Collaborative Ontology Development and its Use for Video Annotation in Elderly Care Domain

Satoshi NISHIMURA†, Member, Julio VIZCARRA†, Yuichi OOTA†, and Ken FUKUDA†, Nonmembers

SUMMARY Multimedia data and information management is an important task according to the development of media processing technology. Multimedia is a useful resource that people understand complex situations such as the elderly care domain. Appropriate annotation is beneficial in several tasks of information management, such as storing, retrieval, and summarization of data, from a semantic perspective. However, the metadata annotation for multimedia data remains problematic because metadata is obtained as a result of interpretation depending on domain-specific knowledge, and it needs well-controlled and comprehensive vocabulary for annotation. In this study, we proposed a collaborative methodology for developing ontologies and annotation with domain experts. The method includes (1) classification of knowledge types for collaborative construction of annotation data, (2) division of tasks among a team composed of domain experts, ontology engineers, and annotators, and (3) incremental approach to ontology development. We applied the proposed method to 11 videos on elderly care domain for the confirmation of its feasibility. We focused on annotation of actions occurring in these videos, thereby the annotated data is used as a support in evaluating staff skills. The application results show the content in the ontology during annotation increases monotonically. The number of “action concepts” is saturated and reused among the case studies. This demonstrates that the ontology is reusable and could represent various case studies by using a small number of “action concepts”. This study concludes by presenting lessons learnt from the case studies.

key words: Ontology construction, Collaborative work with domain experts, Video annotation

1. Introduction

The importance of multimedia data, which is described in a combination of multiple media such as text, audio, images, and video data, increases with the development of media processing technology. Multimedia data, especially video, is a useful resource that humans understand and analyze events that are recorded in the data. For instance, the Elderly Behavior Library [1] was developed for storing video data of the daily living activity of older adults. This data set provides useful information to the manufacturing companies producing products for older adults. The information management of multimedia data, including retrieving specific scenes from video archives, remains an important issue.

Annotating metadata to multimedia data is one of the solutions for retrieving specific scenes. Appropriate metadata contributes to efficient data retrieval by encoding semantic information. Annotation vocabulary should be controlled and reusable. Ontology is one of the tools to provide controlled vocabulary and structure of data. Several works explored applying ontology to data annotation [2]–[4]. Ontology also provides background knowledge that is typically not directly observable from the dataset, such as the subsumption relations of the objects. Such background knowledge can be relevant in retrieving multimedia data from databases.

An annotation and an ontology development for domain-specific data, such as elderly care, still require domain experts’ efforts. Considering the annotation to the Elderly Behavior Library, the interpretation by domain experts in elderly care is required. For instance, detection of care task quality from video data requires interpretation of what actions are performed and the actions are satisfied with certain prescriptive regulation. This knowledge should be built by collaboration with domain experts and ontology engineers.

This paper proposes a collaborative methodology to develop an ontology and annotation for domain-specific data. The methodology includes classification of knowledge types for collaborative construction of annotation data, and division of tasks among domain experts, ontology engineers, and annotators. We applied the proposed methodology to 11 video data in the elderly care domain to realize a staff evaluation support system by using the annotated data. We introduce one case study to examine the lessons learnt.

The paper is structured as follows. Section 2 presents related work on ontology-based annotation, ontology development tools, and methodology with collaborative work. The methodology to construct ontology and annotation is proposed in Sect. 3. Section 4 presents to present the case studies in which we applied the proposed methodology to the elderly care domain. We provide an overview and statistical information of 11 cases and a detailed qualitative description of one case. Section 5 presents a discussion including lessons learnt from the case studies. The conclusion closes the paper.

2. Related Work

Retrieving semantic information from multimedia and annotating semantics to multimedia data is one of the trends. Cavaliere et al. approached the challenge of video recognition [5] from the use-case of scene understanding of video data collected by unmanned aerial vehicles. They combined object and activity detection and a rule-based approach to
detect complex activities, such as “man crosses the road”. Thomsen and Smith proposed a framework for an ontology-based fusion of sensor data and natural language to detect the abnormal situations in the defense domain [6]. Their approach involves generating a knowledge graph from an image sensor and regular report written in natural language. Thus, scene recognition or understanding in multimedia is attracting researchers and ontologies are one of the key technologies.

