A Semantic Representation of Pedestrian Crossing Behavior

He Tan[1,*], Florian Westphal[1]

[1]*Department of Computing, School of Engineering, Jönköping University, Sweden.*

Abstract

In this paper, we focus on the crucial task of understanding and modeling pedestrian behavior, which is essential for numerous applications. We introduce a semantic representation of pedestrian crossing behavior. The representation is to capture the sub-events within a behavior and the spatial-temporal evolution of interactions between pedestrians and other objects involved in crossing events. We demonstrate its practical application by utilizing it to analyze pedestrian crossing behavior from road user movement data (i.e. trajectories). By constructing a knowledge graph from detailed road user dynamics data using this representation, we enable queries that address safety concerns related to pedestrian crossing behavior, aiding traffic engineers in their work on urban traffic infrastructure design.

Keywords

Ontology, Knowledge Graph Construction from trajectory data, Visual Question Answering

1. Introduction

Vulnerable road users, such as children, the elderly, and disabled individuals, are integral to the dynamics of city traffic. They play essential roles in establishing a sustainable, active, and inclusive mobility environment. Therefore, understanding and modeling pedestrian behavior is fundamental for many applications, including traffic flow analysis, traffic safety improvement, urban planning, and intelligent driving systems. Pedestrian crossing behavior is one of the main aspects of pedestrian behavior that has been studied in numerous research studies [1]. It involves the actions and movements of pedestrians while crossing streets or roadways, often guided by traffic signals, road markings, and traffic conditions.

Traditionally, stochastic, linear regression, and discrete choice models are used to build an understanding of how pedestrians make crossing decisions considering various factors related to people, roadway, traffic, traffic controls and traffic rules [1, 2]. Parameters of the models are estimated from survey and/or questionnaire data or manually screened video recordings. More recently, agent-based modeling has been used to model road users as intelligent agents attempting to make rational decisions in uncertain and complex situations. However, most work has focused on modeling vehicle behaviors. Very limited studies dedicated to develop models for other road users such as pedestrians [3]. Often, these studies focus on pedestrian-vehicle conflicts and model pedestrians' collision avoidance mechanism [4]. The parameters of the models are typically calibrated by detecting and tracking road users from video data or using results from literature.

*Corresponding author.

At the same time, researchers across diverse disciplines, such as computer vision, artificial intelligence (AI), cognitive science, and neuroscience, have conducted numerous studies on understanding human activities. Depending on their complexity, human activities can be classified into different levels: gestures, actions, interactions, and group activities [5]. This paper specifically focuses on understanding pedestrian crossing behavior on the level of human-object interactions. Human activity is a spatial-temporal evolution of interactions [6]. Here, we present a semantic representation of pedestrian crossing behavior, capturing sub-events of a behavior and their temporal and spatial structures. Previous studies within computer vision have suggested that a structured spatial-temporal representation can lead to more accurate activity understanding and improve the performance of various computer vision tasks, including image captioning and visual question answering [7]. Studies (e.g., [8, 9]) in the AI area indicate that such a representation can cope with less training data by incorporating prior knowledge and help to understand human activities.

In this paper we present the structured spatial-temporal representation of pedestrian crossing behavior, and describe its application to gain an understanding of pedestrian crossing behavior from recorded road user dynamics data. Utilizing the representation, a knowledge graph is constructed from road user dynamics data. The queries over the knowledge graph can answer safety related questions on pedestrian crossing behavior for traffic engineers and help with their work on urban traffic infrastructure design.

The remainder of this paper is organized as follows: In Section 2, we introduce the methods that have been employed to semantically represent human activities, particularly pedestrian behaviors. In Section 3, we present our approach to semantically representing crossing behaviors. Section 4 outlines the utilization of semantic representation within the context of traffic data analysis for traffic engineers. Finally, the paper concludes the work in Section 5.

2. Related Work

Pedestrian behavior has been widely analyzed in various research works using a plethora of methods. Nevertheless, understanding pedestrian behavior remains challenging due to the inherent complexity of human activities. Despite the diverse analysis methods used, the significance of semantic representation in understanding pedestrian behavior has often been overlooked. Only a limited number of studies have explored the semantic representation for pedestrian behavior.

