Expert-Novice Interaction: Annotation and Analysis

Soumia Dermouche, Catherine Pelachaud
CNRS UMR 7222, ISIR, Sorbonne Universités
75005 Paris, France
{soumia.dermouche, catherine.pelachaud}@upmc.fr

Abstract

In this demonstration, we present the *NoXi* corpus of expert-novice interactions, our annotations and analysis. To analyze the data we apply HCApriori, a Temporal Sequence Mining algorithm to extract relevant behavior sequences for both expert and novice. NoXi provides over 25 hours of dyadic interactions recorded in different languages, mainly English, French, and German. The annotation tool, NOVA, developed by (Baur et al., 2015) allows annotating data using discrete and continuous schema. We use NOVA to manually annotate non-verbal behaviors (discrete annotation) and engagement levels (continuous annotation).

Keywords: Non-verbal behavior; Engagement; Sequence Mining; Virtual Agent

1. Introduction

This work is part of the H2020 project ARIA-VALUSPA (*Artificial Retrieval of Information Assistants - Virtual Agents with Linguistic Understanding, Social skills and Personalized Aspects*). In this project, a corpus of dyadic interactions, named NoXi, has been collected (Cafaro et al., 2017). NoXi is available to the research community from the website: `https://noxi.aria-agent.eu/`. During the interaction, participants exchanged through a large screen in different rooms. One participant assumes the role of an expert on a given topic and the other the role of a novice for this topic. NoXi is composed of 84 sessions recorded in three different countries France, Germany and UK and discussing 58 topics like video games, sports, cooking, etc. In the following sections, we describe our coding scheme for NoXi annotation and how we use the NOVA tool. We have manually annotated several non-verbal behaviors and engagement levels of both expert and novice. The use of sequence mining allowed us discovering relevant patterns for different engagement levels.

2. Annotation

NOVA[1] is an open-source annotation tool developed by (Baur et al., 2015) that we use to annotate the NoXi corpus. NOVA overcomes the limitations of existing annotation tools by exploring richer data like face streams or skeleton and by proposing two annotation schemas at time: discrete and continuous. Moreover, NOVA is a collaborative platform in which annotators from different sites can combine and share their annotations. Discrete annotation schema can be used to label behaviors that can be classified into a set of categories (e.g. gaze direction). Discrete annotation characterizes the starting and the ending time of behaviors. On the other hand, a continuous scale could be more appropriate for describing continuous dimensions, such as, engagement which is expressed all along the interaction. Figure 1 shows one session of NoXi viewed with NOVA. Audio-visual as well as skeleton and face streams of both expert and novice are opened. Using NOVA, continuous and discrete annotations can be visualized at same time.

In this work, we use NOVA to annotate the French part of NoXi database which is composed of 21 sessions. The total duration of all these sessions is 7 hours and 25 minutes. We use a discrete annotation schema to label six non-verbal behavior types: head direction and movement, smile, eyebrow movement, gesture and hand rest positions. Continuous scale is adapted for engagement annotation. In order to avoid content biases from the the verbal stream and prosody when annotating engagement, we have filtered it out, for both expert and novice. According to (Yannakakis et al., 2017) that suggest ordinal annotation for affect modeling, we annotate engagement over five levels: strongly disengaged, partially disengaged, neutral, partially engaged and strongly engaged. One evaluator was asked to rate and associate the engagement level of expert and novice over these levels. Table 1 illustrates the manual annotations that we realize so far. These annotations have been realized by three evaluators: one for engagement annotation, one for gesture annotation and the last one dealt with the remaining annotations. For each modality, we indicate label of annotated signals,the number of annotated sessions, their duration, and the number of annotations for expert and novice.

3. HCApriori Algorithm

Human behaviors are naturally multimodal. Human states, attitude, engagement level, etc, can be displayed through sequences of behaviors (Burgoon and Dunbar, 2006). In order to extract a meaningful multimodal sequences from NoXi, we rely on HCApriori, a temporal sequence mining algorithm (Dermouche and Pelachaud, 2016). This algorithm aims at finding frequent patterns (frequent subsequences) hidden in set of sequences. HCApriori takes as input: the sequence dataset, a minimum threshold (f_{min}), i.e. only patterns that hold within this threshold are considered as frequent, dissimilarity measure like CityBlock and dissimilarity threshold called ϵ.

