Towards Semantic Data Management of Visual Computing Datasets: Increasing Usability of MetaVD

Yasunori Yamamoto[1,2], Shusaku Egami[1], Yuya Yoshikawa[3] and Ken Fukuda[1,*]

[1]National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan
[2]Research Organization of Information and Systems, Tokyo, Japan
[3]Chiba Institute of Technology, Chiba, Japan

Abstract

MetaVD is a meta video dataset that interlinks six existing Computer Vision related datasets such as Charades and Kinetics-700. While MetaVD contributes to enhancing video recognition performance, we found two issues as follows. First, some is-a relationships defined and linked by MetaVD are inconsistent, including one circulation of is-a relationships. Second, all concepts in MetaVD are from the six datasets, and therefore some of them are not well semantically arranged, leading to a possibility of inefficient training of video recognition models. Here, we propose a knowledge graph dataset in Resource Description Framework (RDF), which links MetaVD concepts to those in the Commonsense Knowledge Graph (CSKG) and RDFizing MetaVD itself. By linking it, we can more easily detect inconsistent concept relationships. Furthermore, it allows us to link MetaVD concepts to those of conceptually higher ones. Then, some SPARQL queries were issued to it to evaluate its feasibility. The RDF dataset and SPARQL queries mentioned in this extended abstract are downloadable from https://github.com/aistairc/MetaVD-CSKG .

Keywords

Knowledge Graphs, Video Datasets, RDFization, Data Refine, Data Linking

1. Introduction

Lots of video datasets for human action recognition have been published, such as UCF101 [1] and Kinetics-700 [2]. Each of them covers its specific domain, and we experience poor recognition performance when applying a model trained on a dataset to one of the other domains. MetaVD [3] tackled this issue by interlinking concepts of six popular datasets for human action recognition. They defined three relation types of equality, similarity, and hierarchy, and then annotated 568,015 relation labels in total by hand to the pairs of concepts obtained from the original datasets using these relationships. After interlinking these datasets, they confirmed that video recognition performance were increased by proposing two methods of how to utilize the interlink.

However, we found two issues in MetaVD. First, there are several semantically inconsistent relationships. For example, there are the following relationships in MetaVD: *Playing_ice_hockey*

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece

*Corresponding author.
yy@dbcls.rois.ac.jp (Y. Yamamoto)
0000-0002-6943-6887 (Y. Yamamoto); 0000-0002-3821-6507 (S. Egami); 0000-0001-8584-8470 (Y. Yoshikawa); 0000-0001-7366-1094 (K. Fukuda)

is-a *hit*, and *hit* is-a *Volleyball*. In addition, there is a circulation for three is-a relationships. Since the is-a relationship reflects a semantically hierarchical and directional relation between concepts of the original datasets, a concept in the circulating relationships is claimed to be more general than itself; that is to say, it is inconsistent. Second, MetaVD interlinks the six datasets, and all the concepts are from them. Therefore, there are some cases where a more general concept can be used to group a set of concepts that have a common trait, leading to a possibility of gaining a better recognition performance [4, 5]. For example, there are *playing_ice_hockey*, *playing_lacrosse*, and *playing_basketball* in MetaVD, which can be grouped by a concept of *playing_game*.

We propose integrating MetaVD with the CommonSense Knowledge Graph (CSKG) [6] after RDFizing MetaVD. The CSKG is a commonsense knowledge graph that amalgamates seven widely-recognized sources, such as ConceptNet and Visual Genome and we found that almost all MetaVD concepts can be linked with ConceptNet concepts. The conversion of MetaVD into RDF allows semantic validation of the dataset via the SPARQL query language [7]. Furthermore, SPARQL enables the management of user-specific subset data generation of the MetaVD, which is required to train a special-purpose human-action recognition model tailored for distinct MetaVD user applications, such as in the sports field.

2. Dataset

We use three datasets: MetaVD[1], ConceptNet[2], and CSKG[3]. They are downloadable in CSV, SQLite, and TSV formats, respectively.