2.1 Annotation Based on Ontologies

Ontology is used to annotate various types of data. Baclawski et al. proposed Semantic Media Wiki [2] to annotate text-based web documents based on an ontology. Their framework includes a standard taxonomy on content management. The effort focuses on mainly text information. The challenge of Coustaty et al. [3] aimed to use ontology for image retrieval of historical paintings. They added shape information and historical background knowledge in order to represent the shape of the letter qualitatively. And, the result recognized by the module of image recognition was qualitatively represented. Background knowledge could be combined with the results to present information.

Some studies have also been conducted regarding video annotation. Suchan et al. provided a framework of semantic question answering about the audience of films [7]. It handles spatial information of the video and the attention of the audience in a qualitative way. The system was able to answer the semantic question like “How is the spectator attention shifting, when the camera is moving / after a cut / during a long shot?” Damiano et al. built Drammar [8] as an ontology for annotating data on dramas. In this ontology, a drama is regarded as a series of actions by characters. It was applied to annotate some movie works in [4]. Both these works focused on film analysis, not domain-specific topics.

2.2 Ontology Development Tool

As mentioned above, ontology is used for annotation and the development methodologies has been also an objective of research. Computer-aided support for developing ontologies is one of the research topics. Especially, support for collaborative ontology development has been investigated as follows. Protégé [9] is one of the most famous ontology editor. It has another edition for support collaboration and is called WebProtégé [10]. It provides contextualized threaded discussions and notes attached to entities in the editing ontology. The editing history is also managed. When the entity in which a user watches is modified, then the system provides an alert to the user. User can also import ontologies from BioPortal⁠.⁠

Hozo is another ontology editor handling context-dependent concept in the ontology. Sunagawa et al. extended its functions for collaborative ontology development [11]. Hozo supports a distributed environment via the internet, similar to WebProtégé. They considered the versioning issues related to dependencies between different ontologies. They clarified 17 patterns of ontology edit in a theoretical way and provided countermeasures for each pattern. Some of these countermeasures were implemented in Hozo and they provide a use-case in plant ontology.

Both of these tools provide a function for collaborative ontology development. However, these functions focus on support in a distributed environment not collaboration among ontology engineers and domain experts. As mentioned in [10], methodological aspect like communication forum is necessary to realize collaboration among ontology engineer and domain experts.

2.3 Ontology Development Method

Gene ontology has been developed and maintained by the Gene Ontology Consortium [12]. It is designed to provide a comprehensive vocabulary that can describe the shared elements of molecular biology. To make the elements descriptive, three different taxonomies of molecular function, biological process, and cellular component take a role in guiding developers and users of the ontology. Constructing a community in which the interactions of editing and utilization of the ontology between domain experts is possible to develop comprehensive and consistent ontology. In addition, the tool supports to develop the ontology in a consistent manner [13].

Takhom et al. have proposed an ontology development method in the sustainable development domain [14]. In this field, various stakeholders, such as data analyzers, public researchers, market economy experts, etc., participate in ontology development. In that case, the same term may have different meanings in different background domain. It is the purpose to solve it in the ontology development in their study. For that purpose, they provided a method including a concrete scenario and an online forum for communication between experts from different domains.

Grüninger and Fox have proposed the notion of competency questions for Toronto Virtual Enterprise (TOVE) ontology [15]. The notion is the basis for the rigorous characterization of the problems that the constructed ontology and the ontology-based model are able to solve. In this sense, they can claim that their ontology and models have the competence to represent necessary contents if they can solve these questions. These competency questions are generated by requirements of the ontology for the components of the system.

2.4 Discussion of Related Work

As outlined here, many efforts have been undertaken to annotate data using ontology. Especially, in recent years, the annotation for the video has increased. The challenge on how to construct ontology and relate data-knowledge for the annotation in a specific domain is a still remaining issue. The Gene Ontology project [12] provides methodol-
ogy to construct domain-specific ontology and Protégé [11] can support the ontology construction. However, there is little research that is located in the intersection of a domain-specific ontology and video annotation. This study provides a collaborative methodology of ontology construction with clarification of knowledge types to apply in video annotation in a specific domain with multiple case studies. The lessons learnt will be useful for the community of data and information management for multimedia data.

