2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Multi-label Long Short-Term Memory for construction vehicle activity recognition with imbalanced supervision

Haruka Abe[1], Takuya Hino[2]
Motohide Sugihara[2], Hiroki Ikeya[2], and Masamichi Shimosaka[1]

***Abstract*— Sensor-based activity recognition for construction vehicles is useful for evaluating the skills of the operator, measuring work efficiency, and many other use cases. Therefore, many researches have explored robust activity-recognition models. However, it remains a challenge to apply the model to many construction sites because of the imbalance of the dataset. While it is natural to employ multi-label representation on imbalanced data with a large number of activity categories, multi-label robust classification for activity recognition has yet to be resolved because of the nature of the time-series property.**

In this work, we propose a novel multi-label long short-term memory (LSTM) model, which is effective for the sequence multi-labeling problem. The proposed model has connections to the temporal direction and attribute direction, which exploit both the temporal pattern and co-occurrence among attributes. In addition, by providing a bidirectional connection structure in the attribute direction, the model enables us to alleviate the dependency of the chain order in what we call "classifier chain", which is a classical approach to multi-label classification.

To validate our methods, we conduct experiments using real-world construction-vehicle dataset.

I. Introduction

In this work, we tackle construction-vehicle activity recognition, which is useful for evaluating the skills of the operator, measuring work efficiency, and many other use cases. In similar past studies, many researchers have explored activity recognition for humans and robots [1], [2], [3], [4], [5], [6], [7], [8], [9].

Many studies have reported that modeling temporal patterns is very important for activity recognition [4], [5], [6], [7], [8], [9]. Vail et al. [4], for example, has proposed a conditional random field (CRF)-based activity recognition model for robots. Long short-term memory (LSTM) [10] has been researched for a number of years for the modeling of time series such as part-of-speech tagging [11] and machine translation [12]. Recently, LSTM has also been applied to many activity-recognition researches [7], [8], [9] because of its robustness.

While LSTM has been generally successful, imbalanced data, where some classes have small amounts of data compared to those of other classes, is however still a problem in construction-equipment recognition. Construction machines are used at various sites, such as quarries and coal mines, and thus, activity recognition involving construction machines leads to large-scale classes because various operations are performed at each site. Furthermore, the acquisition cost of construction-vehicle data is very expensive in terms of time and money. Therefore, it is very difficult to obtain sufficient data for each label. When we apply a standard multi-class LSTM model to this domain directly, it will be overfitted to classes which have large amounts of data. Thus, the classification accuracy for classes with less data will be greatly reduced.

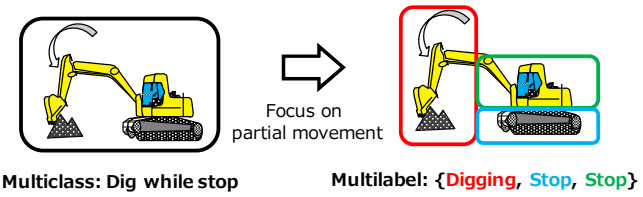


Fig. 1: Example of converting multi-class into multi-label

To tackle this data-imbalance problem, we convert a multi-class dataset to a multi-label dataset by assigning labels to the partial and simple activities of the construction vehicle, as shown in Fig. 1. This method has the following advantages:

- Imbalance of the dataset can be alleviated by exploiting each attribute, which results in the improvement of recognition performance.
- Test data can be classified even if training data is not available.

However, the discussion is insufficient for constructing an effective model toward multi-label time-series datasets like this study.

Lipton et al. [13] have proposed a multi-label diagnoses model based on time-series feature extraction, using LSTM for clinical medical data such as heart-rate data. However, this model uses LSTM only for temporal feature extraction, so it is not a sequence-labeling model, which is what is required in this study.

In one of the few existing studies, Shimosaka et al. [14] have proposed multi-task CRF (MT-CRF), which is a fully connected CRF model for all attributes. Through the introduction of full connections to CRF, it becomes possible to simultaneously consider the relationships among attributes and the dependency on the time series. On the other hand, because CRF is a linear model, it is not sufficient for complex construction-equipment data. In addition, CRF cannot model long-term time-series dependencies because it is based on a first-order Markov property.

In this study, we propose a novel sequence multi-labeling

[1]The authors are with the Department of Computer Science, Tokyo Institute of Technology, Tokyo, Japan. E-mail: {abe, simosaka}@miubiq.cs.titech.ac.jp

[2]The authors are researchers at the Komatsu Corporation, Tokyo, Japan. E-mail: {takuya_hino, motohide_sugihara, hiroki_ikeya} @global.komatsu

model based on multi-label LSTM, which is effective for multi-label time-series datasets. The proposed method can consider the complex and long-term temporal pattern of data and dependency among attributes simultaneously by combining LSTM and a multi-label classification method.