Chai et al. [10] utilized fuzzy logic to model the cognitions and behavioral patterns of pedestrians, in order to understand the effect of age and gender when pedestrians are crossing a signalized crosswalk and jaywalking. Gharebaghi et al. [11] developed a mobility ontology for people with motor disabilities (PWMD). Specifically, it considers the interactions between people and both the social and physical environment. The ontology was used to support the development of assistive technologies for the mobility of PWMD. Fang et al. [12] developed an ontology defining various kinds of road users, including pedestrians, and describing their relationships. The concepts from the ontology are used to define the rules for describing the interactions between road users and to support rule-based reasoning for predicting road users' behavior.

In this paper, we present a semantic representation of pedestrian crossing behavior. The representation describes the dynamic evolution of interactions between pedestrians and objects within the physical environment over time, capturing interactions in both spatial and temporal dimensions.

In 1970, Hägerstrand [13] introduced the concept of a time-space path in understanding human activities. This theory has laid the groundwork for trajectories that have been shown to be useful in representing people's movements. Inspired by Hägerstrand's work, Orellana and Renso [14] developed an interaction ontology. The ontology conceptualizes the characteristics of pedestrian movement behaviour. It has focused on identifying various movement patterns from time-space paths, and the different categories of interactions, spatial and temporal contexts, behavior, and the high-level relations between these concepts. Logic-based reasoning is used to categorize pedestrian movement behavior based on its movement patterns, interactions, and contexts.

Meanwhile, in cognitive science and neuroscience, it has been recognized that segmentation is a fundamental component of perception, playing a critical role in understanding activities. People tend to perceive ongoing continuous activity as series of discrete events (or called segments) [15, 16, 17]. The relationships between segments are encoded in partonomic hierarchies [18]. Coarse segmentation is often related to objects' locations and their goals, and the causal relations between their actions. Fine segmentation is closely linked to changes in the interactions between objects [19]. Building on these findings in cognitive science and neuroscience, Ji et al. [7] proposed a spatial-temporal scene graph to represent human activity and to improve the performance of action recognition and few-shot action recognition using neural networks. Mlodzian et al. [9] presented an ontology that was tailored for representing entities and their spatial and temporal relations in traffic scenes in the nuScenes dataset[1]. A knowledge graph was constructed from the nuScenes dataset using the ontology and provided as a benchmark dataset for developing advanced trajectory prediction models.

In this paper, drawing from these insights in cognitive science, neuroscience and computer vision, we propose a structured spatial-temporal representation for pedestrian crossing behavior and present its application to gain an understanding of pedestrian crossing behavior from road user movement data.

3. Semantic Representation

In this section we present the semantic representation for pedestrian crossing behavior. A pedestrian crossing behavior can be seen as a dynamic evolution of interactions between pedestrians and objects within the physical environment over time. Every crossing behavior can be broken down into segments, each representing a distinct phase of the behavior. These segments capture the changes of the interactions between pedestrians and objects in both physical and temporal dimensions, and together represent the pedestrian crossing behavior. For example, Fig. 1 shows a crossing event, which is extracted from road user behavior measurement performed at a zebra-free crossing in Lindholmen, Gothenburg in Sweden. Fig 1-a displays the trajectories of the pedestrian and other moving objects involved in the event. The red