3. Method

3.1 Overview of the Methodology

Our proposed methodology is intended to be used for (1) collaborative ontology development with domain experts and (2) annotation of video data which is retrieved from the video archives. Ontology development for domain-specific knowledge requires collaboration among domain experts and ontology engineers. The domain experts provide domain-specific knowledge and the ontology engineers formalizes the knowledge into a controlled vocabulary and conceptual schema. Annotation is a task that takes a cost, but it also requires domain experts because the annotation task requires interpretation of the target world from a domain-specific viewpoint. Therefore, we propose a methodology to classify the types of knowledge and to divide tasks by each stakeholder.

The workflow starts from the task performed by a domain expert. The domain expert chooses video for annotation and creates complex queries for post-analysis in parallel. The details of these complex queries is described in the next section. This task is performed in parallel with the annotation task. After choosing a video, the domain expert starts to create an action sequence. In the first iteration, the domain expert creates an action sequence without a controlled vocabulary. Subsequently, an ontology engineer constructs ontologies based on the action sequence. The ontology engineer extracts the terms and formalizes a definition of each term to format the action sequence based on the ontology. The constructed ontology is delivered to the domain expert and the domain expert check the usage and the definition are appropriate as representations of the action sequence in the video. If necessary, the domain expert modifies the terms which are extracted from the constructed ontologies. If the ontology contains all the concepts to represent the action sequence, then the ontology engineer formats the action sequence by using the ontology editor. After the second iteration, the domain expert refers to the constructed ontology and creates action sequences by using the terms in the ontology. If the ontology lacks concepts to represent the action sequence, the ontology engineer modifies the ontology by consulting with the domain expert. Such that, the ontology is created iteratively if the scope is large or ambiguous. The next step is annotating metadata to the video by an annotator. The annotator uses the formalized action sequences based on the constructed ontology.

In parallel with the annotation process, the chosen video is analyzed using object detection algorithms. The detected information is formalized in the RDF format. After the annotation, the annotation data is also formalized in RDF format and both RDF files are integrated into single file for post-analysis.

Finally, formatted complex queries are applied to the integrated RDF file for post-analysis. In this case, the complex query retrieves the specific scene that caregivers perform the action appropriately according to official guidelines or care facility’s policy.

The basic idea of constructing an ontology iteratively comes from the related work, Activity First Method [16]. Our proposed methodology extends this method to the annotation phase and the division of tasks for each stakeholder. The main difference between conventional work is to assign different tasks to each stakeholder. 1) We assigned a domain expert to define the covered range of the target world which concepts should be represented. 2) The ontology engineer organizes the concepts consulting with the domain experts. 3) The well-organized representation enables an annotator to annotate without the specialty of the domain. The division of the roles for each stakeholder is shown in Table 1.

3.2 Knowledge Types

We classify the knowledge types into four layers according to their domain specificity. We extended Layer 2 from conventional constraint-based queries to vector-based representation.

Layer 0 is a top-level ontology independent of the subject domain. The top-level ontology provides guidelines to develop domain-specific ontologies. The reusable ontologies have been published such as Basic Formal Ontology [17], a Descriptive Ontology for Linguistic and Cognitive Engineering [18], and Yet Another More Advanced Top-level Ontology (YAMATO) [19].

Layer 1 is a domain ontology that provides schema and terminology to represent a target world. It is built depending on the application system. A schema of abstract concepts, such as an action, is reusable among interdisciplinary domains, on the other hand, a taxonomy of the concepts de-

| Table 1 | Assigned tasks corresponding to the role of participants. |
|---------|----------------------------------------------------------|
| Task/Role | Domain expert | Ontology engineer | Annotator | System engineer |
| Choose a video | X | - | - | - |
| Create action sequence | X | X | - | - |
| Construct ontologies | - | X | - | - |
| Annotate metadata to the video | - | - | X | - |
| Create complex queries for analysis | X | X | - | X |
| Post analysis | - | - | - | X |
Layer 2 is a complex query that exploits the knowledge graphs. Competency check of an ontology is a common strategy in ontology development to define the scope of the ontology as mentioned in Sect. 2. However, when developing a human-centric intelligent agent that supports human daily living activities, what to be identified or answered is rather unobvious compared to manufacturing components. “Is he fine?” is a simple question but the underlying computation and required information can be quite complex. To answer simple questions that are not predefined, it is likely that the query shall be decomposed into a combination of certain primitives of human nature. While defining these primitives is not within the focus of this paper, there is available information that can be combined to represent human behavior, including end-to-end DNN detectors. We propose a new Layer 2 that is responsible for capturing this type of knowledge. It is intended to be used after the knowledge graphs are generated based on the constructed ontologies. We extend the conventional idea of competency questions [15] from the formal constraint-based representation to any kind of computation including SPARQL and knowledge-graph embedding [20]. The former is a constraint-based query that retrieves information from an ontology-based knowledge base, such as a knowledge graph. The latter is a vector-based computation of a knowledge graph. It can be used to query information from a knowledge graph based on latent semantic features, such as the similarity-based knowledge graph query. The possible use case includes a combination with multimodal embeddings, such as video clips for efficient information retrieval [21].

