a6jbxj3uyf4mf8p/References_LatinX.pdf?dl=0`

References

[1] Mohamed Abouelenien, Verónica Pérez-Rosas, Rada Mihalcea, and Mihai Burzo. Detecting deceptive behavior via integration of discriminative features from multiple modalities. *IEEE Transactions on Information Forensics and Security*, 12(5):1042–1055, 2017.

[2] Mohamed Abouelenien, Verónica Pérez-Rosas, Bohan Zhao, Rada Mihalcea, and Mihai Burzo. Gender-based multimodal deception detection. In *Proceedings of the Symposium on Applied Computing*, pages 137–144. ACM, 2017.

[3] Paul Ekman. *Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition)*. WW Norton & Company, 2009.

[4] Matthew L Newman, James W Pennebaker, Diane S Berry, and Jane M Richards. Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, 29(5):665–675, 2003.

[5] Verónica Pérez-Rosas, Mohamed Abouelenien, Rada Mihalcea, and Mihai Burzo. Deception detection using real-life trial data. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 59–66. ACM, 2015.

[6] Verónica Pérez-Rosas and Rada Mihalcea. Cross-cultural deception detection. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 440–445, 2014.

[7] Zhe Wu, Bharat Singh, Larry S Davis, and VS Subrahmanian. Deception detection in videos. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.