One important challenge in multi-label time-series classification is that the effective architecture of multi-label LSTM is not obvious and thus needs to be explored. Therefore, we comprehensively explore the important factors of model architecture for the sequence multi-labeling problem and incorporate them into the implementation.

In summary, the contributions of this research are as follows:

- We propose a classifier chain (CC)-based LSTM model, which is effective for the sequence multi-labeling problem. The proposed model can consider the relationship between long-term dependency and co-occurrence among attributes simultaneously by connecting hidden nodes in the time-series direction and attribute direction.
- In order to search for an effective architecture, we comprehensively investigate the important factors for the sequence multi-labeling problem. We raise three types of factors to be considered: 1) *the direction of temporal connections*, 2) *the direction of attribute connections*, and 3) *how to connect each layer*. Two types of methodologies are introduced to each of these, leading to eight CC-based multi-label LSTM models.
- To validate our proposed method, we conduct experiments using real-world construction-vehicle dataset. The experimental results show the importance of simultaneously considering the long-term temporal patterns and co-occurrence among attributes. Moreover, the effectiveness and importance of the proposed architecture are also discussed through the experimental results.

II. Related work

Activity recognition: In activity-recognition research, time-series models have been widely studied for handling the temporal dependencies of activities. Thus, Markov-based models have been applied in this research field for many years. According to literature, first-order Markov-based models have been actively applied to this domain to model temporal dependencies [4], [5], [6]. On the other hand, first-order Markov property-based models cannot consider a longer-term context because of the assumptions that are used with the Markov property.

Therefore, research on the application of recurrent neural network (RNN) [15], a deep-learning model that can take into account the long-term context, is in progress. RNN is a model that considers the long-term context while recursively calculating the internal state and has improved accuracy by modeling to a long-term time series. However, because of its model structure, RNN has a problem with the gradient vanishing during long-term time-series learning [16]. In order to solve this RNN gradient-disappearance problem, the application of long short-term memory (LSTM) [10] with a forgetting gate inside the cell has been actively applied in recent years [7], [8], [9].

Multi-label classification: Unlike in conventional activity-recognition researches, we tackle a sequence multi-labeling problem, wherein multiple activity attributes are estimated for each time bin. Problem transformation methods are intensively applied to a multi-label classification problem [17], [18]. Label powersets (LP) [17] are proposed as a simple transformation problem, wherein single labels are annotated into powersets of multi-labels. LP are a simple yet effective method. However, LP become problematic when the number of labels is increased, which leads to an explosion of classes. Therefore, LP are insufficient for large-scale datasets. To tackle this issue, binary relevance (BR) [17] is proposed, where the classifier of each attribute is individually trained. However, BR cannot model the co-occurrence among attributes, which is important for multi-label classification. Therefore, in recent years, the application of classifier chain (CC) [19] has become popular. CC considers the co-occurrence among labels by incorporating a chain structure into the classifier, by inputting the classification results as features.

Multi-label classification towards time-series: In recent years, multi-label classification models for time-series data have been attracting attention from researchers. Chen et al. [20] proposed a CNN-RNN-based multi-label classification model, which extracts local features using convolutional neural network (CNN) and then extracts long-term temporal patterns using RNN for computer-vision-based activity-recognition tasks. Lipton et al. [13] proposed a multi-label diagnoses model for clinical medical data by applying LSTM as a temporal feature extractor. While these models result in a high performance for multi-label time series, these models are not for sequence-labeling models. Therefore, these models cannot be cast into our domain. x Shimosaka et al. [14] proposed MT-CRF for a sequence multi-labeling problem, which is of a similar problem setting to our domain. MT-CRF models the co-occurrence of each attribute label and the temporal dependencies by having connections across all labels. However, CRF is a linear model, so it is not sufficient for learning the complex relationships among sensors. Besides, because it follows Markov property, it cannot model long-term temporal patterns.

III. Activity recognition for construction vehicles

A. Problem setting

Activity recognition in this study is formulated as a sequence-to-sequence mapping problem, where activity tags at each time are inferred from the sequence of sensor values attached to the construction vehicles. We define the number of sensors as d and the discretized time index as $t \in \{1, \cdots, T\}$. The vector $\boldsymbol{x}_t \in \mathbb{R}^d$ represents the concatenation of sensor values at time t. The activity labels at time t are represented by $\boldsymbol{y}_t = \begin{pmatrix} y_t^{(1)} & y_t^{(2)} & \cdots & y_t^{(l)} \end{pmatrix}^{\mathsf{T}} \in$

$Y_1 \times Y_2 \times \cdots \times Y_m$. The objective is to estimate label sequence $\{\boldsymbol{y}_t\}_{t=1}^T$ from the sequence of sensor values $\{\boldsymbol{x}_t\}_{t=1}^T$, with high accuracy.