Layer 3 is an annotation data that represents the semantics of video contents. It is translated into a knowledge graph in the RDF format. In this study, we focus on occurring in the video because one of the features of the video data is fostering in a time interval. The annotation data are constructed using the terminology defined in Layer 1. As Renzel et al. mention [22], domain experts play crucial roles in the annotation of multimedia because of the difference of amateurs’ skill. One of the features of our proposed methodology is to provide an annotation method that collaborated with domain experts, ontology engineers, and annotators.

3.3 Collaborative Methodology for Video Annotation and Ontology Construction

We describe the detail of the methodology according to the overview shown in Fig. 1.

3.3.1 Creation of Action Sequence

We create action sequences from video with collaboration between domain experts and ontology engineers as mentioned in Sect. 3.1. In this study, we define an action as a kind of process that an actor affects the state of an operand according to some teleological viewpoint. This definition is looser than the usual definition that requires the intention of the actor or the actor must be a sub-class of an agent because we would like to apply the same structure of the action to the function of the artifact which is used by the agent. For instance, the function of a wheelchair can be interpreted as moving the patient and its interpretation comes from the user who uses the wheelchair. The exact distinction between the user’s action and the function of the wheelchair is important for clarifying the difference between action and function, but we do not take it in deep in this study.

First, the domain expert chooses a video from the video archive. In this study, we assume that the final application system includes a Deep Neural Network (DNN) detector for object detection and their states in the video and analysis of caregivers’ performance in the video. Based on the target application, the video is selected for the ease with which the DNN detector can detect the objects and meaningfulness of its assessment of caregivers’ performance.

After the video selection, the domain expert creates action sequences that occur in the video. S/he observes the video and detects which actions occur from the viewpoint of evaluating the caregivers’ performance. The sequences are recorded in tabular format as shown in Fig. 2. They include actions that have occurred, its actor, and the time when they occurred based on the narrative descriptions. In this phase, the ontology engineer offers the domain expert to use the terms in ontologies, in case they are available. The domain expert does not consider the terms in the first iteration as mentioned in the next section.

3.3.2 Construction of Ontologies Based on the Action Sequences

In the first iteration, the ontology engineer constructs an ontology based on the action sequences according to the activity first method [16]. This method enables the extraction of concepts that are necessary to represent actions that are occurred in the videos. The method makes the ontology gradually larger with the satisfaction of necessary expressiveness. The basic elements for describing action are provided by YAMATO [19], so the tabular form (shown in Fig. 2) separates the row as the components of actions, e.g. state change,
operands, and actors.

After the second iteration, the ontology engineer formalizes the action sequences by filling the table with terms in the ontologies. If the corresponding terms are not in the ontologies, the ontology engineer consults with the domain expert. This consultation resolves to clarify the meaning of the terms and to check for similar terms in the ontologies. The ontology engineer obtains the domain knowledge and the domain expert understands what terms in the ontologies through the consultation. After this consultation, the ontology engineer formats the information to the ontology-based description such as Web Ontology Language (OWL)†. S/he assigns a uniform resource identifier (URI) to each component of the action sequences through the process. The URIs are used for the next annotation phase.

3.3.3 Annotation Based on the Action Sequences

The annotator annotates the ontology-based action sequences to the video. In this phase, the annotator mainly creates time information when particular actions have occurred in the video. The point is diving the interpretation of actions and annotating the time interval of the action. The former task needs domain knowledge if the target domain needs experiences, such as medicine and elderly care. The latter is not required after the interpretation of the actions. Thus, we divide these tasks to reduce the cost of domain experts.

3.3.4 Creating Complex Queries for Post-Analysis

The latter two phases are performed in parallel. The domain expert and the ontology engineer work collaboratively in this phase. First, the domain expert creates queries in natural language according to the goal of the analysis, such as the evaluation of caregivers. The queries represent specific scenes in which the expert has an interest. The ontology engineer clarifies the queries by consulting with the domain expert to use terms in the ontology and formalized by existing theories. For example, we used Allen’s time theory [23] to represent relationships between time intervals in which actions occur, and we use Region Connection Calculus [24] to represent relationships between regions of objects. This process is related to Layer 2 as mentioned in Sect. 3.2.

