Towards Deep Reasoning on Social Rules for Socially Aware Navigation

Roya Salek Shahrezaie*
Santosh Balajee Banisetty*
rsalek@nevada.unr.edu
santoshbanisetty@nevada.unr.edu
University of Nevada Reno
Reno, Nevada, USA

Mohammadmahdi Mohammadi
sir.m77@gmail.com
WHIP mobility Group
Selangor, Malaysia

David Feil-Seifer
dave@cse.unr.edu
University of Nevada Reno
Reno, Nevada, USA

ABSTRACT

This work presents ideation and preliminary results of using contextual information and information of the objects present in the scene to query applicable social navigation rules for the sensed context. Prior work in socially-Aware Navigation (SAN) shows its importance in human-robot interaction as it improves the interaction quality, safety and comfort of the interacting partner. In this work, we are interested in automatic detection of social rules in SAN and we present three major components of our method, namely: a Convolutional Neural Network-based context classifier that can autonomously perceive contextual information using camera input; a YOLO-based object detection to localize objects with a scene; and a knowledge base of social rules relationships with the concepts to query them using both contextual and detected objects in the scene. Our preliminary results suggest that our approach can observe an on-going interaction, given an image input, and use that information to query the social navigation rules required in that particular context.

CCS CONCEPTS

• **Human-centered computing** → **Social navigation**.

KEYWORDS

Human-Robot Interaction, Knowledge Graph, Social Rules

ACM Reference Format:
Roya Salek Shahrezaie, Santosh Balajee Banisetty, Mohammadmahdi Mohammadi, and David Feil-Seifer. 2021. Towards Deep Reasoning on Social Rules for Socially Aware Navigation. In *Companion of the 2021 ACM/IEEE International Conference onHuman-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3434074.3447225

*Both authors contributed equally to this research.

HRI '21 Companion, March 8–11,2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447225

1 INTRODUCTION

As mobile robots keep growing and foster human-robot collaborative activities in public places, they must navigate efficiently, reliably, and socially around individuals. To successfully accept these robots in human environments, we need to explore robot navigation in the social contexts in which we live. Navigating through a crowded environment using a collision-free path is challenging and has long been a solved problem [15]. However, in human-robot environments, the challenge is no longer about navigating from one point to the other efficiently; it is more about social awareness in navigation behaviors [3–5, 21]. Social navigation in a single context or a few contexts is insufficient for real-world deployments in collaborative environments such as hospitals, shopping malls, and airports. Autonomously detecting the scene/context and adapting an appropriate social navigation strategy is vital for social robots' long-term applicability in dense human environments.

In prior studies, context-aware navigation may have considered object detection using pre-defined rules to define their navigation behavior. We are interested in using environmental context and object information for more appropriate social navigation. We introduce a novel approach to use context recognition and object detection to execute context-appropriate social rules. For example, a robot's social navigation strategy in an art gallery or a museum differs from the social navigation strategy for hallway navigation. Similarly, the social rules within the same context vary based on other factors in the environment. For example, social navigation rules in a gallery with a person are different from social navigation in a gallery with a person viewing artwork. In the former case, the robot may only need to account for the proxemic rules around the person, whereas in the latter case, the robot has to account for both proxemic rules and rules associated with activity space (the space between human and the artwork).

2 RELATED WORK

Previous studies have investigated ways for robot navigation in an indoor context with the presence of a human. We can classify this body of work in two broad areas: context-aware navigation and knowledge base for robot navigation.

2.1 Context-Aware Navigation

Research in HRI, *socially-aware navigation*, codified these unspoken rules into robot path planning algorithms using both analytical [8, 19] and learning-based approaches [16, 21]. However, most of the approaches to social navigation codifies rules for a single

context, such as appropriate hallway behaviors [8, 21], not passing in between two conversing people [19], and avoiding activity zones [16]. With an increased interest in the social navigation research community, researchers have identified the need for a unified socially-aware navigation (USAN) [2, 12], i.e., social navigation not just for a single context but for multiple contexts.