HCApriori operates in two steps: (1) hierarchical clustering in which signals are grouped into the same cluster if and only if their temporal distance is less than ϵ. Temporal distance between two signals is evaluated using a dissimilarity measure such as CityBlock. At the end of this step, the

[1] `https://github.com/hcmlab/nova`

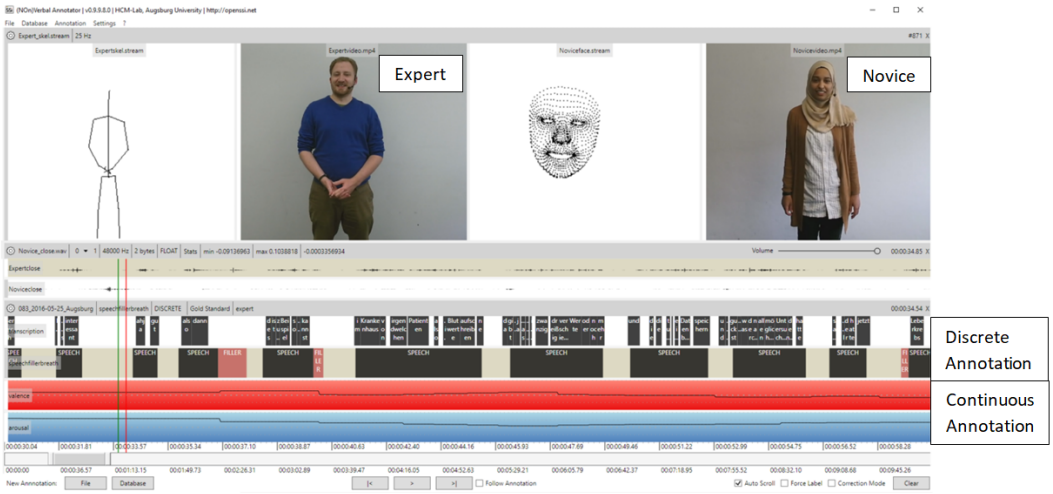

Figure 1: A screenshot of NOVA interface: videos of expert and novice, expert skeleton and novice's face tracking (top). Discrete and continuous annotations tracks are shown (bottom).

Table 1: Number of manual annotations of each non-verbal modality for expert and novice.

Modality	Label of annotated signals	Annotation number	
		Expert	Novice
Head direction and movements	Nod, Shake, Forward, Back, Up, Down, Side, Tilt	72	337
Smiles	Smile	153	157
Eyebrow movements	Frown and Raised	147	44
Gestures	Iconics, Metaphorics, Deictics, Beats, and Adaptors	1223	293
Hand rest positions	Arms crossed, Hands together, Hands in pockets, Hands behind back, and Akimbo	1317	612
Engagement	Strongly disengaged, Partially disengaged, Neutral, Partially engaged, Strongly engaged	1481	1679

cluster centroid represents a pattern of length one. (2) Taking as input the results of the previous stage, Apriori-like procedure is adapted to generate longer temporal patterns.

For NoXi analysis using HCApriori, we can, for example, explore the relationships between non-verbal behavior and engagement perception. For this purpose, we prepared the input dataset of HCApriori by collecting all sequences of non verbal behaviors that appear during a given engagement level. Table 2 presents the number of sequences we obtained for each engagement level for expert and for novice. Then, we have applied HCApriori to extract temporal patterns of nonverbal signals expressing the five engagement levels.

Our demo will consist of a presentation of the data collection, experimental setup of NoXi, as well as the annotation tool used for the manual annotation of various behaviors. It will also provide, based on HCApriori, the data analysis and the investigation of the sequential behavios of both expert and novice.

Table 2: Number of sequences of each engagement level for both expert and novice.

	Level 1	Level 2	Level 3	Level 4	Level 5	Total
Expert	48	373	373	561	126	1481
Novice	116	432	509	558	64	1679

4. Acknowledgements

Funded by European Union Horizon 2020 research and innovation programme, grant agreement No 645378.

5. Bibliographical References

Baur, T., Mehlmann, G., Damian, I., Lingenfelser, F., Wagner, J., Lugrin, B., André, E., and Gebhard, P. (2015). Context-Aware Automated Analysis and Annotation of Social Human–Agent Interactions. *ACM Transactions on Interactive Intelligent Systems*, 5(2):1–33.

Burgoon, J. K. and Dunbar, N. E. (2006). Nonverbal expressions of dominance and power in human relationships. *The SAGE Handbook of Nonverbal Communication*, (September 2014):279–298.

Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres, M. T., Pelachaud, C., Andr, E., and Valstar, M. (2017). The NoXi Database : Multimodal Recordings of Mediated Novice-Expert Interactions. In *ICMI'17,*, pages 350–359, Glasgow, Scotland. ACM.

Dermouche, S. and Pelachaud, C. (2016). Sequence-based multimodal behavior modeling for social agents. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*, pages 29–36, Tokyo, Japan. ACM.

Yannakakis, G. N., Cowie, R., and Busso, C. (2017). The Ordinal Nature of Emotions. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 248–255.