Towards Leveraging Commonsense Knowledge for Autonomous Driving

Sreyasi Nag Chowdhury[1,2], Ruwan Wickramarachchi[3], Mohamed H. Gad-Elrab[1], Daria Stepanova[1], and Cory Henson[4]

[1]Bosch Center for Artificial Intelligence, Renningen, Germany
[2]Max Plank Institute for Informatics, Germany
[3]AI Institute, University of South Carolina, USA
[4]Bosch Research and Technology Center, USA
[1]{firstname.lastname}@de.bosch.com, [3]ruwan@email.sc.edu,
[4]{firstname.lastname}@us.bosch.com

Abstract. Rapid development of autonomous vehicles has enabled the collection of huge amounts of multimodal road traffic data resulting in large knowledge graphs for autonomous driving. These knowledge graphs typically storing factual statements like "*scene 1 includes parking area*", have proved to be useful complements for computer vision models especially for tasks like object prediction in a scene. However, they do not capture inter-object commonsense relationships (e.g., "*car is smaller than truck*" or "*car accelerates faster than bicycle*"), on which human drivers rely subconsciously when making decisions. Existing commonsense repositories target mainly general purpose domains, and their coverage with respect to driving scenarios is very limited, prohibiting a straightforward integration. To bridge this gap, we extend existing autonomous driving knowledge graphs with commonsense knowledge, and demonstrate empirically the benefits of such extension for two downstream applications: object prediction and explainable scene clustering.

Keywords: Autonomous Driving, Commonsense Knowledge Graphs

1 Introduction

Autonomous Driving (AD) has recently emerged as an exciting research area within AI owing to its vast application potential and interdisciplinary nature. The rapid development of AD has enabled the collection and analysis of huge amounts of multimodal road traffic data extracted from semi-structured resources or learned automatically from visual driving scenes. Such data enriched with schema led to the construction of knowledge graphs (KGs) for AD [15, 2, 16]. The existing knowledge repositories for AD contain millions of facts in the form of ⟨*subject*, *relation*, *object*⟩ triples describing entities occurring in AD scenes, e.g., ⟨*scene1*, *includes*, *car*⟩ and mutual relations among them, ⟨*car*, *located_at*, *parkingArea*⟩. KGs for AD naturally complement visual and decision making systems and have been successfully exploited, in particular, via KG embeddings [14] in many applications, e.g., data search [2] and scene similarity [15].

However, current KGs for AD storing primarily factual data extracted from observations, e.g., ⟨*pedestrian*, *located_at*, *crossing*⟩ are limited in terms of commonsense assertions describing, among others, comparative properties between objects, for example, ⟨*car*, *accelerates_faster_than*, *bicycle*⟩. On the other hand, human drivers' ability for seamless decision making is largely attributed to such commonsense knowledge (CSK). Clearly, CSK-enriched KGs could be very beneficial for supporting sophisticated decision-making in AD scenarios – e.g., the CSK facts ⟨*dirt_road, has_property, muddy*⟩, ⟨*rain, causes, mud*⟩, ⟨*mud,causes, wheel_jam*⟩ can lead to a decision to avoid parking on dirt roads on rainy days or on cloudy days with a prediction of rain. Since to the best of our knowledge no KGs for AD that systematically incorporate commonsense relations exist, the goal of our work is to bridge this gap and enrich the knowledge graphs for autonomous driving with CSK assertions.

The usage of CSK in AD scenarios can be divided into two sequential sub-problems: acquisition of relevant domain-specific CSK, and its integration into autonomous driving frameworks. The existing domain-agnostic commonsense KGs (e.g., ConceptNet [11], WebChild [13], Quasimodo [9], etc.), mainly contain assertions about everyday objects and scenes, and their fraction related specifically to autonomous driving is rather limited. In [8] a KG for smart city applications has been extended with parts of WebChild related exclusively to transportation, but the resulting dataset is of limited scope containing CSK relations over only 17 manually specified object classes. Gathering CSK for AD from annotated image captions or description data sets like MSCOCO [6], Conceptual Captions [10], Visual Genome [5], Visual Storytelling [3] etc. yields assertions with spatial relations (e.g., *located_near*); however, such resources are not directly useful for extracting other physical commonsense relations, (e.g., *faster_than*). The integration of reasoning methods into AD applications has been recently targeted by, e.g., exploiting answer set programming for visual sense-making [12]. However, to the best of our knowledge no attempts of systematic collection of CSK for AD and its integration exist.

In this preliminary work, we address the limitations of existing KGs for AD by enriching them with commonsense facts from a wide range of state-of-the-art domain-agnostic CSK repositories. More specifically, the salient contributions of our work are summarized as follows[1]:

- We have enhanced the existing KG for AD [15] based on Pandaset[2] by linking it to assertions from four popular domain agnostic CSK repositories: ConceptNet [11], WebChild [13], Quasimodo [9], and CSKG [4].
- We study the effect of KG enhancement with CSK for the task of scene entity prediction, following [15] as well as explainable scene clustering [1].

