Challenges in Policy Learning for British Sign Language Users in an Example of Navigation-Based Task

ABSTRACT
In the light of natural human interactions, this paper identifies the lack of multimodality in current policy learning approaches for human-robot/human-computer interactions. Larger part of policies, trained for the navigation-based tasks are trained on using either textual or natural language inputs, which are unimodal. This paper suggests and formalises British Sign Language as an input to the offline example-based policy learning algorithm. The policy for how to follow directional instructions from a human to a machine is learned through reinforcement. Currently, the training environment is being set-up and the datasets that will be used for the offline policy learning are being analysed. The novelty of this project is the use of sign languages in policy learning for navigation-based task that exceeds existing research in computing that uses sign languages as well as advances the research on multimodal human-robot/human-computer interactions.

ACM Reference format:
DOI: 10.1145/nmmnn.nnnnnn

1 INTRODUCTION
The field of Human-Robot Interaction is envisioning robots interacting with the general population in various settings, such as household, shops, or museums. This is a significant advancement since robots mainly operate in factories with limited functionality. Such advancements will require complex behaviours on the robots’ part. The challenge is to enumerate all possible robot behaviours that will be required in such complicated environments [13]. Instead of manually describing all the behaviours, robots can learn to approximate certain behaviours by responding correctly to perceived cues. This approximation can be achieved with policy learning.

This method is already being successfully applied in various domains, such as Google’s Gmail Smart Reply that suggests automated replies to your email [16] or Ubuntu’s technical support as well Amazon’s customer service that answers your questions [19, 28], with all these domains being trained on human-human interaction examples. This can be viewed as an attempt to teach machines grounded language based on the context. These approaches make use of the free-form text-based, natural language-based, or video-based interaction for learning policies. Despite the fact that the human interaction is multimodal as we tend to gesture and create facial expressions to support and augment the speech signal, current approaches to policy learning focus on one single modality from the interactions, thus, simplifying the information that is being communicated.

This paper proposes to use British Sign Language to learn policy in a navigation-based task. Generally, sign languages are multimodal in themselves as they use different body parts (manual features, such as hands, fingers, posture) as well as non-manual features (such as facial expressions) to communicate information as an individual modality [9]. The use of sign languages should not be seen as only an application for a specific target group, but rather as the superset of other modalities for human-robot/human-computer interactions, such as speech, gestures, or gaze.

The paper presents the desired aim of the project and gives an overview and the development of the ideas behind policy learning in the navigation domain, mentioning a number of general-purpose approaches and their applications. Section 2.2 introduces sign languages, their collection, and use in computing. The proposal of this paper for initial proof of concept is formalised as a reinforcement learning problem. Two datasets that will be used for this project are described and, finally, a set of the expected challenges is listed for a policy learning algorithm that is necessary for the long-term human-robot/human-computer interaction.

2 RELATED WORK
Learned policies are usually applied in specific contexts, some of which are navigation [21, 35] or dialogue [29]. In this paper, we are interested in spatial route directions, which is a special case of instructions that specifies how to get from A to B and the understanding of these instructions depends heavily on spatial context.

2.1 Policy Learning in Navigation-Based Tasks
Approaches to policy learning in the field of human-robot or human-computer interaction can be divided into two streams. Tradition-als rely on past examples and foresee that future examples will be similar to that of the past. This approach performs direct policy estimation given interaction examples. Although, there is usually some similarity between the future and the past events, this similarity is not guaranteed during the live stochastic interactions, which, consequently, leads to progressive approaches that do not assume the similarity between the past and the future knowledge. These approaches perform free-form language parsing into structured form from which policy is constructed. Both streams align input with the action selection in different ways. With parsing, the alignment is explicit, whereas with policy estimation, alignment is learned by selecting correct state-action mapping. This section will follow with specific examples of the two streams in navigation-based
tasks, touching on appealing general purpose approaches that are currently being used in dialogue management. These specific examples have concentrated mostly on policy learning for unimodal interaction with the main modality being either speech, text, or vision.

2.1.1 Reinforcement Learning Approaches. Vogel A. and Jurafsky D. as well as Branavan S.R.K et al. made use of reinforcement learning for mapping instructions to actions in navigational map task or for mapping manual text onto actions [6, 35]. Dipendra K. Misra used the same method to train a neural network with reward shaping for mapping visual observations and textual descriptions onto actions in a simulated environment [22]. In all these cases, the constructed model did not include semantic or syntactic knowledge, nor did it require linguistic annotations for model training.