In this research, we employ a sliding-window-based feature extraction for the input of LSTM. We extend the input vector using a mk size window ($m, k \in \mathbb{N}$). Here, m is the rate of sampling and k is the dimension of a vector obtained after downsampling. That is, the vector at time t is first extended by collecting sensor vectors focusing on the t-th frame using the mk size window. The extended vector is then downsampled with an interval of m frames. Finally, the concatenated input vector data from the series of vector data are obtained by normalizing the extended vector to make its mean equal to 0 and its variance equal to 1 for each column from the raw sensor vector $\boldsymbol{x}_t$ to $\tilde{\boldsymbol{x}}_t$. The formulation of this procedure can be written as follows: $\boldsymbol{X}_t = \left(\tilde{\boldsymbol{x}}_{t-mk}^\mathsf{T} \quad \tilde{\boldsymbol{x}}_{t-m(k-1)}^\mathsf{T} \quad \cdots \quad \tilde{\boldsymbol{x}}_{t+mk}^\mathsf{T}\right)^\mathsf{T} \in \mathbb{R}^{d(2k+1)}$.

B. Standard long short-term memory

We formulate a standard LSTM model used in multi-class sequence labeling. LSTM receives input sequence $\{\boldsymbol{X}_t\}_{t=1}^{t=K}$ from 1 to $K \in \mathbb{N}$ and output estimation sequence $\{y_t\}_{t=1}^{t=K}$ for the given input.

LSTM captures the long-term dependency, propagating temporal information via the hidden nodes. From this characteristic, LSTM has been frequently applied to activity-recognition researches in recent years [8], [9], [7]. The hidden nodes and outputs are formulated as follows:

$$\boldsymbol{i}_t = \sigma\left(W_{xi}\boldsymbol{X}_t + W_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right), \quad (1)$$

$$\boldsymbol{f}_t = \sigma\left(W_{xf}\boldsymbol{X}_t + W_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right), \quad (2)$$

$$\tilde{\boldsymbol{c}}_t = \tanh\left(W_{x\tilde{c}}\boldsymbol{X}_t + W_{h\tilde{c}}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{\tilde{c}}\right), \quad (3)$$

$$\boldsymbol{c}_t = \boldsymbol{i}_t \circ \tilde{\boldsymbol{c}}_t + \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1}, \quad (4)$$

$$\boldsymbol{o}_t = \sigma\left(W_{xo}\boldsymbol{X}_t + W_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o\right), \quad (5)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh\left(\boldsymbol{c}_t\right). \quad (6)$$

$W_{\cdot}$ and $\boldsymbol{b}_{\cdot}$ represent weight matrices and bias vectors, respectively. Operator $\circ$ indicates element-wise production, and $\sigma(x) = \frac{1}{1+\exp(-x)}$ represents the sigmoid function. $\boldsymbol{i}_t$, $\boldsymbol{o}_t$, and $\boldsymbol{f}_t$ represent the input gate, output gate, and forget gate, respectively. These gates select what kind of information should be discarded or not from the input data. $\boldsymbol{h}_t$ is a memory cell which stores the context information. This memory cell will be used to model the temporal dependency. Estimation for the data at frame t is given by the label which maximizes the probability $\boldsymbol{p}_t = \mathrm{softmax}\left(\boldsymbol{h}_t\right)$.

C. Conventional multi-label classification models

The goal of this study is to construct a multi-label LSTM, which is effective for multi-label time-series datasets. This section outlines the conventional multi-label classification method, which is the basic architecture of the proposed model.

1) Label powersets (LP): Label powersets is a technique that converts multi-label classification to a multi-class classification problem by reassigning a single label to a set of multi-labels. Because it converts the problem into a simple multi-class problem, existing methods can be applied easily.

On the other hand, the number of single labels increases in proportion to $|Y_1||Y_2|\cdots|Y_m|$. Thus, the dataset will have an imbalance, which leads the model to an over-fitting problem if $|Y|$ is large.

2) Binary relevance (BR): Binary relevance trains a model for each attribute $y_t^{(i)}$ ($i \in \{1, \cdots, m\}$) $\in Y_i$ $F_i \colon \mathbb{R}^d \to Y_i$ independently. Because binary relevance performs training for each attribute independently, it can suppress the imbalance of dataset better, compared to label powersets. However, because the independence of each attribute is assumed, the consistency of the label combination is not guaranteed (e.g., inconsistent output such as digging while running) because of the independence of the recognition phase.