2.1. MetaVD to ConceptNet

As ConceptNet has a large number of concepts that can be linked to those in MetaVD, we first attempted to align them. Of the 966 MetaVD concepts, 397 were fully aligned and 560 were partially aligned. "Fully aligned" here means that both concepts exactly match each other after normalization described below. The seven concepts that could not be aligned include *Powerbocking*, *Shotput*, or *barbequing*. “Partially” here means that there were some words aligned to a MetaVD concept consisting of multiple words such as *Turning on a light*. In addition, since some aligned concepts were identical words such as *accordion* for *playing_accordion* [Kinetics-700] and *Playing_accordion* [ActivityNet], the total number of concepts is 868.

We used ConceptNet Numberbatch[4] to obtain a corresponding ConceptNet ID. The version we obtained is 19.08 English. We extracted headwords from it and constructed an index based on them. Before looking it up, we normalized the terms as follows:

- camel case to multiple words
- all letters to lowercase
- dash symbol to underline
- expressing multiple senses using slash symbols expanded to each concept

[1] https://github.com/STAIR-Lab-CIT/metavd

[2] https://github.com/ldtoolkit/conceptnet-lite

[3] https://zenodo.org/record/4331372

[4] https://github.com/commonsense/conceptnet-numberbatch

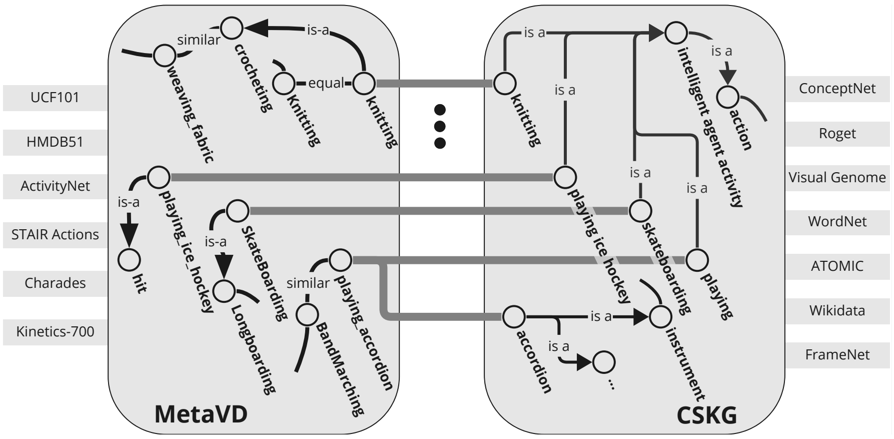

Figure 1: Schematic diagram of relationships between CSKG and MetaVD. We made links between CSKG nodes and MetaVD ones, which are depicted as thick gray lines.

2.2. MetaVD to CSKG

Next, for each concept, we obtained ConceptNet ID such as `/c/en/accordion` and retrieved CSKG edges and nodes. These CSKG edges link these ConceptNet IDs to the CSKG nodes. Note that the retrieved CSKG nodes are within one hop from the ConceptNet IDs. Fig 1 delineates relationships between CSKG and MetaVD in our work. We linked MetaVD nodes to CSKG ones. As for *accordion* as an example, the following statements can be obtained.

1. Accordion is an instrument.
2. Accordion is a man-made object.
3. Accordion is a musical instrument.

We used Knowledge Graph Toolkit (KGTK) [8] to retrieve designated data from CSKG. More specifically, for each ConceptNet concept linking from MetaVD, we issued a query to retrieve nodes whose IDs begin with that concept. For example, we issued a query that retrieves nodes whose IDs matched the regular expression of `/c/en/accordion(/.+)?` along with their edges and nodes linked by these edges. The resultant node IDs were the following.

- /c/en/accordion
- /c/en/accordion/n
- /c/en/accordion/n/wn/artifact
- /c/en/accordion/v

As a result, we obtained 831 CSKG nodes linked to MetaVD. In addition, 47 293 nodes and 90 506 edges linked from these nodes were retrieved.

2.3. Degree of abstraction

Although it is not trivial how to measure the granularity of each MetaVD concept, we assume that ConceptNet graph structure can provide supporting evidence. Each ConceptNet node often has *IsA* relationships, such as *an accordion IsA instrument*, and semantically higher concepts have more child nodes linked with the IsA relationship. The number of children of *accordion* is three, and that of *instrument* is 68.