---

[1]https://www.nuscenes.org/

trajectory represents a pedestrian, the blue trajectory represents a cyclist, and the cyan trajectory represents a light vehicle. Fig 1-b1 to b8 show a sequence of distinct segments that capture the changes in interactions between pedestrians and objects over time during the event. These interactions are expressed in a set of triples, as shown in Fig 2. Each triple follows the format `((id, object_1), spatial_relation, (id, object_2))`, where the object 1 is a moving object such as pedestrian, cyclist, and vehicle, and the object 2 can be a moving object or a static object such as crossing, area and sidewalk, and id is the unique identifier for each object.
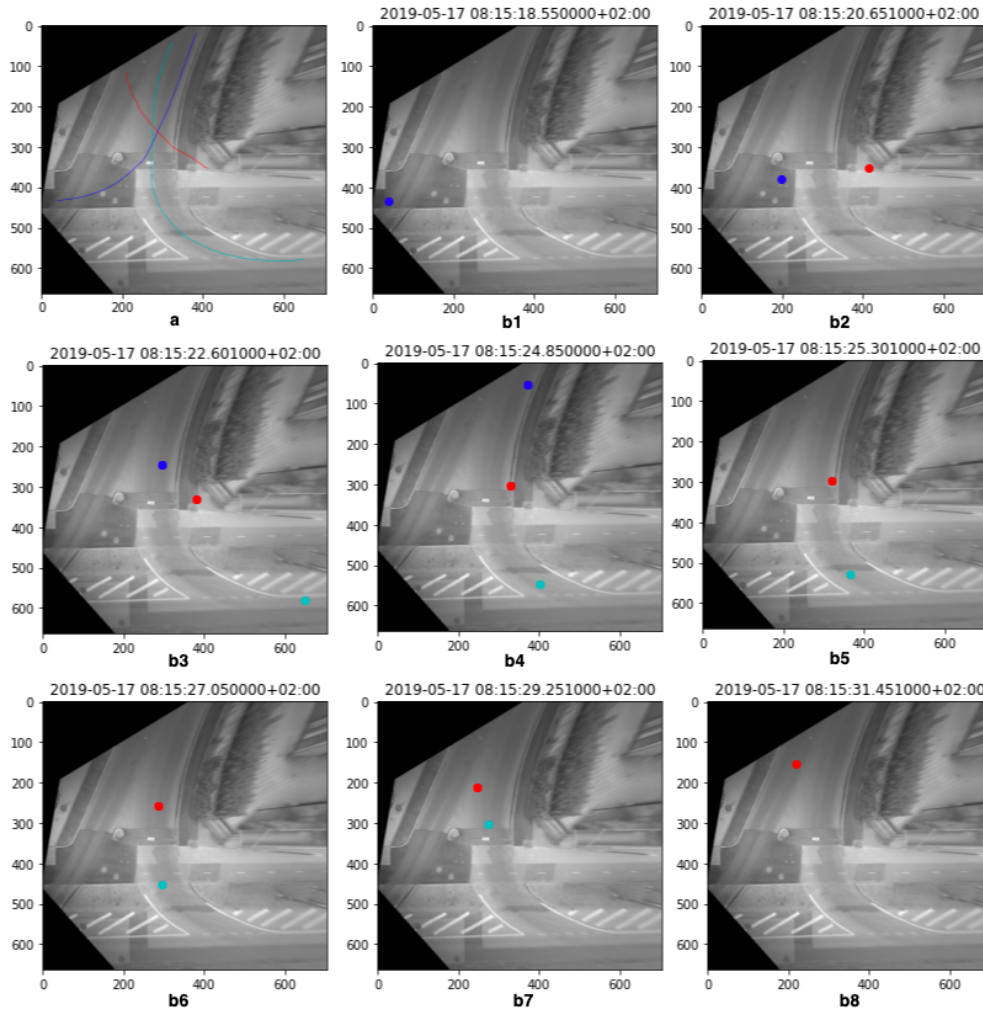

Figure 1: An example of pedestrian crossing behavior.

Fig 3 illustrates the current version of the ontology designed to represent the spatial-temporal evolution of crossing behavior. This ontology is accessible on GitHub[2]. Since *segment* is often

---

```
b1:   ((1152, bicyclist), on, (0, crossing))

b2:   ((1152, bicyclist), on, (0, crossing)
      ((1153, pedestrian), right_far_away, (3, right_sidewalk))

b3:   ((1152, bicyclist), on, (0, crossing))
      ((1152, bicyclist), far_away, (1154, light_vehicle))
      ((1153, pedestrian), right_close_to, (3, right_sidewalk))
      ((1153, pedestrian), far_away, (1154, light_vehicle))
      ((1154, light_vehicle), out_of_crossing_area, (1, crossing_area))

b4:   ((1152, bicyclist), on, (0, crossing))
      ((1152, bicyclist), far_away, (1154, light_vehicle))
      ((1153, pedestrian), on, (0, crossing))
      ((1153, pedestrian), far_away, (1154, light_vehicle))
      ((1154, light_vehicle), out_of_crossing_area, (1, crossing_area))

b5:   ((1153, pedestrian), on, (0, crossing))
      ((1153, pedestrian), far_away, (1154, light_vehicle))
      ((1154, light_vehicle), out_of_crossing_area, (1, crossing_area))

b6:   ((1153, pedestrian), on, (0, crossing))
      ((1153, pedestrian), far_away, (1154, light_vehicle))
      ((1154, light_vehicle), on, (0, crossing))

b7:   ((1153, pedestrian), on, (0, crossing))
      ((1153, pedestrian), close_to, (1154, light_vehicle))
      ((1154, light_vehicle), on, (0, crossing))

b8:   ((1153, pedestrian), on, (0, crossing))
```

