Team Intent Understanding through Latent Representation Learning for Underground Search and Rescue

Fei Han and Hao Zhang

Abstract— **Team intent awareness plays an important role in human-robot teaming applications. The autonomous robots need to infer the team intent accurately before they can collaborate with human peers in the challenging underground search and rescue tasks. In this paper, we introduce a novel latent representation learning method for team intent awareness based on both individual human behaviors and the relationship between human peers in the team. It is formulated as a sparse learning problem, where each individual behavior is represented by the sparse latent representation. In addition, the teammate relationship is modeled by the Laplacian regularization to enforce the team structure. Our approach enables to model both individual human behaviors and the relationship between humans, which has significant capability to infer team intents. Real-world experiments in the underground search and rescue scenario demonstrate the superior performance of our proposed approach.**

I. Introduction

Underground search and rescue efforts are always carried out by groups of rescuers working together as a team. Accordingly, maintaining team safety is a shared responsibility, and the robots need to be aware of the human teammates to ensure team situational awareness. In addition, in a mixed human-robot rescue team, the human rescuers are usually professionally trained for rescue missions, and are too busy to elaborate their intentions to robots (sometimes impossible when facing a hazard). Robot teammates need to intelligently interpret the activities and intentions of human peers, to provide direct support to rescuers or assist with the ongoing rescue task without burdening the human rescuers. In the underground search and rescue team, each individual rescuer can perform a different activity; but together as a team, the rescuers work toward the same goal with the same intention. Our work investigates the team intent awareness problem in the challenging underground search and rescue scenario.

There are robotic perception literature addressing the task of single-person activity reasoning and recognition of a group of people with similar activities (e.g., waiting in line, fighting with each other, etc.) [1], it was not well studied to simultaneously reason about the activities of multiple teammates and their shared intention, especially in unstructured environments with reduced visibility.

In this research, we introduce a novel approach for team intent awareness based on both individual human behaviors and the relationship between human peers in the team. It is formulated as a sparse learning problem, where each individual behavior is represented by the sparse latent representation. In addition, we also encode the team structure by modeling the relationship of human peers within the team using the Laplacian regularization. Our approach enables to model both individual human behaviors and relationship between humans, which has significant capability to infer the shared team intents.

Fei Han and Hao Zhang are with the Department of Computer Science, Colorado School of Mines, 1500 Illinois Street, Golden, CO 80401, USA fhan@mines.edu, hzhang@mines.edu

This research was supported in part by Army Research Office under Grant W911NF-17-1-0447.

II. The Aproach

Given m human teammates within a team, each observation (i.e. data instance) of the entire team is represented as $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_m] \in \mathbb{R}^{d \times m}$, where $\mathbf{x}_i \in \mathbb{R}^d$ represents the feature vector of the observation from the i-th individual. We represent the team intent label vector of the observation $\mathbf{X}$ as $\mathbf{y} \in \mathbb{Z}^{c_1}$: if $\mathbf{X}$ belongs to the l-th team intent category, the l-the element within the intent label vector $\mathbf{y}$ satisfies $\mathbf{y}^{(l)} = 1$; otherwise $\mathbf{y}^{(l)} = 0$, where c_1 is the number of team intents. Since team intent is an abstract concept that is collectively reflected by the activities of all individual human members in the team, to model the behavioral hierarchy of human teaming, we propose to introduce a latent variable vector $\mathbf{z}_i \in \mathbb{Z}^{c_2}$ to encode the individual activities of the i-th human team member: if $\mathbf{x}_i$ belongs to the l-th individual activity category, the l-th element of the vector $\mathbf{z}$ satisfies $\mathbf{z}^{(l)} = 1$; otherwise $\mathbf{z}^{(l)} = 0$, where c_2 is the number of individual activities. Then, given a set of n training data instances $\left\{(\mathbf{X}^j, \mathbf{y}^j)\right\}_{j=1}^{n}$, we define the individual activity matrix for each data instance j as $\mathbf{Z}^j = [\mathbf{z}_1, \cdots, \mathbf{z}_m] \in \mathbb{R}^{c_2 \times m}$. We formulate the problem of team intent awareness as an optimization problem by minimizing the following cost function:

$$\min_{\mathbf{W},\mathbf{U},\mathbf{Z},\mathbf{b},\mathbf{p}} \sum_{j=1}^{n} \Big(\alpha_j \left\| \mathbf{W}^\top \mathbf{X}^j + \mathbf{b}\mathbf{1}_m^\top - \mathbf{Z}^j \right\|_F^2 + \left\| \mathbf{U}^\top \mathbf{Z}^j + \mathbf{p}\mathbf{1}_m^\top - \mathbf{y}^j \mathbf{1}_m^\top \right\|_F^2 \Big) \quad (1)$$

where $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_{c_1}] \in \mathbb{R}^{c_2 \times c_1}$ and $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_{c_2}] \in \mathbb{R}^{d \times c_2}$ are coefficient matrices (parameters to learn) for team intent and individual activity estimation, $\alpha_j, j = 1, \cdots, n$ and $\gamma_l (l = 1, 2, 3)$ are trade-off hyper-parameters, $\mathcal{R}(\cdot)$ is the regularization term with respect to $\mathbf{Z}$, which encodes the peer-to-peer characteristic of human teaming, $\mathbf{b} \in \mathbb{R}^{c_2 \times 1}$ and $\mathbf{p} \in \mathbb{R}^{c_1 \times 1}$ are intercept vectors, and $\mathbf{1}_m \in \mathbb{R}^{m \times 1}$ is the constant vector of all 1's.

To encode the peer-to-peer characteristic of human teaming, we define the regularization for individual activity estimation as $\sum_{k=c_2}^{n} \sum_{i=1}^{m} \|\mathbf{z}_i^k\|$, which is an entrywise ℓ_1-norm over $\mathbf{Z}^j$, encoding that all humans are equally important (i.e., peers), while enforcing the sparsity of individual activities (i.e., our formulation is capable of modeling concurrent activities, which assumes a human teammate only simultaneously performs a small number of such activities).

To further enhance the modeling capability of our approach beyond modeling the team behavioral hierarchy, we also propose to model the relationship of human peers within the team, by explicitly enforcing a Laplacian regularization over the latent individual activity label matrix $\mathbf{Z}$, which models the affinity structure of human members. Given m human peers in a team, we first construct a relationship matrix $\mathbf{R} \in \mathbb{Z}^{c_2 \times c_2}$ such that $\mathbf{R}_{\alpha\beta} = 1$ if and only if the α-th and β-th individual activities are the same or similar; otherwise, $\mathbf{R}_{\alpha\beta} = 0$. We intend to find an embedded subspace $\mathbf{Z}$ for individual activities such that the distances of pairwise peers' activities which are similar or exactly the same in the team are minimized. We can compute the Laplacian matrix $\mathbf{L}$ by $\mathbf{L} = \text{diag}(\mathbf{R1}) - \mathbf{R}$. Then, we introduce a Laplacian regularization to integrate the teammate relationship (encoded by $\mathbf{L}$) into the regularizer of $\mathbf{Z}$ as follows:

$$\begin{aligned}\mathcal{R}(\mathbf{Z}) &= \sum_{j=1}^{n}\left(\beta_1 \left\|\mathbf{Z}^j\right\|_1 + \beta_2 \sum_{\alpha,\beta=1,\cdots,c_2} \mathbf{R}_{\alpha\beta}\|\mathbf{z}_\alpha - \mathbf{z}_\beta\|_2^2\right) \\ &= \sum_{j=1}^{n}\left(\beta_1 \left\|\mathbf{Z}^j\right\|_1 + \beta_2 \mathbf{Tr}\left(\mathbf{Z}^{j\top}\mathbf{L}\mathbf{Z}^j\right)\right) \qquad (2)\end{aligned}$$

Besides that, we also introduce two additional regularization terms with respect to $\mathbf{W}$ and $\mathbf{U}$ to address the overfitting issue, which are defined as $\|\mathbf{W}\|_F^2$ and $\|\mathbf{U}\|_F^2$, respectively.

Therefore, the final team intent awareness problem is formulated as the following optimization problem with three regularization terms

$$\begin{aligned}\min_{\mathbf{W},\mathbf{U},\mathbf{Z},\mathbf{b},\mathbf{p}} &\sum_{j=1}^{n}\left(\alpha_j \left\|\mathbf{W}^\top\mathbf{X}^j + \mathbf{b}\mathbf{1}_m^\top - \mathbf{Z}^j\right\|_F^2\right. \\ &+ \left.\left\|\mathbf{U}^\top\mathbf{Z}^j + \mathbf{p}\mathbf{1}_m^\top - \mathbf{y}^j\mathbf{1}_m^\top\right\|_F^2\right) \\ &+ \gamma_1\|\mathbf{W}\|_F^2 + \gamma_2\|\mathbf{U}\|_F^2 + \gamma_3\mathcal{R}(\mathbf{Z}) \qquad (3)\end{aligned}$$

where $\mathcal{R}(\mathbf{Z})$ is defined in Eq. (2).