3.3.5 Object Detection and Post-Analysis of Data

We also apply DNN detectors for human and object classification and recognition of the regions where they are located in the video. We used YOLO [25] and 3D Multi-Person Pose Estimation (3DMPPE) [26] in this phase. The results and annotated data were formatted to Resource Description Framework (RDF)†† format for the post-analysis. The details are described in [27].

4. Case Studies

We performed 11 case studies for the assessment of the proposed methodology. One of the cases is described in detail and after which, we provide statistical information that results from ontology construction through the cases.

4.1 Resource

First, we introduce the resources that we used in these case studies as a premise.

The Elderly Behavior Library is a video archive of elderly behavior developed by the Ministry of Economy, Trade and Industry in Japan, which was developed by business related to commerce transaction optimization/product safety (action data acquisition project of older adults, etc. for the realization of the vintage society) [1], [28].

The archive includes video data with attributes that describe the behavior of older adults. The videos were recorded at nursing homes and in ordinary households. The attribute includes physical and cognitive functions, behavior types, and behaviors of each elderly person in the video. The expected usage is to retrieve related data on product-usage and safety and to give idea generation support for the product makers. A user can retrieve the video data by setting the condition of physical/cognitive functions, behavior type, product appearing in the video. The number of collected video data was 1,699 on April 14, 2020.

We employed YAMATO [19] as a guide to construct domain concepts, as in our previous study [29] which was performed on the customer-contact service domain. One of the roles of top-level ontology is enhancing ontology construction by giving developers general guidelines about

†https://www.w3.org/TR/owl2-overview/

††https://www.w3.org/TR/rdf11-primer/

![Fig. 2 Action sequence in tabular format](image-url)
how to view the target domain, and some domain ontologies, including sustainability science [30], the abnormal states in clinical domain [31], color emotions in psychology [32], [33] and ontology of gene [34] are constructed with reference to YAMATO.

4.2 Overview of Case Studies

We applied the proposed methodology to the elderly care domain. The stakeholders included four people, specifically a domain expert, an ontology engineer, an annotator, and a system developer. The domain expert possessed experience as a manager of an elderly care facility. However, there is no expertise or experience in ontology construction. The ontology engineer possessed experience in cooperative research in the nursing domain but was not an elderly care specialist. The annotator and the system developer have no knowledge of elderly care. The assigned tasks are shown in Table 1 as mentioned in Sect. 3.1.

The purpose of this case study is to provide proof of the concept system of a service skill assessing platform [27], which supports to analyze caregivers’ performance using videos massively annotated with semantics about human interaction. The massive annotations are provided by manual annotation and a DNN detector. This paper focuses on manual annotation based on the ontology and its collaborative construction.

First, the domain experts retrieved 11 videos from the video archive. The 11 videos were selected based on the criteria (i) the interaction among caregiver and care receiver is included, (ii) the care can be distinguished as good or bad clearly, (iii) the video is DNN detector friendly, e.g. Most of the caregiver’s body is visible to detect the location where they are present in the frame. The selected videos are listed in Table 2.

Then, we performed each task according to the proposed methodology respectively for each video file. It makes each person process their tasks in parallel. The detail of one case study is provided in the next section. As a result, we obtained 11 annotated videos and confirmation of the analysis result on the Proof of Concept system. Table 3 shows the transition of the statistical information of the ontology that was constructed iteratively through the 11 cases.1

---

1The number of cases 7 and 9 were completed by the way shown in Appendix A.