Figure 2: The interactions and their changes in the crossing behavior.

related to regions in an image in computer vision, the term *frame* is used instead. In computer vision, a video can be divided into a sequence of frames. Each frame represents a single still image in the video sequence. The blue arrow represents `subclass` relations between concepts. Currently, the ontology includes only a limited number of categories for both moving and static objects. However, additional categories will be integrated as the ontology continues to undergo further development.

4. Usage of the Representation

In this section, we describe an application of the semantic representation of pedestrian crossing behavior mentioned beforehand in Section 3. The application aims to provide information support for traffic engineers during traffic infrastructure planning and development, with a particular focus on pedestrian safety. In the application, pedestrian crossing behaviors are described using the semantic representation, and a knowledge graph is constructed for these behaviors. Subsequently, a number of queries serve as question-answering tools to provide information for traffic engineers.
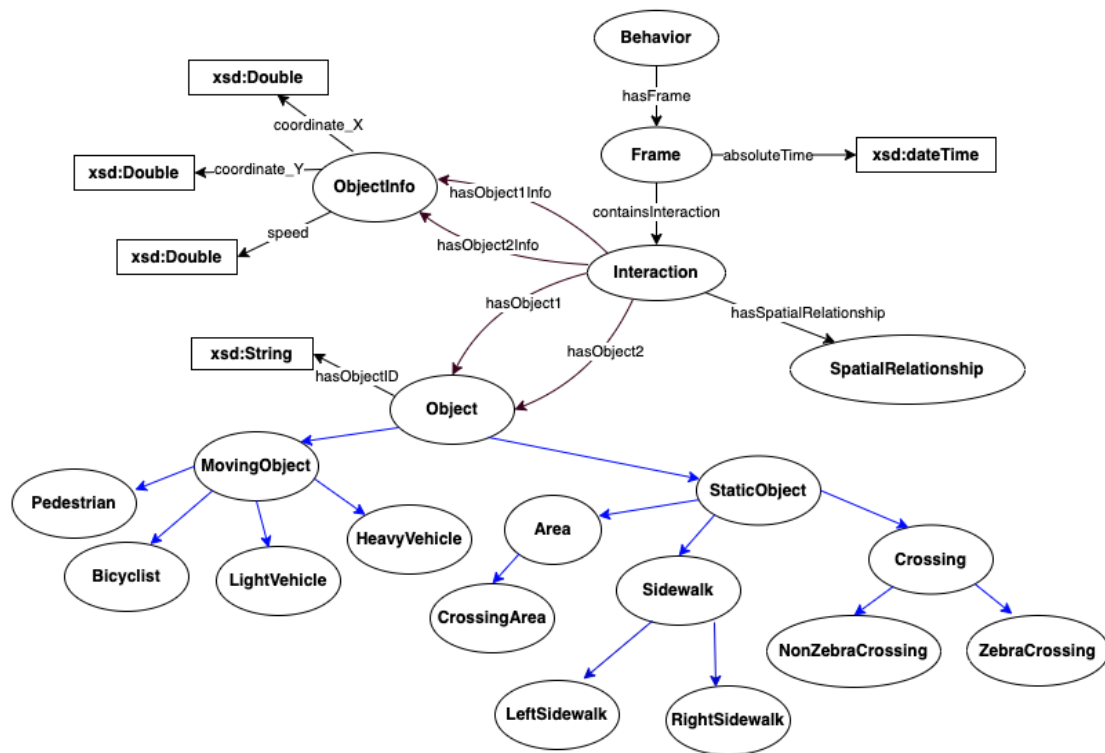
Figure 3: The ontology representing temporal and spatial structures of the interactions in pedestrian crossing behavior.