The solution of the optimization problem in Eq. (3) are the parameter matrix $\mathbf{W}, \mathbf{U}, \mathbf{b}, \mathbf{p}$, which describe the mapping from the observation of human team peers to the team intent.

After obtaining the optimal parameters from the training data, we can utilize the learned model to estimate individual activities of all human peers and predict the team intent in an online fashion. The team intent is obtained according to the following method:

$$\text{team intent} = \arg\max_{l} \mathbf{y}^o(l), l = 1, 2, \cdots, c_1 \qquad (4)$$

where $\mathbf{y}^o(l) = \frac{1}{m}\mathbf{U}^\top\mathbf{Z}^o\mathbf{1}_m + \mathbf{p}$, $\mathbf{Z}^o = \mathbf{X}^{o\top}\mathbf{W} + \mathbf{1}_m\mathbf{b}^\top$, $\mathbf{U}, \mathbf{W}$ are optimal coefficient matrices learned in the training process, and $\mathbf{X}^o$ is the query team observation.

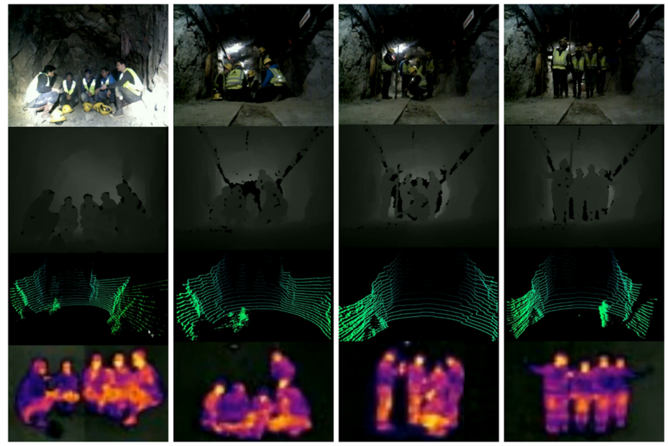

Fig. 1. Multimodal perception data captured in the dark side of the Eagle Mines by RGB-D camera, LiDAR, and thermal camera simultaneously. Each column shows different team intents. Left to right: donning, patient care, timbering, and traversing.

III. Experiments

We evaluated the proposed method under a real-world search and rescue scenario in underground mines. In this experiment, we focus on the team intent awareness application by robots in the challenging dark environments, where the Husky robot needs to infer the team intents by recognizing the behaviors performed by each team member. We define five team intent categories, including *donning, patient care, team stop, timbering, and traversing*. The perception data were captured using various sensors, including RGB-D camera, LiDAR, and thermal camera, by which we recorded color images, depth images, point clouds, and thermal images simultaneously. Fig 1 illustrates all these multimodal perception data recorded on the dark side of the Eagle Mine (the Colorado School of Mines Experimental Mine) when the team was performing donning, patient care, timbering, and traversing, respectively.

In this experiment, each team behavior was performed by different team configurations, by which subjects with different body scales and motion patterns were involved. Each team behavior was performed 20 times, where half were used for training and the remaining half testing. Ground truth was manually labeled and used to quantitatively evaluate the performance of our proposed approach. In this experiment, we applied our proposed team intent awareness approach using the recorded color image data. After extracting the features from the individual behaviors during training, we obtained the optimal weight matrices $\mathbf{W}$, $\mathbf{U}$, $\mathbf{b}$, and $\mathbf{p}$ using the training input data. Then, Eq. (4) was applied to infer the team intent in the execution phase.

After applying our proposed approach, we achieve an average accuracy of 97.6%, which demonstrates the superior performance of our proposed approach in team intent awareness by observing each individual human behavior.

References

[1] T. Lan, Y. Wang, W. Yang, S. N. Robinovitch, and G. Mori, "Discriminative latent models for recognizing contextual group activities," *TPAMI*, vol. 34, no. 8, pp. 1549–1562, 2012.