Prior work in SAN deals only with a single context; to the best of our knowledge, no method can handle multiple SAN contexts on the fly. Lu *et al.* work on layered costmaps is an approach that closely relates to the goals of USAN [14]. However, this approach does not include a method to sense a context autonomously; hence, costmaps associated with a specific context cannot be selected automatically. In a similar work [3], researchers used a non-linear multi-objective optimization approach to socially-aware navigation that enabled socially appropriate navigation behaviors in multiple scenarios such as *hallways*, *art galleries*, *O-formations*, and *standing in a line*. In follow-up work [1], an image classification based context classifier (for *hallways* and *art galleries*) along with a based-based classifier using linear SVM (for *O-formations* and *standing in line*) enabled autonomous selection of cardinal objectives that matter most for an autonomously sensed context thereby enabled the optimization-based planner to work in multiple contexts.

Inverse Reinforcement Learning (IRL)-based approaches [10, 11, 16, 18] are promising in a single context and can be trained to handle multi-context SAN but will require a lot of human training data for each context. Even though the work on multi-context socially-aware navigation [1] generates social trajectories for an autonomously sensed context, it is constrained and does not maintain a knowledge base that can scale.

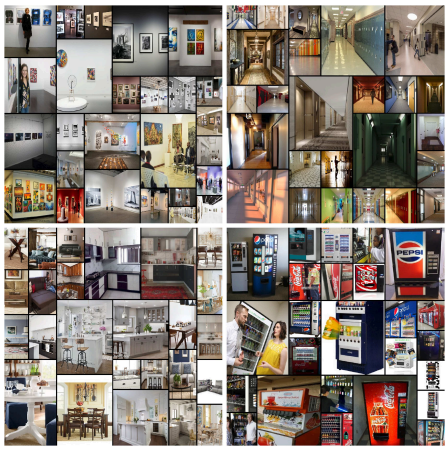

Figure 1: A sample of images from the internet that constitute images of hallways, artwork, vending machines, and other categories used for training our model.

2.2 Knowledge Base For Robot Navigation

There has been a growing interest in the knowledge-based methods and their application in robotics, such as socially aware navigation [13]. Semantic awareness has presented new frontiers in robot navigation, enabling more powerful tools for abstraction in representing information [7]. A location-based mobile service was developed and evaluated to study an indoor navigation service [22], which helped people to navigate around with physical difficulties. This work uses navigation context to enhance navigation behavior similar to our work, but our application is autonomous robot navigation instead of an online service for people with disabilities.

In another work [20], authors propose a knowledge engine that learns and shares knowledge representations for robots to complete various responsibilities. This work presented system structure and how it supports different tools for users and robots to interact with the knowledge engine. The authors extensively discussed the role and need of such knowledge engines in the real-world application in three major research areas: training natural language, perception, and planning, which are all critical in many robotic tasks.

3 APPROACH

3.1 Context Classification

3.1.1 Dataset. We trained a CNN model to distinguish between four contexts (classes), art galleries, hallway, vending machines, and others (anything that is not a hallway, art gallery, or vending machine - we utilized images of kitchens, living rooms, and dining rooms). We collected a total of 4773 images from the internet, as shown in Figure 1 and split them into training (.75), validation data (.25) and 400 test images. The images collected were all in color, resized to 256x256, and normalized before feeding to the network. Data augmentation was incorporated to ensure model generalization as the dataset is small. Augmentation includes image manipulations like zoom, shear, a shift in width, a shift in height and horizontal and vertical flip. Apart from the training, validation, and test data, we also collected real-world data at a mid-sized university in the United States to further test the model.