Crossing Behavior Dataset

The crossing behavior dataset is prepared from the traffic measurement aforementioned in Section 3. Fig 4 shows an example frame extracted from the dataset. The road user positions and trajectories are displayed in a camera view, overlaid on the anonymized video frame. The measurement is performed by Viscando AB[3] using the 3D&AI based infrastructure sensor OTUS3D. The total period of the measurment is 11 hours and 5 minutes. The data contains trajectories of all road users recorded 10 times per second. Trajectories contain the unique track ID for each object, the UTC time stamp, position (i.e. X-coordinate and Y-coordinate), velocity (i.e. object speed in the direction of motion (km/h)) and object type. Currently, the object types include pedestrian, cyclist, light vehicle and heavy vehicle. Vision data are processed in the embedded computational unit and removed within 20 ms from being captured. Thus, the dataset is stored fully anonymously, ensuring compliance with the General Data Protection Regulation (GDPR) of the European Union[4], because personal information is neither stored in the sensors nor transmitted.

---

[3]www.viscando.com
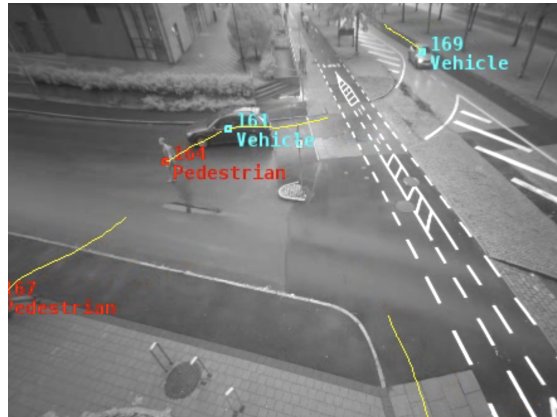[4]https://gdpr-info.eu/

Figure 4: An example frame extracted from the dataset.

Knowledge Graph Construction

In this section we describe the construction of the knowledge graph that describes the pedestrian crossing behaviors recorded in the aforementioned dataset. Since the application is to support the traffic infrastructure planning and development prioritizing pedestrian safety, the construction has focused on the crossing events involving pedestrians/cyclists but also vehicle(s). The spatial relationship between objects was calculated based on the physical distance between them. The current spatial relationships include the ones between moving objects, i.e. *close_to* and *far_away*, and the ones between a moving object and a static object, i.e., *left_close_to*, *right_close_to*, *left_far_away*, *right_close_to*, *on*, *out_of_area*. If the x-coordinate of one object is smaller than that of another, the former is positioned to the left of the latter; otherwise, it is positioned to the right.

When the information was extracted from the aforementioned dataset, the ontology described in Section 3 was populated, and the knowledge graph is set up. Fig 5 shows the fragment of the knowledge graph that represents the pedestrian crossing behavior presented in Section 3.

Question Answering

In this section, we present the SPARQL queries to retrieve answers from the knowledge graph or to inquire the information from it to formulate responses to a few example questions that traffic engineers might pose.

First, two prefixes are predefined for the following SPARQL queries, i.e., tsdata (http://www.example.com/ontology/traffic_scene_kg# and ts (http://www.example.com/ontology/traffic_scene_ontology.owl#).