3) Classifier chain (CC): Classifier chain models the co-occurrence among attributes by introducing a chain structure for the classifier. This model enables us to model co-occurrence by chaining each classifier to the others.

A chain structure is introduced by adding the classification result for an attribute as an input feature to the input of another classifier. In this study, there are cases where multi-class labels, rather than binary labels, are assigned to each attribute. Thus, we conduct encoding classification via one-hot encoding. That is, feature vector $\boldsymbol{X}_t^{(i)'} = \left(\boldsymbol{X}_t^\mathsf{T} \quad \hat{\boldsymbol{y}}_t{}^{(1)\mathsf{T}} \quad \cdots \quad \hat{\boldsymbol{y}}_t{}^{(i-1)\mathsf{T}}\right)^\mathsf{T}$ is used for training classifier for the i-th attribute $f_i \colon \mathbb{R}^d \times \{0,1\}^{|Y_1|} \times \ldots \times \{0,1\}^{|Y_{i-1}|} \to Y_i$. Here, $\hat{\boldsymbol{y}}_t{}^{(j)}$ is a one-hot vector using classification result for attribute j.

CC has been widely used in multi-label classification because of its ability to model co-occurrence among attributes. On the other hand, the performance of the model highly depends on the structure of the chain. However, it is not obvious what structure is the best.

In the following sections, we discuss the design principles of the multi-label LSTM based on this CC model.

IV. Multi-label LSTM for sequence multi-labeling

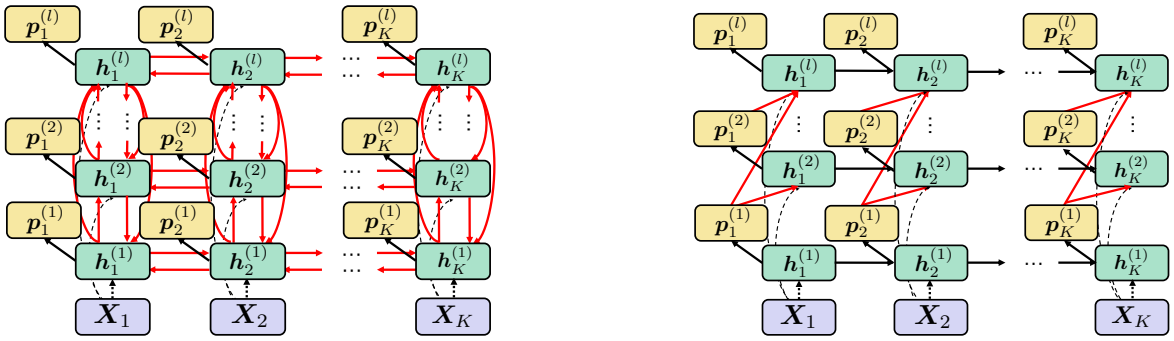
Unlike previous multi-label classification researches [21], [22], [23], [24], this study deals with a sequence multi-labeling method that assigns a multi-label to each frame. Therefore, modeling temporal dependency and handling dependency across attributes are both important.

Existing methods [21], [22], [23], [24] are classification models that are not based on time-series data, and thus cannot be directly applied to sequence multi-labeling as in this work. In addition, these models are not sufficient because they do not consider temporal patterns.

In this study, we propose a multi-label LSTM model that is effective for sequence multi-labeling problems.

TABLE I: Design guidelines towards implementation of Multi-label LSTM

Points	Temporal connection	Attribute connection	Connection target
Way of implementation	Unidirection, Bidirection	One-way-ordered, Arbitrary-ordered	Hidden node, Softmax



(a) ACC-BLSTM (b) SCCC-LSTM

Fig. 2: Example of variants of cascaded classifier chain LSTM (CCC-LSTM). (a) represents the arbitrary-ordered cascaded classifier chain bidirectional LSTM (ACC-BLSTM), which chains temporal and attribute layers bidirectionally. (b) represents the softmax version cascaded classifier chain LSTM (SCCC-LSTM), which chains softmax outputs for attribute layer.

A. Design policy towards multi-label LSTM

In this study, we propose a novel multi-label LSTM model for multi-label time series.

The basic architecture for the multi-label LSTM model is based on the classifier chain (CC) model introduced in section III-C. The purpose of this work is to build a model that simultaneously takes into account the long-term temporal patterns and the co-occurrence among attributes by having a connection between the time series and the attributes. However, it is not obvious what kind of CC-based LSTM architecture would be effective. In this study, we will raise factors that are considered to be important for sequence multi-labeling and incorporate these factors into the proposed architectures. We focus on three types of components: the direction of temporal connections, the direction of attributes connections, and how to connect each layer.