3.1.2 Model. Our approach to a context classifier is a CNN architecture that resembles VGGnet [6] but with a shallow depth (only eight convolution layers, three max-pooling layers, and four fully-connected layers). The CNN takes a 3-channel color image as input and outputs a probability that the image belongs to one of the four classes. The proposed CNN model consists of 8 convolution layers, each with 32 filters, a kernel size of 3, a stride of 1x1, same padding, and ReLU activation. There are three max-pooling layers with a pool size of 2x2 to downsample between layers 2-3, 5-6, 8-9. The network also includes dropout regularization with every max-pooling layer and between layers 9 and 10 (between the first two fully connected layers). All the fully connected layers use ReLU activation except for the last layer, which uses soft-max activation to make the predictions.

3.2 Object/Person Detection

The input images for context classification were also used with YOLOv3 to perform object detection and localization. The 'You Only Look Once' v3 (YOLOv3) method is among the most broadly used deep learning-based object detection approaches [23]. We need to detect objects and persons because understanding context alone is not enough to extract related rules as the objects within the context and the interactions between them play a vital role. For example, in

an empty gallery, the agent does not observe activity space unless a person and objects like artwork are detected. Similarly, in a hallway context, general rules to navigate the right side of the hallway applies, but the strategy can be different when a person is also in the scene as the robot should also account for the human's personal space. In other words, more specific rules can be added to a context on top of general contextual social rules by detecting objects and people in the environment (see Figure 2).

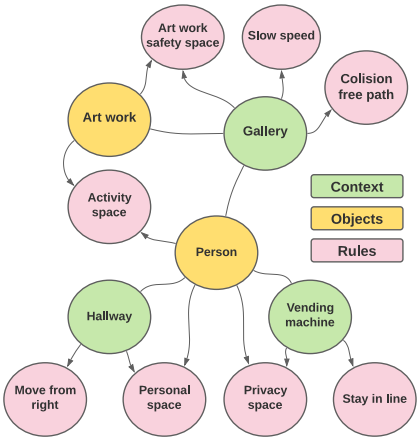

Figure 2: A social navigation ontology illustration, showing the relationship between context, objects, and rules.

3.3 Knowledge Base Representation

To share a common understanding of the information of the environment, we developed a social navigation ontology. An ontology is a way of describing knowledge as a collection of concepts in a domain and the relationships among them. An ontology can formalize high-level representations of knowledge of various concepts [9]. The authors use OWL language to build and expand a knowledge graph of concepts and relationships between them. We used context, objects, and rules associated with them to form an ontology in our approach. Relationships between contexts, objects are represented as a knowledge graph. A sample of related rules based on the observed context and detected objects can be seen in Table 1.

Environment	Social Rules
Gallery	Do not get too close to the artwork Respect peoples' personal space Respect activity zones if any
Hallway	Stay on the right Respect peoples' personal space
Vending Machine	Respect peoples' privacy Wait in line

Table 1: Sample social rules for specific environments.

3.4 Extracting Social Rules

With the output label from the context classifier and the objects detected, we query the knowledge base to extract applicable social rules associated with the context, given the objects within the detected context, i.e., the relationships between the objects and the context are used to get the associated social rules. We used the SPARQL language, a protocol using RDF query language, which is a semantic query language for databases to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

4 RESULTS

In this section first we present results of each component of our method and then the outcome of the whole package. For our context classifier, the model was trained on the training set and validated on the validation set over 500 epochs. Our model achieved a 96.44% training accuracy and 94.33% accuracy on validation data. The model generalized to real-world images (collected on campus) that it has not seen; the accuracy for an art gallery (15 samples), hallway (33 samples), and vending machine (12 samples) categories are 93.33%, 100.0%, and 91.66%, respectively. Table 2 shows performance on the real-world data.

Class	Precision	Recall	F1-Score
Art Gallery	1.0	0.93	0.97
Hallway	0.97	1.0	0.99
Vending Machine	1.0	0.92	0.96

Table 2: Performance of the CNN based context classifier on real-world images collected on campus.