Example 1: describe a crossing behavior. The query will return an RDF dataset describing a specific crossing behavior. Fig 5 shows a visualization of such an RDF dataset. It was generated by using the Stardog Studio visualization tool[5]. Such an RDF dataset can also be converted into text, allowing traffic engineers to easily access and understand the information [20].
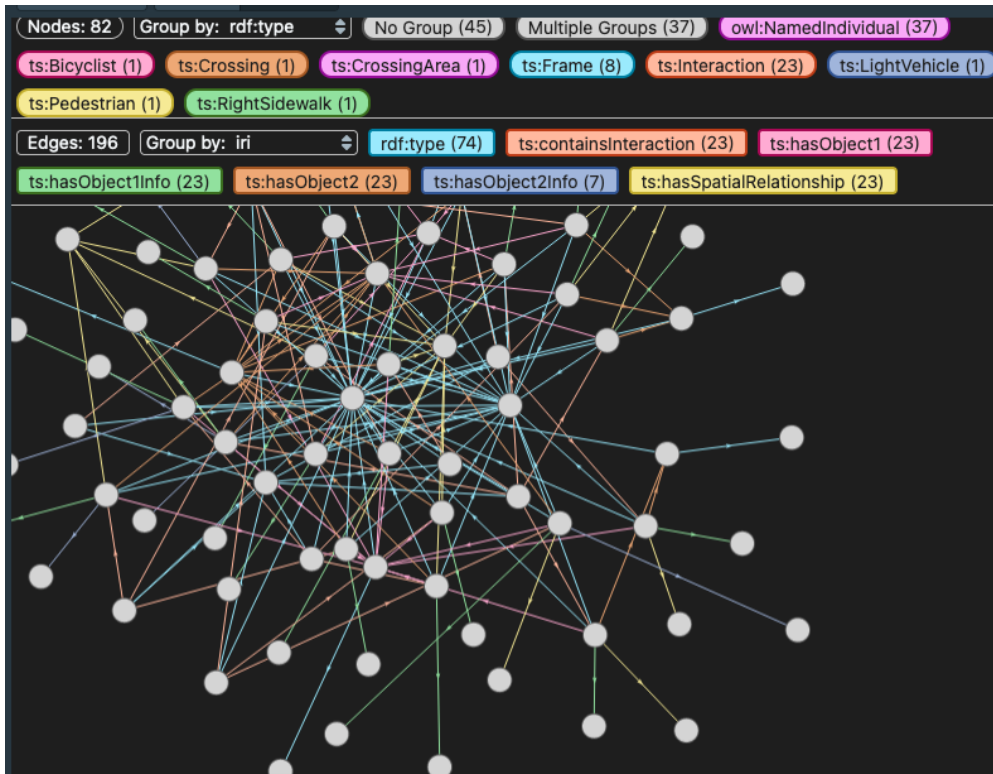
---

[5]https://cloud.stardog.com/

Figure 5: The fragment of the knowledge graph that represents a pedestrian crossing behavior.

```
DESCRIBE ?f ?i ?obj1 ?obj2
WHERE {
  tsdata:behavior_103 ts:hasFrame ?f.
  ?f ts:containsInteraction ?i .
  ?i ts:hasObject1 ?obj1.
  ?i ts:hasObject2 ?obj2 }
```

Example 2: find and describe the crossing behaviors within a specified time period. The query will return an RDF dataset containing the crossing behaviors within the specified time period. For each behavior, using the query given in Example 1, it can be described in text, allowing traffic engineers to access and understand the information.

```
SELECT DISTINCT ?b
WHERE {
  ?b a ts:Behavior .
  ?b ts:hasFrame ?f.
  ?f ts:absoluteTime ?t.
  FILTER (?t >= "2019-05-17 08:00:00"^^xsd:dateTime
       && ?t <=  "2019-05-17 08:20:00"^^xsd:dateTime)
}
```

Example 3: find the crossing events where pedestrians/cyclists are close to vehicles and return the frames when this happens.

```
SELECT DISTINCT ?b ?f
WHERE {
  ?b rdf:type ts:Behavior   .
  ?b ts:hasFrame ?f .
  ?f ts:containsInteraction ?i .
  ?i ts:hasSpatialRelationship ts:close_to .

  { ?i ts:hasObject1 ?obj1 .
      {?obj1 rdf:type ts:Pedestrian}
        UNION  {?obj1 rdf:type ts:Bicyclist}.
    ?i ts:hasObject2 ?obj2 .
      {?obj2 rdf:type ts:HeavyVehicle}
       UNION {?obj2 rdf:type ts:LightVehicle}}
  UNION
  {?i ts:hasObject2 ?obj2 .
      {?obj2 rdf:type ts:Pedestrian}
        UNION {?obj2 rdf:type ts:Bicyclist}.
    ?i ts:hasObject1 ?obj1 .
      {?obj1 rdf:type ts:HeavyVehicle}
        UNION {?obj1 rdf:type ts:LightVehicle}}.
}
ORDER BY ?b
```