Firstly, we consider two types of connections in the direction of temporal connections: unidirectional and bidirectional. By incorporating bidirectionality, it is possible to use more temporal information, and it enables the model to increase the expressive ability for time series.

Secondly, we consider two types of connection directions in the direction of attribute connections: one-way-ordered and arbitrary-ordered. By incorporating bidirectionality, we will obtain the possibility of reducing the dependency of chain structure in CC, making the model more stable.

Finally, we consider two ways of connecting in the attribute direction: chaining hidden layers of softmax values. The basic architecture is to consider co-occurrence among attributes by connecting hidden layers. On the other hand, as the number of dimensions of the hidden nodes increases, the input dimension becomes enormous and the risk of overfitting is increased. Therefore, we introduce chaining softmax values to reduce the number of parameters.

In summary, we construct models following the design guidelines shown in table I. In this study, these factors are combined to construct models, and we have a total of $2^3 = 8$ CC-based models. We will give the formulation of the proposed architecture in the next section.

B. Proposed method: multi-label LSTM

1) Cascaded classifier chain LSTM (CCC-LSTM): The basic model of classifier chain (CC)-based LSTM. The difference from the standard LSTM model is that this model makes use of feature vector $\boldsymbol{v}_t^{(l)}$ for the l-th attribute to model co-occurrence among attributes. This feature vector is made by concatenating the outputs from previous hidden layers and input data. The internal gates in the LSTM for the l-th attribute are calculated as follows:

$$\boldsymbol{v}_t^{(l)} = \left(\boldsymbol{X}_t^{\mathsf{T}} \quad \boldsymbol{h}_t^{(l-1)\mathsf{T}} \quad \boldsymbol{h}_t^{(l-2)\mathsf{T}} \quad \cdots \quad \boldsymbol{h}_t^{(1)\mathsf{T}}\right)^{\mathsf{T}} \tag{7}$$

$$\boldsymbol{i}_t^{(l)} = \sigma\left(W_{xi}^{(l)}\boldsymbol{v}_t^{(l)} + W_{hi}^{(l)}\boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}_i^{(l)}\right), \tag{8}$$

$$\boldsymbol{f}_t^{(l)} = \sigma\left(W_{xf}^{(l)}\boldsymbol{v}_t^{(l)} + W_{hf}^{l}\boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}_f^{(l)}\right), \tag{9}$$

$$\tilde{\boldsymbol{c}}_t^{(l)} = \tanh\left(W_{x\tilde{c}}^{(l)}\boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}_{\tilde{c}}^{(l)}\right), \tag{10}$$

$$\boldsymbol{c}_t^{(l)} = \boldsymbol{i}_t^{(l)} \circ \tilde{\boldsymbol{c}}_t^{(l)} + \boldsymbol{f}_t^{(l)} \circ \boldsymbol{c}_{t-1}^{(l)}, \tag{11}$$

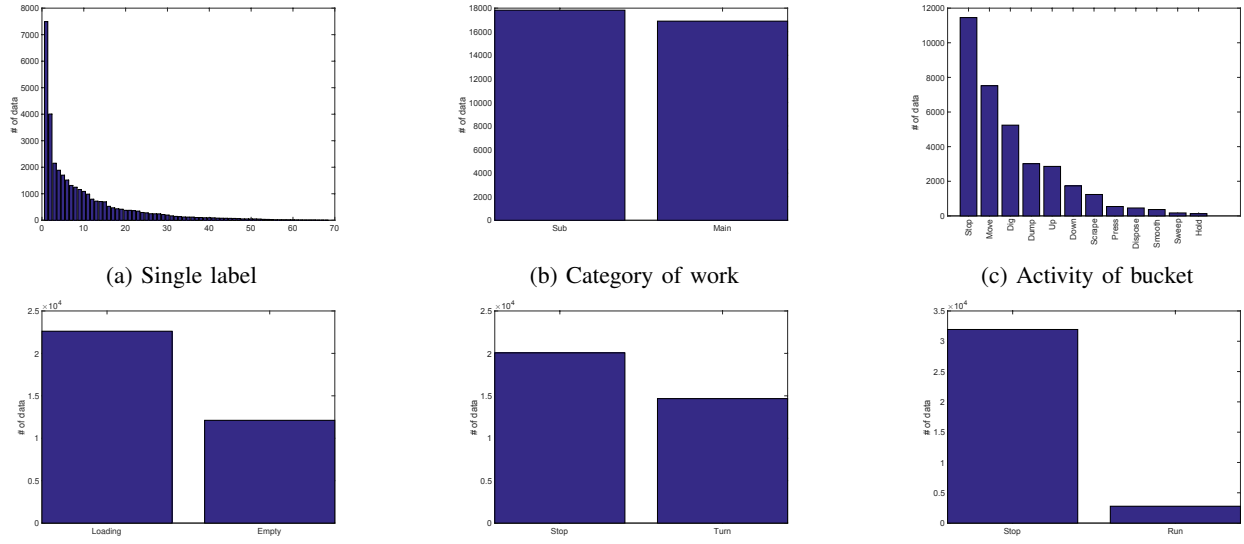
$$\boldsymbol{o}_t^{(l)} = \sigma\left(W_{xo}^{(l)}\boldsymbol{v}_t^{(l)} + W_{ho}^{(l)}\boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}_o^{(l)}\right), \tag{12}$$