---

| Case #  | Content                                      |
|--------|----------------------------------------------|
| 1      | Standing up and moving support 01            |
| 2      | Lying, sitting, standing and moving support 01 |
| 3      | Lying, sitting, standing and moving support 02 |
| 4      | Moving support 01                            |
| 5      | Sleeping support                             |
| 6      | Standing up and moving support 02            |
| 7      | Lying, sitting, standing and moving support 03 |
| 8      | Moving support 02                            |
| 9      | Moving support 03                            |
| 10     | Moving support 04                            |
| 11     | Moving support 05                            |

Table 2  List of cases

| Case #  | # of Classes | # of Properties |
|---------|--------------|-----------------|
| 1       | 2,028        | 1,705           |
| 2       | 2,138        | 1,796           |
| 3       | 2,384        | 2,160           |
| 4       | 2,559        | 2,379           |
| 5       | 2,762        | 2,646           |
| 6       | 2,853        | 2,767           |

We constructed an ontology from Case 1 and began revising the ontology for Case 2 after the concepts in the ontology satisfied for representing Case 1. The statistics contain the number of classes and properties derived from YAMATO. The classes and the properties were increased around 100 to 200 to represent the new case. The added classes represent actions and artifacts which are referred to in the new case. For instance, “make a person stand” was added for Case 2. Then, the abstract class “change a person’s standing position” was added for organizing the existing concept “stand up.” The former one denotes that the actor and its operand are the same, and the latter one denotes that the actor and the operand are different. Another example is that “remote controller” was added linking to the addition of the concept “switch oﬀ the remote controller” as its operand. Thus, the new concepts for representing new cases are added to the ontology.

4.3 One Case Study: Standing Up and Moving Support

4.3.1 Video Selection

The video archive [28] does not provide IDs to refer to the video clip uniquely. We provide lists of tags to refer to individual video clips. The tag consists of the attribute name and its value. The archive provides 14 tag attributes and approximately 100 values in total. We selected the video clips retrieved according to the following tag set (The attribute and value are separated by “:”). Place: Elderly care facility, Environment: Move, Furniture: Chair, Behavior: Standing. The selected video shows good care that standing up and moving support by one caregiver to one care receiver in a private room of an elderly care facility. The video is 59 s long. The point is that the caregiver provides the service by showing dignity to the care receiver. The purpose of the post-analysis is to detect the point.

4.3.2 Creation of Action Sequence and Construction of Ontologies

The domain expert listed the actions that are performed by both the caregiver and the care receiver in the video. The criteria for listing were determined as follows for post-analysis. (1) It was easy to distinguish between good care and bad care, and (2) the action was externally distinguishable so that the DNN detector can be recognized in the future. The determination of what constitutes a unit of action was performed according to YAMATO’s criteria which is to change
the state of the target world.

The result was confirmed by the ontology engineer, and the vocabulary was unified and each meaning was clarified. Modularization was attempted by describing the operand of the action and its state change separately. As a result, 55 action concepts and 106 operand concepts of the action were obtained. The action sequence was rearranged and formalized according to the ontology construction. We converted from tabular format to OWL format by using Protége [9]. We referred to the concepts defined in the constructed ontology and assigned unique URIs to the components of the action sequences.

Figure 2 shows an example of the created action sequences. The domain expert recognized who performed which action to whom or what. In this case, he recognized that a caregiver touched both arms of the care receiver softly at the starting time of the video clip. The caregiver did the other five actions simultaneously. We also described the action of the care receiver. The participant, action, and operand were described in unified vocabulary given by the ontology.

We classified the actions into subclasses of physical actions, cognitive actions, and composite actions. Table 4 shows representative examples of each class. I will explain the definition with an example "Move to left" action. The action is a subclass of physical action. This action is described as that the location of one object was changed after this action was performed. To describe this, we prepared the properties to describe "has_object," "has_reference_point," and "has_terminal_state." The object of "has_object," "has_reference_point," and "has_terminal_state" are "operand" whose location is changed, "reference_object" which defines the left, and "left_side." These properties are specialized from the properties defined in YAMATO ontology.

### 4.3.3 Annotation Based on the Action Sequences

We used ELAN [35] as an annotation tool enabling users to annotate arbitrary text for time series data such as video and audio. The selected video is read into ELAN, and the actions were annotated while watching the video. In this study, the URI allocated in the previous section is directly input to the ELAN tool to facilitate the linkage with the ELAN file and the ontology file in the post-analysis. The annotator can carry out the annotation to the video without hesitation because which terms to be used are determined by the domain expert in the previous section.

### 4.3.4 Creating Complex Queries for Post-Analysis

The purpose of the analysis of the video in this case study is to retrieve the specific scene that displays the caregiver’s skill based on the actions which they performed. We created complex queries for retrieving specific scenes that they performed a particular skill. These complex queries were composed of temporal relation of the actions and the regional relation among the objects (including human bodies) in the video. We describe an example of a complex queries considering moving support. A care receiver should direct the direction to move with an oral explanation before moving. At first, the domain expert creates such criteria to represent the scenes that shows good or bad care in natural language, then the ontology engineer formalized them with the terms in the ontology for the system engineer who processes the data at the post-analysis.