Example 4: find the crossing events where pedestrians/cyclists are close to vehicles and their speed is too fast. Such behaviors are considered unsafe. The query is an extension of the one given in Example 3, with the addition of the following triple patterns and filter.

```
  ?i ts:hasObject1Info ?obj1info .
  ?i ts:hasObject2Info ?obj2info.
  ?obj1info  ts:speed ?s1.
  ?obj2info  ts:speed ?s2
  FILTER (?s1 >= highest_safe_speed || ?s2 >= highest_safe_speed )
```

Example 5: find the crossing behaviors where pedestrians take a shortcut to the crossing, specifically by crossing diagonally across the street. Such a behavior is considered unsafe. This query is separated into two steps. The first step is retrieving the crossing events and the frames where pedestrians are involved. In the second step, the y-coordinates of the pedestrians during the crossing are retrieved. If the changes of the y-coordinates exceed a certain threshold, the pedestrians are considered as being taking a shortcut to the crossing. As an example, the following query shows how to retrieve the y-coordinates of the pedestrian involved in the crossing event presented in Section 3.

```
SELECT DISTINCT ?f ?y
WHERE {
  tsdata:behavior_103 ts:hasFrame ?f.
  ?f ts:containsInteraction ?i .
  ?i ts:hasObject1 ?obj1.
  ?obj1 rdf:type ts:Pedestrian.
  ?i ts:hasObject1Info ?obj1info.
  ?obj1info ts:coordinate_Y ?y
}
```

The queries over the knowledge graph are not limited to the ones listed in this paper. More complex queries can be constructed when traffic engineers require more intricate information. For instance, cyclists swinging out at a crossing are considered as unsafe behavior. Such behavior can be identified by using a number of queries in a simple program.

5. Conclusions

In this paper we have introduced a structured spatial-temporal representation of pedestrian crossing behavior and demonstrated its application in understanding such behavior from recorded road user dynamics data. By leveraging this representation, we construct a knowledge graph from the road user dynamics data. Queries made over this knowledge graph can address safety-related inquiries regarding pedestrian crossing behavior for traffic engineers, supporting them in urban traffic infrastructure design work.

In future work, we aim to enhance the ontology by incorporating more granular categories of road users and other spatial relations between objects. Additionally, we plan to develop a tool that enables traffic engineers to pose text-based questions and receive text-based answers, thereby enhancing their workflow support. This way of interacting with the road user dynamics data could be implemented with the help of large language models (LLMs) and retrieval-augmented generation (RAG) [21]. In such a system, the user's question would be translated into a query against the knowledge graph and the returned information would be transformed into natural language text by the LLM.

Apart from querying the constructed knowledge graph to gain insights into the behavior of different traffic participants, the proposed semantic representation could also serve as base for trajectory prediction approaches. With increased interest in the development of self-driving cars, predicting the behavior of other traffic participants has come more into focus [22]. For this task, it is important to understand the spatial relationships between different actors. Hence, different approaches have been investigated to integrate these relationships into trajectory prediction, including simple graph structures [23], heterogeneous graphs [24], and knowledge graphs [9]. While clearly belonging to the latter category, our representation focuses particularly on static objects, such as road infrastructure elements, to capture their impact on trajectories of traffic participants. Presumably, this will not only improve trajectory predictions, but also help traffic engineers to understand the impact different road infrastructure elements will have on traffic. Therefore, another direction of the future work is to investigate the incorporation of the constructed knowledge graphs with graph neural networks for trajectory prediction.

Acknowledgments

References

[1] E. Papadimitriou, G. Yannis, J. Golias, A critical assessment of pedestrian behaviour models, Transportation research part F: traffic psychology and behaviour 12 (2009) 242–255.

[2] E. Papadimitriou, G. Yannis, J. Golias, Theoretical framework for modeling pedestrians' crossing behavior along a trip, Journal of transportation engineering 136 (2010) 914–924.

[3] P. Nasernejad, T. Sayed, R. Alsaleh, Modeling pedestrian behavior in pedestrian-vehicle near misses: A continuous Gaussian Process Inverse Reinforcement Learning (GP-IRL) approach, Accident Analysis & Prevention 161 (2021) 106355.

[4] A. Rasouli, I. Kotseruba, Intend-wait-cross: Towards modeling realistic pedestrian crossing behavior, in: 2022 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2022, pp. 83–90.

[5] J. K. Aggarwal, M. S. Ryoo, Human activity analysis: A review, Acm Computing Surveys (Csur) 43 (2011) 1–43.

