Knowledge graph retrieval and analysis for the evaluation of customer service in video

Julio Vizcarra[1]　西村悟史 [1]　福田賢一郎 [1]
Julio Vizcarra [1], Satoshi Nishimura[1], Ken Fukuda[1]

[1] 産業技術総合研究所
[1] National Institute of Advanced Industrial Science and Technology (AIST)

Abstract: In this work, we present a methodology that measures the customer service expertise and performance in video by applying several rules and metrics on knowledge graphs (KGs). In our approach, the KGs represent human behavior performed in the video through conversations and actions which are described from the knowledge base (KB). The definition of rules, baselines, and metrics are written by specific notations (Allen's for representing time interval relations and RCC8 for defining location relations).
The methodology is composed of four stages: in 1) "behavior pattern definition" the rules, baselines, and metrics/scores are defined for assessment of behaviors in customer service. The 2) "knowledge graph constructions" process video files, extracts, and represent human activities and interactions with objects in video. During the 3) "knowledge graph retrieval" the user behavior is retrieved from a knowledge graph by means of SPARQL queries. Finally, in 4) "knowledge graph analysis" the rules and scores are applied. In order to measure the expertise following certain rules, the methodology implements inferences, queries, filters, and temporal processing on the knowledge graphs. The purpose of this step is to measure expertise in customer service. Consecutively, the user performance in the video is compared with other baselines (user expert and average).
As a case study, the work was applied to elderly care customer service using public videos from the elderly behavior library.

1. Introduction

Analyzing behavior recorded in the video can become a crucial goal in areas such as customer service, marketing, training programs, human sciences, and so on. In the video analysis, there is a necessity to represent several aspects of the human activities in minuscules lapse times as small as video frames. Human behavior is by nature multimodal which can cover a wide number of human expressions (facial and corporal gestures, tone of voice, sight, posture, and so on). Tackling this research area, the main objective of current work is to retrieve and analyze human behavior represented through a knowledge graph (KG). The methodology proposed 1) defines formally the metrics that have to be applied on the KG, 2) create a knowledge graph that represents the human behavior recorded in the video. 3) retrieves the knowledge graph, and 4) measures the expertise of users in customer service. Additionally, a user interface was developed for the visualization and tracking of user activities validated through rules.

2. Methodology

In this section the methodology is introduced and divided into three stages: "behavior pattern definition", "knowledge graph construction"," "knowledge graph retrieval", and "knowledge graph analysis". At first, the work defined formally the task for a desirable behavior pattern in customer service. The next step is to represent human behavior on video through a knowledge graph. Consequently, the entities and relations are by means of queries. Finally, the rules defined are applied in the KG. As result, the users can be classified and compared with experts and averages users (Figure 1).

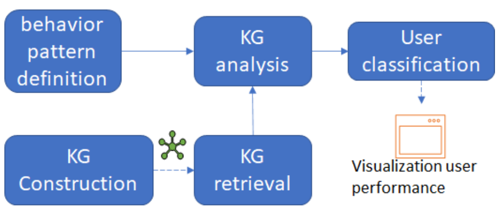

Figure. 1. methodology.

2.1 Behavior pattern definition

In this section, the definition and classification of rules are introduced. The rules are defined formally by using 1) Allen's algebra[1] that represents a time interval. the RCC8 calculus[2] serves for defining location and relations. The rules are classified into the domains: conceptual, temporal physical. The conceptual refers to terms that represent actions or concepts in the case study.

The temporal rules describe the relationship among events on time. Finally, the physical defines real measures or physical features. Table 1 presents some examples of rules.

Table. 1. Type of rules.

Rule name	Rule definition	domain	Description
Get Agreement 2	(同意を見る < 持つ ~ max=20sec & min=10sec)	Temporal	See agreement reaction before Grab during 10-20 sec
pull back left foot	(左膝の角度 <90度 : 座位)	Physical	Left knee degree less than 90.
Grab elbow	(持つ => 肘)	Function	Grab elbow
Pull	引っ張る	Conceptual	Pull

The user behavior is evaluated according to the rules, metrics (groups of rules), and scores (groups of metrics). (Table 2).

Table. 2. Classification of rules.

Score	metric	rule
actions allowance	Do	Get Agreement 2
actions allowance	Don't	Pull
actions allowance	Personalized Do	pull back left foot
actions allowance	Personalized Don't	Grab wrists

The rules and baselines are written in text files. The definitions are stored in the triple store[3] for being process during the KG analysis stage.

2.2 Knowledge graph construction

Once the definition of behavior patterns was performed through the writing of rules. The next step is to construct a knowledge graph that represents human behavior in the video. The process 1) starts processing video components a) conversations, b) human-objects features obtained by vide analysis, and c) annotations taken from a knowledge base (ontology). 2) During the KG construction hierarchical structures, custom-made rules and knowledge base extended definitions are taking into account. (Figure 2).