For object detection, YOLO-v3 outputs detected objects along with their confidence score; however, in this work, we used the confidence score to filter the noise by thresholding at 50%, and the label data is used with the query system. These outcomes are used as text information for making SPARQL queries to extract social rules.

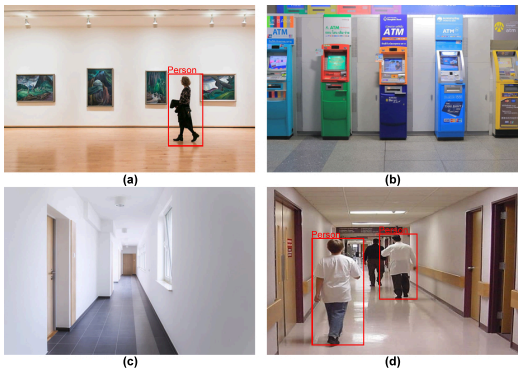

Figure 3: In a sample of images from the internet used to test the approach, we found "gallery, hallway, and vending machine" as context. "Person" was also detected in these contexts, which made differences in executed rules.

Figure 3 presents sample results on images of three categories: an art gallery, hallway, and the vending machine. Image (a) is an art gallery context with a person in it so, our system extracted *do not get too close to the artwork*, *respect peoples' personal space*, and *respect activity zones* as social rules for this situation. Similarly, image (b) is a vending machine (ATMs) situation where the robot would apply general rules like *wait in line* (even when others are not waiting, in which case it is called approach behavior); however, when people are using the ATMs, other specific social rules like *respect privacy* and *respect peoples' personal space* should be considered. To illustrate this, consider images (c) and (d) both are hallway context; however, one of the images is just a hallway without any people; in this case, the general rule of *stay on the right side* is applicable. In the other case, a hallway context with people in it, specific rule of *respecting peoples' personal space* is also extracted by our system along with the general rule of *stay on the right*, as shown in Table 3.

Context	Social Rules
Empty hallway (c)	Stay on the right
Person in hallway (d)	Stay on the right
	Respect peoples' personal space

Table 3: Results of extracted social rules in various environments. Most importantly, this table shows the extraction of general and specific social rules depending on the context and the objects detected in the contexts.

5 DISCUSSION

This paper presented the ideation and preliminary results of our approach towards deep reasoning of social rules for social navigation applications. Our results show that using deep learning methods such as image classification and object detection can query a knowledge base to extract both general and specific social rules pertaining to the detected context and the observed objects within the sensed context. We also showed how specific rules could apply depending on the objects detected in a context. Our preliminary results show evidence that our approach could be used as an objective selection mechanism for a unified socially-aware navigation system.

5.1 Limitations/Future Work

Some of our work's limitations are that we used a few contexts, and therefore, the knowledge base is small. However, our ongoing efforts include building a broader knowledge base using MIT Indoor Scenes dataset [17]. Future work will augment our system to autonomously build a knowledge graph by learning the relationships between contexts and objects within the context. Once we have this high-level knowledge graph system, we will integrate it with an optimization-based social navigation planner [3, 8]. In the aftermath of the COVID-19 pandemic, we plan to develop the system and validate our proposed method on a real-world robot.

ACKNOWLEDGMENTS

The authors would also like to acknowledge the financial support of this work by the National Science Foundation (NSF, #IIS-1719027).