[6] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma, et al., Visual genome: Connecting language and vision using crowdsourced dense image annotations, International journal of computer vision 123 (2017) 32–73.

[7] J. Ji, R. Krishna, L. Fei-Fei, J. C. Niebles, Action genome: Actions as compositions of spatio-temporal scene graphs, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10236–10247.

[8] K. Ramirez-Amaro, M. Beetz, G. Cheng, Transferring skills to humanoid robots by extracting semantic representations from observations of human activities, Artificial Intelligence 247 (2017) 95–118.

[9] L. Mlodzian, Z. Sun, H. Berkemeyer, S. Monka, Z. Wang, S. Dietze, L. Halilaj, J. Luettin, nuScenes Knowledge Graph - A Comprehensive Semantic Representation of Traffic Scenes for Trajectory Prediction, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2023, pp. 42–52.

[10] C. Chai, X. Shi, Y. D. Wong, M. J. Er, E. T. M. Gwee, Fuzzy logic-based observation and evaluation of pedestrians' behavioral patterns by age and gender, Transportation research part F: traffic psychology and behaviour 40 (2016) 104–118.

[11] A. Gharebaghi, M.-A. Mostafavi, G. Edwards, P. Fougeyrollas, S. Gamache, Y. Grenier, Integration of the social environment in a mobility ontology for people with motor disabilities, Disability and Rehabilitation: Assistive Technology 13 (2018) 540–551.

[12] F. Fang, S. Yamaguchi, A. Khiat, Ontology-based reasoning approach for long-term behavior

prediction of road users, in: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE, 2019, pp. 2068–2073.

[13] T. Hägerstrand, What about people in Regional Science?, Papers of the Regional Science Association 24 (1970) 6–21. URL: http://dx.doi.org/10.1007/bf01936872. doi:10.1007/bf01936872.

[14] D. Orellana, C. Renso, Developing an interactions ontology for characterising pedestrian movement behaviour, in: Movement-aware applications for sustainable mobility: Technologies and approaches, IGI Global, 2010, pp. 62–86.

[15] D. Newtson, Attribution and the unit of perception of ongoing behavior., Journal of personality and social psychology 28 (1973) 28.

[16] L. Spector, J. Grafman, Planning, neuropsychology, and artificial intelligence: cross-fertilization, Handbook of neuropsychology 9 (1994) 377–392.

[17] C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, K. A. Norman, Discovering event structure in continuous narrative perception and memory, Neuron 95 (2017) 709–721.

[18] J. M. Zacks, B. Tversky, G. Iyer, Perceiving, remembering, and communicating structure in events., Journal of experimental psychology: General 130 (2001) 29.

[19] N. K. Speer, J. M. Zacks, J. R. Reynolds, Perceiving narrated events, in: Proceedings of the Annual Meeting of the Cognitive Science Society, volume 26, 2004.

[20] C. Gardent, A. Shimorina, S. Narayan, L. Perez-Beltrachini, The WebNLG challenge: Generating text from RDF data, in: Proceedings of the 10th International Conference on Natural Language Generation, 2017, pp. 124–133.

[21] P. Zhao, H. Zhang, Q. Yu, Z. Wang, Y. Geng, F. Fu, L. Yang, W. Zhang, B. Cui, Retrieval-augmented generation for ai-generated content: A survey, arXiv e-prints (2024). URL: http://arxiv.org/abs/2402.19473v1. arXiv:2402.19473v1.

[22] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, H. Chen, A Survey on Trajectory-Prediction Methods for Autonomous Driving, IEEE Transactions on Intelligent Vehicles 7 (2022) 652–674. URL: http://dx.doi.org/10.1109/tiv.2022.3167103. doi:10.1109/tiv.2022.3167103.

[23] J. Gao, C. Sun, H. Zhao, Y. Shen, D. Anguelov, C. Li, C. Schmid, VectorNet: Encoding HD Maps and Agent Dynamics From Vectorized Representation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[24] D. Grimm, P. Schörner, M. Dreßler, J.-M. Zöllner, Holistic Graph-based Motion Prediction, in: 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2023. URL: http://dx.doi.org/10.1109/icra48891.2023.10161468. doi:10.1109/icra48891.2023.10161468.