For instance, the domain expert extracted the sentence such as “Body language is a good way to communicate with older adults who have difficulty communicating verbally. Nonverbal communication tends to convey more information rather than only verbal communication” [36]. The domain expert decided to retrieve the scenes that a caregiver properly provided a care service based on this general statement. The recommendation can be represented as a constraint-based query, so we formalized the statement in a formal notation by using the notion of Allen’s interval algebra and Region Connection Calculus. An example of the complex query is shown below.

\[
\text{Show moving direction by gesture := (Explain next action < Move) AND (Direct) AND (Explain next action) AND (has_Action.Operand(Direct, Direction))}
\]

The italicized words show the terminology defined in the ontology and the symbol “<” shows the temporal relation that an action A ”precedes” an action B referring to Allen’s temporary interval algebra [23]. has_Action.Operand(X, Y) denotes by X an action and by Y an operand of the action X. This complex query means that explain next action is performed before move action with direct action. This query was translated into a SPARQL query, as shown in Table 5, for scene retrieval from the resulting knowledge graph. By using this complex query, the domain experts can get a specific scene from the video archive for...
evaluation of staff’s skill.

We created 25 such queries for this case study. We retrieved the scenes by using the complex queries from the video clips that we used in this case study.

4.3.5 Application Interface

The domain expert prepared a set of complex queries retrieving a desired scene. Figure 4 shows the results of scene-retrieval by one complex query. The left part shows that there are three scenes that satisfy the complex query as results. When a user clicks a green button a specific scene is displayed as shown on the right side. The user can understand that this staff performs desirable care appropriately. When the system is used as a training system, the system provides scenes that different staffs perform the same care activity in parallel. Such a comparison of different scenes of the same activity will help a trainee consider how to perform their own activities.

5. Discussion

We discuss the case studies from the perspectives of 1) reuse rate of concepts, 2) lessons learnt and 3) instance management as a remaining issue.

5.1 Reuse Rate of Concepts

The reuse rate $R$ was calculated based on Eq. (1), i.e. the ratio of concepts that are necessary for the case to the concepts that are available in the constructed ontology for the previous case study.

$$R = \frac{|C_{case} \cap C_{onto}|}{|C_{case}|}$$

where $C_{case}$ is a set of classes that are necessary to represent the case. $C_{onto}$ is a set of classes that were constructed in the previous cases. $|A|$ is a cardinality of the set $A$.

We calculate the $R$ for each case. The result is shown in Table 6. We show the two rows of the reuse rate. The upper row denotes the reuse rate of all concepts that include actions, actors, objects as an operand of an action, or location. The lower one denotes the reuse rate of action which is a focused concept in this study.

We discuss the reuse of action concepts in detail. Our proposed methodology was intended to increase the concepts incrementally through the case studies. It is assumed that the number of additional action concepts decreases as the number of cases increases and saturates at a certain point. The results are shown in Fig. 5. 23 action concepts were required to annotate the 1st video. From there, as the number of videos increased, it increases to 8, 4, 5, 3, 2, 1, 3, 2, 2, and 2. As a result, 55 action concepts were used to annotate of 11 videos. The reuse rate remained around 50%–60% until the first 4 cases, however, it was more than 75% after the ninth case. The number of re-used concepts depended on the complexity and types of care which occurring in each video.

5.2 Lessons Learnt from Case Studies

Through these case studies, we experienced difficulties and
beneficial outcomes as follows.

- **Difficulty**: The interaction among the domain expert and the ontology engineer occurred even though the role was divided. For instance, during the first half, the ontology engineer needs domain knowledge to interpret the action sequence that is created by the domain experts. We needed interaction among both stakeholders to solve this.

- **Benefit**: The ontology engineer could understand the output from the domain expert should be related to action because the domain expert and the ontology engineer got an agreement that the action sequence is the order of actions that are occurred in the video. The agreement helps us to make the coverage clear.

5.3 Instance Management

In these case studies, instance management was not considered. However, it was necessary to manage the identity of people and objects between videos. That is, performance evaluation of caregivers cannot be performed unless the same person is identified in different locations, at different times, and with videos from different cameras. This requires the integration of external knowledge, such as lists of caregivers and care receivers, with the constructed knowledge graph. Even for the same video, identity management of people and objects in different frames is required. However, the conventional computer vision concentrates on class recognition other than individual recognition.