2.1.2 Probabilistic Approaches. Probabilistic approaches to model training have used available linguistic information to train underlying linguistic parsers to perform inference by searching for maximum of joint distributions among all possible actions in navigation scenarios. Kollar T. et al. pursued the traditional approach to direction following with directions being presented in natural language together with the environment and the system inferring the most probable path in the environment [17]. In the case of Duvallet F. et al., the map is not fully observable and the system has to obtain knowledge through exploration, update the prior, and infer with more complete knowledge [10]. Andreas J. and Klein D. use predictors to generate trajectories from text in different contexts [3].

2.1.3 Supervised Learning Approaches. Similar to the reinforcement learning approaches others employed supervised learning for the same purpose without the need for linguistic annotations. These methods can be used as probabilistic models for mapping from route descriptions to actions. Matuszek C. et al. learn a parser that translates natural language into concise path description [20]. Andreas J. and Klein D. show how supervised learning can be used in different contexts to align language with the environment at the word level [4]. Shimizu N. followed traditional approach of learning model that is capable of following directions in natural language on a given map [30]. Mei H. shows how navigational instructions can be mapped to actions sequences without annotations or parsers using supervised generative sequence-to-sequence (Seq2Seq) model with long short-term memory (LSTM) [21]. In fact, Seq2Seq models are becoming more popular than goal-oriented models, because Seq2Seq models are general-purpose and are used for various components of standard dialogue systems. The main disadvantage with Seq2Seq models is that they require large amounts of data for training [29].

2.1.4 Summary. All above mentioned studies have learned policies offline and the majority have had prior knowledge of the environment, described in terms of, for example, landmarks, their labels and geometrical relationships as well as the specific domain-dependent interaction examples. Goldwasser D. and Roth D. acknowledge the drawback of the example-based learning and propose an instruction interpretation approach, which grounds interactions in the domain, whereas action inference happens at the grounded concepts and entities level [12]. Other works have followed this approach, focusing on machine translation methods that were mentioned previously, such as [3, 10, 17].

The biggest disadvantage of direct policy learning for navigation-based task is that the policy becomes restricted to certain actions that were deemed appropriate to that state during policy learning. This means that the method is restricted to learning examples, while parsing approaches can cope with possible combinations and permutations of inputs that were seen during training.

Recently, Weston J. suggested restructuring datasets that are being used for example-based learning, making learners learn from feedback that contains both reward and additional information for learning [36]. The work even suggested types of possible training strategies, extending existing learning approaches. This is a significant step towards real-time policy learning with user interactive feedback.

2.2 Sign Languages in Computing

As mentioned in Section 1, sign languages are multimodal by their nature. Signers simultaneously use their posture, arms, hands, fingers, facial expressions, and gaze to communicate information. Different body parts operate in harmony during signing, enriching the overall meaning as much as supporting or reinforcing each other.

2.2.1 Sign Language Analysis. The majority of research on sign languages in computing has focused on the task of recognising the sign languages, mostly concentrating on estimating configuration of hands and fingers from video data. Although limited to single modality, results have shown that it is possible to recognise signer-independent signs the models were trained on [9, 24, 25, 38].

2.2.2 Sign Language Synthesis. Recently, the aim has switched from analysis of the sign languages to their synthesis for avatar-based interpretation of spoken or written resources for the purposes of teaching sign languages or fostering interaction between deaf and hearing people [1, 7, 14]. With the recent popularity of robots and the long-term aim of robots thriving in household environments, some work has been done on teaching robots how to sign [26, 34, 39]. This can have a major positive impact on deaf community once the results become acceptable.

2.2.3 Capturing Sign Languages. Most approaches to analyse sign languages use vision for information extraction, because vision is relatively cheap and widely-available sensor. However, since signing is happening in 3D, the depth information is being lost when recording with a single video camera. Such approaches are, therefore, insufficient for capturing all the signing information. To overcome this limitation, other approaches use either more than one camera or such devices as Microsoft Kinect [37], motion capture systems [15], or Leap motion sensor [8] that capture depth information by emitting and capturing reflected light. Another option is to use dataglove to recognise signs [18]. Although, these methods improve recognition precision, they obstruct the signing process as signing involves contact between the hands and the body.
2.2.4 Sign Languages for Policy Learning. Despite all the perception challenges associated with sign languages analysis, this paper argues that sign languages are a suitable modality for policy learning in order to achieve human-like multimodal interaction. The use of sign languages in policy learning for navigation-based task will transcend the existing research in computing that uses British Sign Language as well as advance the research on multimodal interaction.