2.4. RDFization

We used TogoDB [5] to build the MetaVD RDF dataset with a subset of CSKG linked to MetaVD. TogoDB accepts table data in CSV or TSV formats and provides a GUI-based environment where we can edit configuration that defines how to generate RDF data from a given table. We took a general approach of RDFization from a table dataset. A row becomes a set of triples whose subject is from the cell value of the ID column and other column names and the corresponding cell values denote its properties. RDF dataset from the MetaVD-CSKG data were also built in the same way. The number of triples for MetaVD and MetaVD-CSKG were 25 218 and 1 190 904, respectively. We also built an ancillary RDF dataset to link MetaVD data IDs to ConceptNet IDs including partial words of MetaVD concepts. We loaded the RDF dataset into Fuseki Version 4.7.0[6].

3. Use Cases

We constructed several SPARQL queries to verify the usefulness of the RDF dataset to fulfill the two purposes mentioned above.

3.1. Inconsistency Checking

It seems that the current MetaVD has some semantically inconsistent is-a relationships. Here, we say inconsistent in terms of the degree of abstraction obtained by counting children of a concept in ConceptNet. For example, while *Volleyball* has one child and *hit* has 13, *hit is-a Volleyball* in MetaVD.

3.2. MetaVD subsetting for customization

Another use case is to make an MetaVD subset to train a human action recognition model tailored to one's specific purpose. As a feasibility study, we issued a SPARQL query to retrieve all MetaVD concepts related to *intelligent agent activity*, which returned 200 results. It includes *Playing_ice_hockey* from ActivityNet, *roller_skating* from Kinetics-700, and *playing_guitar* from STAIR Actions.

4. Conclusion

We made an RDF dataset consisting of MetaVD and subset of CSKG linking to MetaVD. In addition, we defined degree of abstraction based on the ConceptNet graph structure. By using these data, we propose a way of showing potentially semantically inconsistent relationships in MetaVD. In addition, we propose a method of making an MetaVD subset in terms of a given abstract concept such as intellignet agent activity. This method enables us to train a video recognition model for a specific purpose.

[5] http://togodb.org/

[6] https://jena.apache.org/

On the other hand, we need to evaluate the result. There are 1010 relationships defined in MetaVD, and we are considering whether we can check all of them manually or not. Future works include utilizing CSKG graph structure to consider datasets other than ConceptNet. In addition, we will consider a method of suggesting a relation in MetaVD based on the CSKG relationships.

Acknowledgments

This paper is based on results obtained from a project, JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), and JSPS KAKENHI Grant Number JP22K18008.

References

[1] K. Soomro, A. R. Zamir, M. Shah, UCF101: A dataset of 101 human actions classes from videos in the wild, arXiv preprint arXiv:1212.0402 (2012).

[2] J. Carreira, A. Zisserman, Quo vadis, action recognition? a new model and the kinetics dataset, in: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6299–6308.

[3] Y. Yoshikawa, Y. Shigeto, A. Takeuchi, Metavd: A meta video dataset for enhancing human action recognition datasets, Computer Vision and Image Understanding 212 (2021) 103276. doi:https://doi.org/10.1016/j.cviu.2021.103276.

[4] A. Dhall, A. Makarova, O.-E. Ganea, D. Pavllo, M. Greeff, A. Krause, Hierarchical image classification using entailment cone embeddings, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020) 3649–3658.

[5] T. Yamazaki, S. Ito, K. Ohara, Hierarchical image classification with conceptual hierarchies generated via lexical databases, in: Proc. of the The 29th International Workshop on Frontiers of Computer Vision, 2023.

[6] F. Ilievski, P. Szekely, B. Zhang, Cskg: The commonsense knowledge graph, Extended Semantic Web Conference (ESWC) (2021).

[7] E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF, W3C Recommendation, 2008. URL: http://www.w3.org/TR/rdf-sparql-query/.

[8] F. Ilievski, D. Garijo, H. Chalupsky, N. T. Divvala, Y. Yao, C. Rogers, R. Li, J. Liu, A. Singh, D. Schwabe, P. Szekely, KGTK: A toolkit for large knowledge graph manipulation and analysis, in: International Semantic Web Conference, Springer, 2020, pp. 278–293.