To address this issue, we develop an identification system between frames as shown in Fig. 6. The system provides lists of entities that are annotated to the video and bounding boxes (bboxes) are automatically detected by a DNN detector. The annotator recognizes which entity corresponds to a given bbox and creates a link on the GUI. The system calculates the moving average of the bboxes’ locations in different frames and if the value is less than a given threshold then the system detects the identity of bboxes in different frames. These two functions solve the second instance management issue.

6. Concluding Remarks

We have proposed a collaborative methodology for developing ontology and annotation for multimedia data. The methodology includes (1) classification of knowledge types for collaborative construction of annotation data, (2) division of tasks among domain expert, ontology engineer and annotator for collaborative work, and (3) incremental approach for ontology development. We applied the proposed methodology to 11 video files in the elderly care domain in order to realize a staff evaluation support system by using the annotated data. Through the case studies, we confirmed that the proposed methodology contributes to the construction of ontology incrementally and in collaboration with domain experts. The ontologies which developed via this methodology are expected to contribute to retrieve information from video archive as information management. The demonstration system shows pieces of the proof-of-concept. The contribution of this study also includes lessons learnt from the application of the proposed approach to the elderly care domain.

Acknowledgments

The part of this paper is based on results obtained from a project, JPNP18002, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] Y. Nishida and K. Kitamura, “Smart living lab for living function resilient society adapting to changes in living functions,” Journal of The Society of Instrument and Control Engineers, vol.58, no.2, pp.89–95, 2019 (in Japanese).
[2] K. Baclawski, “Semantic annotation of collaborative work environments,” Applied Ontology, vol.12, no.3-4, pp.313–322, 2017.
[3] M. Coutatuy, N. Tsopoz, A. Bouju, K. Bertet, and G. Louis, “Towards ontology-based retrieval of historical images,” Applied Ontology, vol.10, no.2, pp.147–167, 2015.
[4] V. Lombardo, R. Damiano, A. Pizzo, and C. Terzulli, “The intangible nature of drama documents: An FRBR View,” Proceedings of the 2017 ACM Symposium on Document Engineering, pp.173–182, 2017.
[5] D. Cavaliere, V. Loia, A. Saggese, S. Senatore, and M. Vento, “A human-like description of scene events for a proper uav-based video content analysis,” Knowledge-Based Systems, vol.178, pp.163–175.
Completion of the Case Studies

As a result of practical issues, we lacked the information on ontology in Cases 7 and 9. The information in Table 3 was completed in the following way. First, we extracted the concepts that were used to represent Cases 7 and 9, respectively. We call the set of concepts "Set A." Second, we extracted the concepts that are used to represent cases before Cases 7 and 9, respectively (We call the set of concepts "Set B"). Third, we compared the Sets A and B and extracted the concepts that were not included in Set A from Set B. We defined the extracted set of concepts was newly added to represent Case 7 and 9.
Satoshi Nishimura received his Ph.D. in Communication Technology from Osaka University, Japan in March 2015. He works at National Institute of Advanced Industrial Science and Technology (AIST) in Japan since 2015. His research interest is a formalization of the procedural knowledge and production rules based on ontologies.

Julio Vizcarra received his Ph.D. in Computer Science from Center of Computing and Research, Mexican National Polytechnic Institute, Mexico in March 2016. He works at National Institute of Advanced Industrial Science and Technology (AIST) in Japan since 2019. His research interests are knowledge graphs, natural language processing, pattern recognition, artificial intelligence, and sentiment analysis.

Yuichi Oota graduated and received Dr. Engineering (Underwater Sound) from Tokai University Graduate School in 1983 and 1984 respectively. He received a certified-caregiver in 2016. He had worked in OKI Electric Industry Co. Defense System Division (1984–2011), worked as Director of Group Home Motouchi (2012–2017), and now works at AIST since 2017.

Ken Fukuda received his Ph.D. in Information Science from University of Tokyo and was awarded Research Fellowship for Young Scientists as a doctoral student. He joined AIST as a research scientist in 2001 and was a visiting lecturer at University of Tokyo and a visiting associate professor at Waseda University. He is the Team Leader of Data Knowledge Integration research team and expert on ontology application design, knowledge representation, text mining, Semantic Web, and Service Science with particular interests in computing human activities in daily living.