2 Commonsense Knowledge for Autonomous Driving

CSK Acquisition. Defining a domain is a challenging problem which we elude in this work. As a starting point, a set of 38 road entities identified in the

---

[1] The dataset and results are available at https://github.com/boschresearch/ad_cskg
[2] https://scale.com/open-datasets/pandaset

Table 1: Examples of CSK triples for AD

Subject	Relation	Object
Bus	*is_more_efficient_than*	Car
Motorcycle	*are_more_popular_than*	Bus
Pedicab	*conceptually_related_to*	Bicycle
Standing	*located_near*	Motorcycle
Pedestrian	*are_more_susceptible_than*	Car
Walking	*related_to*	Pedestrian

Table 2: Link Prediction

		HolE	TransE
MRR	w/o CSK	0.9266	0.4133
	w/ CSK	**0.9357**	**0.4353**
Hits@1	w/o CSK	0.8729	0.2747
	w/ CSK	**0.8874**	**0.3123**
Hits@3	w/o CSK	0.9796	0.4582
	w/ CSK	**0.9846**	**0.4613**

PandaSet dataset are considered as seeds. We then expand these entities with their synonyms from WordNet [7], considering only noun senses that contain the original entity. These constitute 36 additional seeds.

The expanded set of entities (referred to as Features of Interest or FoIs) are then used to collect an AD-CSK slice from three existing CSK knowledge graphs – ConceptNet [11], WebChild [13], Quasimodo [9], as well as from a consolidated source named Commonsense Knowledge Graph (CSKG) [4]. CSKG integrates seven sources – ConceptNet, Visual Genome (a dataset of annotated images for visual CSK), ATOMIC (a procedural source), Wikidata, and three lexical sources, namely WordNet, Roget, and FrameNet. In total there are 98,738 CSK assertions ($\langle subject$, $relation$, $object \rangle$ triples) where either the subject or the object is a FoI, consisting of 7217 unique CSK relations. For experiments we work with a subset of these assertions where both the subject and the object are FoIs. This subset consists of ∼1000 assertions with ∼600 unique relations.

Removal of Synonymous CSK Relations via Clustering. Since AD-CSK is collected from various sources, it naturally contains many synonymous relations, which need to be further canonicalized. We exploit the DB-scan algorithm to cluster semantically similar relations based on their cosine similarities calculated from their respective Word2Vec embeddings. The relation clusters are further manually refined. In this way, 440 out of 600 relations are grouped into 60 dominant relation clusters. Table 1 shows some example assertions from our curated AD-CSK knowledge graph.

CSK Integration. Our proposed AD-CSK KG can then be integrated into existing KGs. For our experiments, we enrich the AD KG [15] constructed from the PandaSet consisting of 8343 scenes. Although our PandaSet-specific AD-CSK KG consists of only 1000 triples, at scene level each CSK relation connects millions of entities. E.g., the most frequent CSK relation (*located_near*) connects 8,589,486 entities adding as many edges in the CSK-enriched driving scene KG.

3   Results

We evaluate the KG embeddings learned with and without the harvested CSK assertions on two downstream tasks – scene entity prediction, and explainable scene clustering. For both tasks, embeddings learned from the CSK-enriched KG outperform those learned from the KG without CSK. This is the first evidence of the benefits of CSK integration for AD applications.

Table 3: Quality of the explainable clusters learned using ExCut

	Clusters Quality	Explanations Quality		
	Silh. Width	Cov@1	Exc@1	WRA@1
without CSK	0.0423	0.8539	0.2362	0.0157
with CSK	**0.1103**	**0.8713**	**0.2991**	**0.0698**

Table 4: Sample explainable clusters

Cluster	Explanation (without CSK)
C_1	$e_1 :< X, partOf, \texttt{scene\_123} >$
C_2	$e_2 :< X, includes, Y >, < Y, type, \texttt{Motorcycle} >$
C_3	$e_3 :< X, includes, \texttt{PedestrianWithObject} >, < X, includes, \texttt{TrafficCone} >$

Cluster	Explanation (with CSK)
C_1	$e_1 :< X, includes, \texttt{TrafficCone} >, < X, includes, \texttt{Pedestrian} >$
C_2	$e_2 :< X, includes\_ent\_slower\_than, \texttt{Car} >$
C_3	$e_3 < X, includes\_ent\_located\_near, \texttt{Walking} >, < X, includes, \texttt{VehicleWithoutRider} >$

Scene Entity Prediction. An important application for AD is predicting missing entities based on existing road entities. For example, if a ball rolls out in front of a car in a residential area, there is a high probability of children suddenly appearing in the scene. In [15] this entity prediction task has been modified to a link prediction problem which aims to derive potentially missing links in a KG from learned embeddings. Table 2 shows the results on the PandaSet KG for two popular embedding models[3], TransE and HolE. One can observe that predicting links relying on the embeddings learned from the CSK-enhanced KG yields better results across all standard metrics for evaluating link prediction.