3 PROPOSED APPROACH

Section 2.1 identified existing division in domain-specific policy learning research. The first being example-driven policy learning and the second being linguistically analysed (parsed) language, represented as symbols that are later grounded in the environment with the corresponding responses. In the case of the sign languages as the input, linguistic parsing is a challenge on its own. The task would involve signer independent segmentation, possibly with the eponthesis modelling, fusion of multimodal data, use of linguistics theories to recognize individual phonemes and signs with generalisation to complex datasets [9, 25]. Therefore, direct policy learning has been chosen as a reasonable method for the problem, acknowledging the complexity of syntactic parsing for sign languages.

We propose an approach that extends the method for following navigational directions, applied by Vogel A. and Jurafsky D. to a map task [35] and a method, used by Branavan S.R.K. et al. [6] to follow textual instructions in operating system. Our approach to policy learning is a proof of concept for learning how to respond to navigational instructions, presented in annotated British Sign Language.

3.1 Formulation

The aim of this project is to associate navigation instructions, given in unrestricted British Sign Language SL(t) = \{m_1(t), ..., m_n(t)\}, with sequence of actions A = \{a_1, a_2, ..., a_n\} executable on a virtual map from states S = \{s_1, s_2, ..., s_n\}. SL(t) is a system of individual modalities m_1, ..., m_n that are executed in parallel during signing, such as body posture, hand shape and orientation, facial expression, etc. All modalities can be expressed as separate functions of non-linear combination of parameters, such as 3D information, orientation, velocity, acceleration at a specific time step t. This is in accordance with lexical features for describing a sign, identified by Stokoe et al. and considered to be universal: location of the sign in space, hand shape, body posture, and non-manual features [32].

3.1.1 Inputs. The first input to the algorithm is a set of signing instructions U ∈ SL from instruction giver to instruction follower. u ∈ U is a sequence (u_1, ..., u_n) with u_i being a single phoneme (called chereme in the past), surrounded by epontheses. By phoneme we refer to an isolated sign that may involve movement and may not necessarily have the final intended meaning, but may belong to a more complete sign that would consist of more than one phoneme, which will have the final intended meaning. By epontheses we refer to everything that is happening between phonemes, which commonly are transitions between the individual phonemes. One example of an eponthesis may be hands movement into another configuration during fingerspelling.

3.1.2 Environment. Semantics of the environment are expressed in terms of landmarks L = \{l_1, ..., l_n\} and their relationships, such as distance d(l, l'), ∀l ∈ L and the navigation route order, provided by the instruction giver during the training. Every landmark is expressed as a phoneme in British Sign Language. As is done in Mocialov et al. for gestures, all the landmarks are encoded in corresponding matrices M(l) [23]. Matrices contain similarity measures between extracted kinematic features (manual sign language features) and facial features (non-manual sign language features) from every timestep during SL(t). This encoding enforces concise and uniform-length representations of phonemes and supports comparison of phonemes.

3.1.3 States. State s ∈ S is expressed as s = (u, l), where u is a set of phonemes (an utterance) from instruction giver and l is the landmark encoding M(l).

3.1.4 Actions. Action a ∈ A indicates the destination landmark l' on the map from the current state s, expressed as the landmark encoding M(l'). l' can be empty, which would mean that the destination landmark was either not specified by the instruction giver or it was not detected by the M comparisons. Same as in Vogel A. and Jurafsky D. [35], we assume that at most one instruction is given per utterance u.

3.1.5 Transition Function. Transition function T(s, a) = s' presents a new state s' when executing action a in state s. This transition results in the next state and a scalar reward, given by the reward function.

3.1.6 Reward Function. Reward function R(s, a) presents a scalar reward after an action a is executed in a state s. The reward represents the immediate utility of executing that action in that state.

3.1.7 Feature and Parameter Vectors. Vogel A. and Jurafsky D. [35] use feature vectors \( \phi \) for every state that associate information from the environment with the utterance from the instruction giver. This feature vector influences updates of the parameter vector \( \theta \) that, in turn, influences action selection for a given state. In our case, the feature vector would play a valuable role, because we want to associate phoneme u_i with the map. This means that the segmentation of signs in U is necessary. In addition, the feature vector in the original work provides a set of spatial terms, such as right, left, up, down, etc. The set of spatial terms is used to learn positioning in respect to the landmarks (allocentric feature). This is advanced functionality that we do not consider in our implementation as we are focusing on the proof of concept of navigational instructions, presented in British Sign Language.