REFERENCES

[1] Santosh Balajee Banisetty. 2020. *Multi-Context Socially-Aware Navigation Using Non-Linear Optimization.* Ph.D. Dissertation. University of Nevada, Reno.
[2] Santosh Balajee Banisetty and David Feil-Seifer. 2018. Towards a unified planner for socially-aware navigation. *arXiv preprint arXiv:1810.00966* (2018).
[3] Santosh Balajee Banisetty, Scott Forer, Logan Yliniemi, Monica Nicolescu, and David Feil-Seifer. 2019. Socially-aware navigation: A non-linear multi-objective optimization approach. *arXiv preprint arXiv:1911.04037* (2019).
[4] Santosh Balajee Banisetty, Meera Sebastian, and David Feil-Seifer. 2016. Socially-Aware Navigation: Action Discrimination to Select Appropriate Behavior.
[5] Aniket Bera, Tanmay Randhavane, and Dinesh Manocha. 2019. The Emotionally Intelligent Robot: Improving Socially-aware Human Prediction in Crowded Environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.*
[6] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316* (2016).
[7] Jonathan Crespo, Jose Carlos Castillo, Oscar Martinez Mozos, and Ramon Barber. 2020. Semantic Information for Robot Navigation: A Survey. *Applied Sciences* 10, 2 (Jan 2020), 497. https://doi.org/10.3390/app10020497
[8] Scott Forer, Santosh Balajee Banisetty, Logan Yliniemi, Monica Nicolescu, and David Feil-Seifer. 2018. Socially-aware navigation using non-linear multi-objective optimization. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, 1–9.
[9] Thomas R. Gruber. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies* 43, 5 (1995), 907 – 928. https://doi.org/10.1006/ijhc.1995.1081
[10] Beomjoon Kim and Joelle Pineau. 2016. Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics* 8, 1 (2016), 51–66.
[11] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. 2016. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research* 35, 11 (2016), 1289–1307.
[12] Thibault Kruse, Amit Kumar Pandey, Rachid Alami, and Alexandra Kirsch. 2013. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems* 61, 12 (2013), 1726–1743.
[13] Séverin Lemaignan, Mathieu Warnier, E. Akin Sisbot, Aurélie Clodic, and Rachid Alami. 2017. Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence* 247 (2017), 45 – 69. https://doi.org/10.1016/j.artint.2016.07.002 Special Issue on AI and Robotics.
[14] David V Lu, Dave Hershberger, and William D Smart. 2014. Layered costmaps for context-sensitive navigation. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on.* IEEE, 709–715.
[15] Eitan Marder-Eppstein, Eric Berger, Tully Foote, Brian Gerkey, and Kurt Konolige. 2010. The office marathon: Robust navigation in an indoor office environment. In *2010 IEEE international conference on robotics and automation.* IEEE, 300–307.
[16] Billy Okal and Kai O Arras. 2016. Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning. In *2016 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 2889–2895.
[17] Ariadna Quattoni and Antonio Torralba. 2009. Recognizing Indoor Scenes. *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 413–420. https://doi.org/10.1109/CVPR.2009.5206537
[18] Rafael Ramon-Vigo, Noe Perez-Higueras, Fernando Caballero, and Luis Merino. 2014. Transferring human navigation behaviors into a robot local planner. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on.* IEEE, 774–779.
[19] Jorge Rios-Martinez, Anne Spalanzani, and Christian Laugier. 2011. Understanding human interaction for probabilistic autonomous navigation using Risk-RRT approach. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE, 2014–2019.
[20] Ashutosh Saxena, Ashesh Jain, Ozan Sener, Aditya Jami, Dipendra K. Misra, and Hema S. Koppula. 2015. RoboBrain: Large-Scale Knowledge Engine for Robots. arXiv:1412.0691 [cs.AI]
[21] Meera Sebastian, Santosh Balajee Banisetty, and David Feil-Seifer. 2017. Socially-aware navigation planner using models of human-human interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* IEEE, 405–410.
[22] Vassileios Tsetsos, Christos Anagnostopoulos, Panagiotis Kikiras, and Stathes Hadjiefthymiades. 2006. Semantically enriched navigation for indoor environments. *IJWGS* 2 (01 2006), 453–478. https://doi.org/10.1504/IJWGS.2006.011714
[23] Liquan Zhao and Shuaiyang Li. 2020. Object Detection Algorithm Based on Improved YOLOv3. *Electronics* 9 (03 2020), 537. https://doi.org/10.3390/electronics9030537