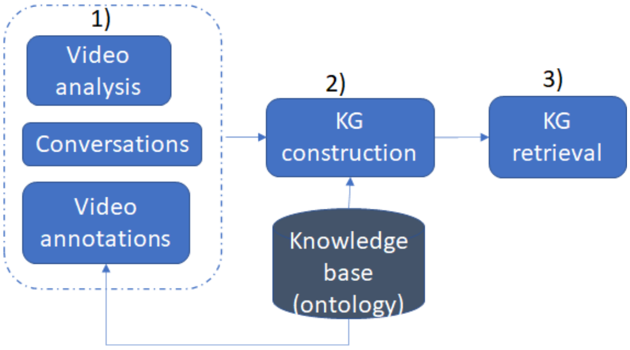

Figure. 2. KG construction.

As example for this paper, Figure 3 introduces an excerpt from the KG that corresponds to a conversation. The text processed is “山本さん、東京で育った街の思い出は何ですか?” (Mr. Yamamoto, What do you remember about the place you grew up in Tokyo). This comment was processed by NLP [4] focused on the Japanese language [5]. The type of entities person, person, and location can be distinguished. The action represented is “敬意を示す” (show respect).

Additionally, the content of the KG is introduced in Figure 4 [6]. The elements represented are physical objects (person-object) analyzed by video analysis. The main features are extracted and represented (upper side).

Moreover, video contents (elements in the conceptual domain) as actions, user objects names or events that occurred on video as lapse times are also included in the representation (lower side).

The elements pointed in figure 4 are the physical objects caregiver. chair, and their attributes such as “roll”. Similarly, the action represented “move” and their description is retrieved from the knowledge base.

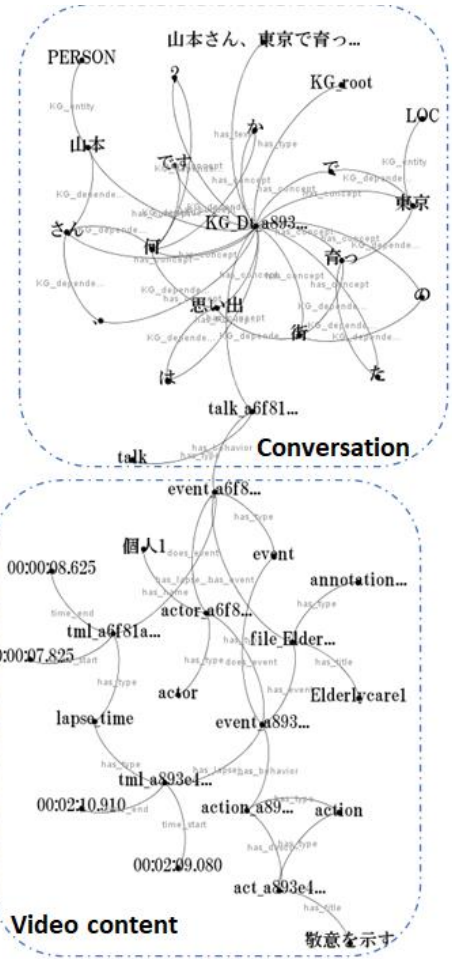

Figure 3. Excerpt KG of human activity.

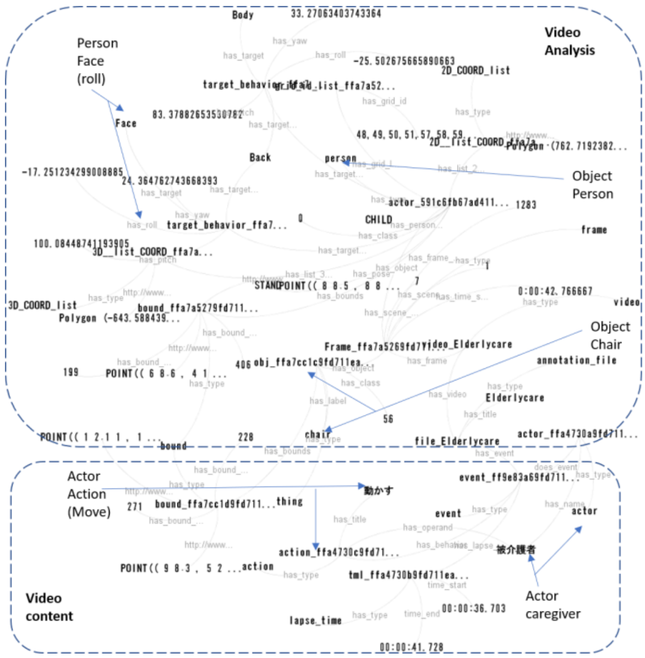

Figure 4. KG video analysis and content[6].

2.2 Knowledge graph retrieval

The next process is to retrieve the user behavior represented as a knowledge graph by SPARQL queries [4]. (Figure 5).

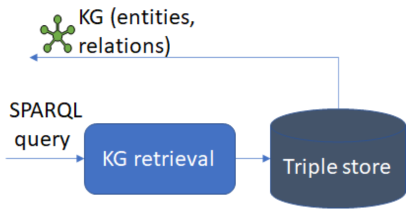

Figure. 5. Table user actions validated.