$$\boldsymbol{h}_t^{(l)} = \boldsymbol{o}_t^{(l)} \circ \tanh\left(\boldsymbol{c}_t^{(l)}\right). \tag{13}$$

By using a vector formulated as (7) and previous hidden nodes $\boldsymbol{h}_{t-1}^{(l)}$, we define the calculation process of CCC-LSTM cell (7) - (13) as

$$\boldsymbol{h}_t^{(l)} = \mathrm{CCCLSTM}(\boldsymbol{v}_t^{(l-1)}, \boldsymbol{h}_{t-1}^{(l)}). \tag{14}$$

It should be noted that CCC-LSTM is equivalent to the well-known standard LSTM, except that this model has another connection along with attributes. This connection makes it possible to model dependencies among the attributes. Based on this CCC-LSTM, we formulate factors in the table I into the following sections.



(a) Single label (b) Category of work (c) Activity of bucket

(d) Loading state (e) Turning state (f) Running state

Fig. 3: Data distribution for (a) single label, and (b) - (f) each attribute after conversion to the multi-label

2) Arbitrary-ordered Cascaded Classifier Chain LSTM (ACC-LSTM): A variant of the CCC-LSTM model, in which bidirectionality was introduced to the connection for the direction of the attribute. In the CC model, classification performance is highly dependent on the chain structure. In order to alleviate this dependency, we introduce bidirectionality.

We define equation (14) as the calculation process for *upward* CCC-LSTM cell ${}^{(\mathrm{U})}\boldsymbol{h}_t^{(l)}$. Similarly, we define the calculation process for *downward* CCC-LSTM cell as

$$ {}^{(\mathrm{D})}\boldsymbol{h}_t^{(l)} = {}^{(\mathrm{D})}\,\mathrm{CCCLSTM}({}^{(\mathrm{D})}\boldsymbol{v}_t^{(l+1)}, {}^{(\mathrm{D})}\,\boldsymbol{h}_{t-1}^{(l)}). \tag{15} $$

Final output is obtained by concatenating the upward and downward vector $\boldsymbol{h}_t^{(l)} = \left({}^{(\mathrm{U})}\boldsymbol{h}_t^{(l)\mathsf{T}} \quad {}^{(\mathrm{D})}\boldsymbol{h}_t^{(l)\mathsf{T}}\right)^{\mathsf{T}}$.

3) Cascaded classifier chain bidirectional LSTM (CCC-BLSTM): A variant of CCC-LSTM model, in which bidirectionality was introduced for the time direction. This architecture enables models to use more temporal information, which leads to a high classification ability. This kind of model has been firstly proposed as BLSTM [25].

In combination with the aforementioned ACC-LSTM model, we implement a model which introduced all connections bidirectionally shown in Fig. 2a.

4) Softmax version cascaded classifier chain LSTM (SCCC-LSTM): A variant of CCC-LSTM model that uses softmax output for the chaining of hidden nodes. In CCC-LSTM, the dimension of the feature vectors is significantly increased at the suffix of the chain, carrying the risk of overfitting due to the curse of dimensionality. Therefore, we implement SCCC-LSTM, which is a variant of CCC-LSTM model that links output obtained by softmax output $\boldsymbol{p}_t^l = \mathrm{softmax}\left(\boldsymbol{h}_t^l\right)$ instead of hidden nodes $\boldsymbol{h}_t$. By using softmax values, as shown in Fig. 2b, instead of directly chaining hidden nodes, it considers the co-occurrence among attributes, while reducing the number of dimensions of input features even at the suffix of a chain.

V. Experiments

A. Construction-vehicle dataset

The dataset used for this experiment consists of multi-label time-series data that record the work of construction vehicles for around 55 minutes. The sensor values are acquired at intervals of 100 ms. As a result, we have 34,782 frames dataset. The acquired data contain engine speed, engine torque, pressure in each direction of the control lever, and so on. For the experiment, 42 types of sensors are used.