Our feature vector \( \phi \) has a binary feature \( \phi_1 \) that indicates whether any of the signed phonemes correspond to the name of the target landmark M(u_i) ≈ M(l'). In case if \( \phi_1 \) is false, this means that the signer did not mention any target landmarks during the signing sentence u or it was not detected, in which case the second binary feature \( \phi_2 \) becomes true and the agent should not perform any actions. The second feature \( \phi_2 \) indicates whether there should be no action. Our feature vector is simpler than the one from Vogel A. and Jurafsky D. [35], because we want to begin by testing the
feasibility of this approach. The parameter vector $\theta$ has the same size as the $\phi$ vector and the utility of taking action $a$ in state $s$ is determined by $Q(s, a) = \theta \cdot \phi$. There is one parameter vector $\theta$ for every state-action pair and it is updated every time the same action is chosen in the same state, whereas the action selection in a state depends on the exploration strategy chosen. The advantage of using feature vector is that it can be extended with additional features in case if we would like our policy support additional functionality.

3.2 Datasets

From the Section 2.1 we see that the common evaluation task for learned navigation-based policies is The HCRC Map Task [2]. Unfortunately, as pointed out by the Matthies S. et al., an initial plan to use The HCRC Map Task revealed practical problems with sign languages users. The signers were using the signing space to refer to landmarks, which reduced intended interaction between participants and simplified the task. In addition, since sign languages are perceived visually, signers were losing eye contact, referring to the map on which they had to draw. Yet another issue is since The HCRC Map Task requires drawing, participants had to hold a pen, which in turn interfered with their signing [31].

To our knowledge, there are no public datasets that would contain 3D annotated free-form navigational directions signing data due to the specificity of the data and limited popularity of the sign languages. However, there are transcribed 2D video recordings of sign language users explaining directions on a map in the DGS Corpus Project [11] that were used instead of The HCRC Map Task or in Deaf-Deaf Map Task recordings [27]. As it was already mentioned in Section 2.2 on sign languages, the data should be in 3D format to convey the complete knowledge of the signs. In the case of the DGS Corpus Project, the signers are recorded with multiple cameras from different angles, which makes it possible to reconstruct 3D information with triangulation method. Currently, we are analysing mentioned datasets, while setting up the training environment.

Alternatively, in the case if automatic 3D inference will not be feasible, the 2D to 3D mapping could be either crowdsourced or reconstructed using models that were trained on a large set of human postures [33].

4 CHALLENGES

More generally than the domain-specific policy learning, the learned policy for multimodal interaction should comply with a set of requirements as follows:

1) Support concept drift: learned policy should be trained to expect change in intention, because human interaction dynamics are non-linear. What is considered as appropriate or correct and is reinforced at one instance may be wrong later.

2) Adapt to stochastic environment: real environment is unpredictable that is influenced not only by concept drifts induced by humans, but also by the environment itself.

3) Filter noisy perception: any fluctuation in the environment causes noise in perception, which in turn distorts acquired data.

4) Explore in case of partial observability: real environment is too big and too complex to be known in advance. Learner should explore the environment and update its prior beliefs.

5) Extend existing knowledge: existing knowledge might be insufficient for the desired behaviour or an explicit command could be initiated during the interaction to extend the existing knowledge.

6) Shared responses: a set of states can have same responses.

As a proof of concept, the proposed approach in Section 3 aims to satisfy requirements 3 and 6. The long term aim of the project is to tackle as many requirements as possible, beginning with the requirements 1 and 5 after a successful proof of the concept.

5 CONCLUSIONS

This paper describes a long-term project to advance multi-modal interaction between humans and computers/robots. This paper proposes to use British Sign Language as a communication medium between humans and robots. The rationale behind the use of the sign languages is the availability of many modalities, present in them. Although, the requirement for the policy learning for the purpose of human-computer/human-robot interaction is quite ambitious, the project proposes to use domain-specific navigation-based task and reinforcement learning for policy learning as a proof of concept that the British Sign Language can be used to communicate navigational information and the learner can learn to respond appropriately to the instructions. Preliminary work had begun with setting the learning environment and analysis of candidate datasets for the policy learning.

REFERENCES

Hee-Deok Yang. 2014. Sign language recognition with the kinect sensor based
Nobuyuki Shimizu and Andrew R Haas. Learning to Follow Navigational Route
Cynthia Matuszek, Dieter Fox, and Karl Koscher. 2010. Following directions
Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins,
Tommi Jantunen, Birgitta Burger, Danny De Weerdt, Iria Seirola, and Tuja
Kollar, Stefanie Tellex, Deb Roy, and Nicholas Roy. 2010. Toward understanding
natural language directions. In Human-Robot Interaction (HRI), 2010 5th
ACM/IEEE International Conference on. IEEE, 259–266.
data-glove for sign language recognition.
Nobuyuki Shimizu and Andrew R Haas. Learning to Follow Navigational Route Instructions.