As an example of content retrieval, Table 3 presents a SPARQL that retrieves the users' body position (roll) for the frame 37.

Table. 3. SPARQL query.

Query
select ?fp as ?body_roll where {
graph <integrated_graphs> {
?tb <has_roll> ?fp.
?tb <has_target> <Body>.
?o <has_target_behavior> ?tb.
?o <has_type> <person>.
?f <has_object> ?o.
?f <has_frame_number> <100> .
} }

The result of the previous query is presented in Table 4. The body roll (angle in degrees) for a user is displayed.

Table. 4. SPARQL result.

body_roll
26.611355830735853

2.3 Knowledge graph analysis

In this stage the following processes are carried out: 1) the parsing of rules, 2) the retrieval of users' actions performed in the video, and represented as KG. 3) The validation of rules using the Boolean operators (AND OR NOT). 4) The calculation and measurement of behavior patterns in order to assign a numerical value to user performance. Finally, 5) the users are classified as an expert (green) or novice (red) according to the expertise in care service (Figure 6).

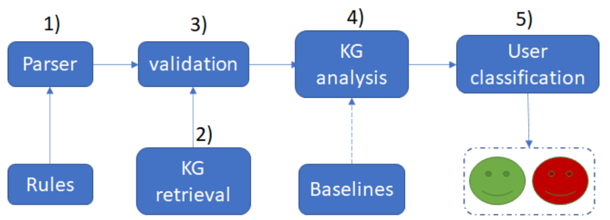

Figure. 6. KG analysis.

3. Implememtation

The system implementation and web visualization were developed using python. The system is able to load the rules and baselines as text files. The KG's are retrieved using queries on HTTP. As result, the users are classified according to their performance in elderly care services as experts and users. An example is presented in Figure 7.

Actor	file	Classification	Expertise value
介護者	Elderlycare	expert	0.04573074494949495
被介護者	Elderlycare	novice	0.005555555555555556
Actor	file	Classification	Expertise value

Figure. 7. Table user classification.

Additionally, user actions can be visualized on the web. For a user, the performed are validated are displayed. The combination of actions/events that followed certain rules is presented in red (false) and green (true). Moreover. The time in the video when the actions occurred is also presented (Figure 8).

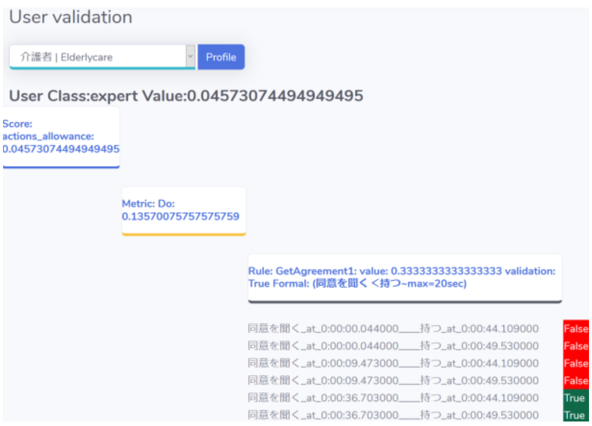

Figure. 8. Table user actions validated.

4. Conclusions

In this paper, a methodology that retrieves and analyses knowledge graphs from video content was introduced. The approach was focused on analyzing multi-modal human data. The definition of desirable behavior patterns was formally described by a domain expert and introduced to the system. Moreover, the user expertise in care service could be compared with user experts and average and thus classify them. The domains exploited in the knowledge graph were the conceptual, physical, temporal using a semantic approach. A graph exploitation tool was developed in order to verify the application of rules and its suitability with the case study. The results obtained presented a way to define behavior patterns on the semantic layer which are applied to knowledge graphs that represent human behavior.

Acknowledgments

The part of this paper is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] Allen, James F, "Towards a general theory of action and time" Artificial intelligence, Elsevier. vol.23, no.2, pp.123–154, 1984.

[2] Randell, David A and Cui, Zhan and Cohn, Anthony G. A spatial logic based on regions and connection. Journal KR. Vol.92. pp.165–176.1992.

[3] Quilitz, B., Leser, U.: Querying distributed rdf data sources with sparql. In: European semantic web conference. pp. 524–538. Springer (2008).

[4] Chowdhury, Gobinda G: Natural language processing. Annual review of information science and technology, Wiley Online Library, Vol. 37, No. 1, pp. 51-89, 2003.

[5] Hiroshi, Mai, and Masayuki: 短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習，言語処理学会 第 25 回 年 次 大 会 . http://www.anlp.jp/proceedings/annual_meeting/2019/pdf_dir/F2-3.pdf, 2019.

[6] Julio Vizcarra. Satoshi Nishimura, Ken Fukuda. Multi-modal representation of human activities via knowledge graph. The 22nd ACM International Conference on Multimodal Interaction (ICMI 2020). Paper under review.