The dataset contains 68 labels as a single label. Towards this single label, we annotated multi-labels focusing on the partial activity of the construction equipment. Specifically, a total of five types of attributes are annotated: category of work, movement of the bucket, loading state, turning state, and running state. Among these attributes, "movement of the bucket" is composed of 12 classes, while the others are binary classes. Fig. 3 shows the data distribution for (a) single label as multi-class representation and (b) to (f) for each attribute. As we can see, the distribution for the single label is highly imbalanced and most labels have almost no data. On the other hand, in the multi-label, although there are some classes which have less data, imbalance is alleviated compared to with the single label.

To compare these data imbalances, we introduce the simple index. Define $C^{\mathrm{a}}_{\max}$ as the maximum value of the number of data points in the attribute a and $C^{\mathrm{a}}_{\min}$ as the minimum value. We define simple data imbalance as $C^{\mathrm{a}}_{\mathrm{bal}} = \frac{C^{\mathrm{a}}_{\max}}{C^{\mathrm{a}}_{\min}}$. When this index is used, the imbalance for a single label is $C^{\mathrm{Single}}_{\mathrm{bal}} = 7491.0$, while the imbalances for each of the attributes of a multi-label are $C^{\mathrm{State}}_{\mathrm{bal}} = 1.06$, $C^{\mathrm{Load}}_{\mathrm{bal}} = 1.87$, $C^{\mathrm{Work}}_{\mathrm{bal}} = 89.5$, $C^{\mathrm{Run}}_{\mathrm{bal}} = 11.47$, and $C^{\mathrm{Turn}}_{\mathrm{bal}} = 1.37$. This index also shows that the imbalance is highly reduced.

TABLE II: Accuracy for the classes wherein no training data is available.

Models	BR-LSTM	CCC-LSTM	SCCC-LSTM	ACC-LSTM	CCC-BLSTM	ACC-BLSTM
Subset-Accuracy	0.122	0.012	0.0008	0.01	0.122	0.122

B. Experimental setting

We employ simple comparative methods: combination models of logistic regression (LR), multilayer perceptron (MLP), and multi-label classification. We also employ label powersets LSTM (LP-LSTM) and binary relevance LSTM (BR-LSTM). We employ the 8 variants of CCC-LSTM models discussed in IV-A as the proposed models. Feature extraction is performed via a time-sliding window, where size is set to 5. For the parameters of LSTM, we set 256 for the dimension of each hidden layer and 0.5 for the dropout of each layer. The sequence of LSTM is set to 50. The MLP has 3 hidden layers and set to 256, 128, and 64 dimensions for each of the hidden layers. LR is regularized via L2 regularization, with a regularization coefficient of 1.0. These parameters are empirically determined through experiments. We conduct 10-fold cross-validation for the validation.

We employ subset accuracy and Hamming accuracy as evaluation metrics. Subset accuracy indicates an exact match rate between estimation and ground truth, while Hamming accuracy indicates a partial match rate between estimation and ground truth. For the given instance $\boldsymbol{x}$ and ground truth $\boldsymbol{y}$, each index is formulated as follows:

$$\mathrm{Subset}(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x})) = [\![\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})]\!], \quad (16)$$

$$\mathrm{Hamming}(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x})) = \frac{1}{l}\sum_{i=1}^{l}[\![y_i = f_i(x)]\!]. \quad (17)$$

Here, $[\![\cdot]\!]$ denotes the indicator function, which returns 1 if condition $\cdot$ is true and 0 otherwise.

C. Results

1) Comparing basic performance of each model: Fig. 4 illustrates the results of plotting the Hamming accuracy on the horizontal axis and subset accuracy on the vertical axis. In Fig. 4, the model is revealed to become better as it proceeds to the upper right. Fig. 4 shows that the proposed multi-label LSTM model outperforms other models such as LR and MLP. In particular, the CCC-BLSTM model that incorporates bidirectionality in the time-series connection demonstrates good performance. This result establishes the importance of considering temporal information.

Focusing on the multi-label method, the classifier-chain-type model shows a relatively good performance compared to those of the other models, excluding MLP. This result shows the importance of considering co-occurrence among the attributes.

The effect of the arbitrary-ordered model is small in the CCC-LSTM models. Meanwhile, this effect improves when SACC-BLSTM and SCCC-BLSTM are compared. The improvement comes from reducing the dimension of features via chaining softmax values, while the number of parameters is increased via the arbitrary-ordered architecture.