Explainable Clustering. Intuitively, the addition of CSK to KGs for AD should provide better understanding of the scene context, which might not necessarily be visual. To assess this hypothesis, we adapt the explainable clustering approach ExCut [1], and examine whether the addition of CSK relations helps in clustering a given set of scenes based on their semantic similarity. More specifically, we cluster a set of 824 driving scenes based on the KG embeddings computed using the HolE method, and learn explanations for the discovered clusters from the underlying AD KG with and without introduced CSK relations.

Table 3 reports the quality of the clusters with and without CSK, which is measured using average silhouette width. In addition, we present the average quality for the top explanation assessed using cluster coverage (Cov), exclusive quality (Exc), and weighted average accuracy (WRA) as described in [1]. The CSK relations have a positive impact on the overall quality of the clusters and their explanations. In Table 4, we show samples of the explanations derived from the KG with and without CSK. One can observe that the explanations derived from the CSK-enriched KG contain more general relations among scene entities.

---

[3] Available in Ampligraph library https://github.com/Accenture/AmpliGraph

4   Conclusion

This preliminary work aims at studying the role of commonsense knowledge (CSK) in autonomous driving (AD) applications. To this end, we have constructed a domain-specific CSK knowledge graph (KG) and integrated it into an existing AD KG. To verify the usefulness of the constructed CSK-enriched KG, we exploit KG embeddings for two tasks – *Scene Entity Prediction*, which involves predicting the presence of entities in the scene that were not detected by computer vision tools, and *Explainable Clustering*, which clusters scenes based on their semantic similarity and generates explanations for the clusters. The preliminary results are promising, and demonstrate the importance of integrating CSK in AD applications.

References

1. Gad-Elrab, M.H., Stepanova, D., Tran, T., Adel, H., Weikum, G.: Excut: Explainable embedding-based clustering over knowledge graphs. In: ISWC (2020)
2. Henson, C., Schmid, S., Tran, A.T., Karatzoglou, A.: Using a knowledge graph of scenes to enable search of autonomous driving data. In: ISWC Satellites (2019)
3. Huang, T.K., Ferraro, F., Mostafazadeh, N., Misra, I., Agrawal, A., Devlin, J., Girshick, R.B., He, X., Kohli, P., Batra, D., Zitnick, C.L., Parikh, D., Vanderwende, L., Galley, M., Mitchell, M.: Visual storytelling. CoRR (2016)
4. Ilievski, F., Szekely, P.A., Cheng, J., et al: Consolidating commonsense knowledge. CoRR (2020)
5. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Kenji Hata, e.a.: Visual genome. Int. J. Comput. Vis. (2017)
6. Lin, T., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context. In: ECCV (2014)
7. Miller, G.A., Hristea, F.: Wordnet nouns: Classes and instances. Comput. Linguistics **32**(1), 1–3 (2006)
8. Persaud, P., Varde, A.S., Robila, S.A.: Enhancing autonomous vehicles with commonsense: Smart mobility in smart cities. In: ICTAI (2017)
9. Romero, J., Razniewski, S., Pal, K., Pan, J.Z., Sakhadeo, A., Weikum, G.: Commonsense properties from query logs and qa forums. In: CIKM. ACM (2019)
10. Sharma, P., Ding, N., Goodman, S., et al: Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning. In: ACL (2018)
11. Speer, R., Havasi, C.: Conceptnet 5: A large semantic network for relational knowledge. In: The People's Web Meets NLP. Theory and Applications of NLP, Springer (2013)
12. Suchan, J., Bhatt, M., Varadarajan, S.: Commonsense visual sensemaking for autonomous driving. CoRR (2020)
13. Tandon, N., de Melo, G., Suchanek, F.M., Weikum, G.: Webchild: harvesting and organizing commonsense knowledge from the web. In: WSDM. ACM (2014)
14. Wang, Q., Mao, Z., Wang, B., Guo, L.: Knowledge graph embedding: A survey of approaches and applications. IEEE Trans. Knowl. Data Eng. (2017)
15. Wickramarachchi, R., Henson, C.A., Sheth, A.P.: An evaluation of KG embeddings for AD data. In: AAAI Symposium: Combining ML with KE (1) (2020)
16. Zhao, L., Ichise, R., Mita, S., Sasaki, Y.: Core ontologies for safe autonomous driving. In: ISWC Posters & Demos (2015)