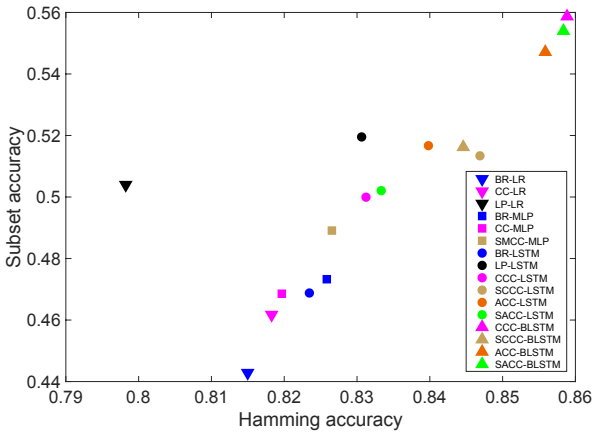


Fig. 4: Results with scatter plot: Horizontal axis shows the Hamming accuracy; vertical axis shows the subset accuracy

The LP models show good performance in terms of subset accuracy but low performance in terms of Hamming accuracy. These results indicate that LP models are strongly influenced by imbalanced data because the LP models solve the problem as a multi-class classification, and thus, tend to be overfitted to the class with a large number of data points.

2) Recognition performance for missing class: In construction-vehicle activity recognition, there are cases where some sites cannot provide any training dataset. Hence, it is important to recognize classes which are missing in the training dataset. To evaluate the performance for such a situation, we conduct an additional experiment.

We remove the class from the training dataset, and then we train our proposed model by using the remaining dataset. We evaluate the performance of the proposed model via the removed classes. The parameters are the same as those described in V-C.1. We remove a class which contains a total of 1,313 frames in the dataset, and use it for test.

Tab. II shows the result of this experiment. According to the results, our proposed model has the ability to recognize classes which have no training dataset. This result validates the effectiveness of our proposed method. Note that the LP model cannot classify any given classes because no training dataset is available. Therefore, this result also indicates the effectiveness of multi-label, rather than multi-class, for construction-vehicle activity recognition.

VI. Conclusion

In this study, we discussed the construction of a multi-label LSTM that is effective for multi-label time-series data. Because the effective architecture was not obvious, the architecture was formulated and implemented focusing on three points: time-series connection, attribute direction connection, and connection method. In addition, a comparison experiment was conducted using actual data that recorded construction-equipment work to verify the effectiveness. As

a result of the experiment, the CCC-BLSTM model incorporating time-series bidirectionality demonstrated good performance. From this observation, it can be said that it is important to simultaneously consider the relationship between attributes and time-series dependency.

Future issues include support for various construction-equipment models. Because the current model is effective for only one type of equipment, a model that can be applied to multiple types of equipment is desired. Another issue includes validating our proposed towards other real-world datasets.

REFERENCES

[1] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. of PerCom*. Springer, 2004, pp. 1–17.

[2] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. of AAAI*, vol. 5, no. 2005, 2005, pp. 1541–1546.

[3] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *Transactions on biomedical engineering*, vol. 44, no. 3, pp. 136–147, 1997.

[4] D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proc. of AAMAS*. ACM, 2007, p. 235.

[5] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," 2005.

[6] K. Han and M. Veloso, "Automated robot behavior recognition," in *Robotics Research*, 2000, pp. 249–256.

[7] F. J. Ordòñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, 2016.

[8] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *Proc. of the IJCAI*. AAAI Press, 2016, pp. 1533–1540.

[9] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *IMWUT*, vol. 1, no. 2, pp. 11:1–11:28, 2017.

[10] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.

[11] X. Ma and E. Hovy, "End-to-end sequence labeling via bi-directional lstm-cnns-crf," in *Proc. of ACL*, 2016, pp. 1064–1074.

[12] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. of NIPS*, 2014, pp. 3104–3112.

[13] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzel, "Learning to diagnose with lstm recurrent neural networks," *arXiv preprint arXiv:1511.03677*, 2015.

[14] M. Shimosaka, T. Mori, and T. Sato, "Robust action recognition and segmentation with multi-task conditional random fields," in *Proc. of ICRA*, 2007, pp. 3780–3786.

[15] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. of INTERSPEECH*, 2010, pp. 1045–1048.

[16] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pp. 107–116, 1998.

[17] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in *Data mining and knowledge discovery handbook*, 2009, pp. 667–685.

[18] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp. 1–13, 2007.

[19] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine learning*, p. 333, 2011.

[20] T. Chen, Z. Wang, G. Li, and L. Lin, "Recurrent attentional reinforcement learning for multi-label image recognition," in *Proc. of AAAI*, 2018.

[21] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern recognition*, pp. 1757–1771, 2004.

[22] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, "CNN-RNN: A unified framework for multi-label image classification," *CoRR*, vol. abs/1604.04573, 2016.

[23] M.-L. Zhang, "ML-RBF: RBF neural networks for multi-label learning," *Neural Processing Letters*, vol. 29, no. 2, pp. 61–74, 2009.

[24] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.

[25] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional lstm and other neural network architectures," *Neural Networks*, pp. 602–610, 2005.