ML1473

An Intelligent Data Labelling Tool that Supports Machine Learning for Frost Forecast

Tokio Tanaka[1,*], Shugo Yoshida[1], Liya Ding[1], Kosuke Noborio[2] and Kazuki Shibuya[2]

[1] School of Science and Technology, Meiji University, Japan   [2] School of Agriculture, Meiji University, Japan

Introduction

The damage caused by frost to crops is significant, and countermeasures need to be taken, but the labor and cost of such measures is a heavy burden on farmers.

Therefore, highly accurate frost forecast makes it possible to take pinpoint countermeasures. Highly accurate forecast is possible by using machine learning to predict the occurrence of frost, which requires a training dataset labelled with the presence or absence of frost.

Currently, we use a manually labelled dataset based on measured meteorological factors. However, more labelled data is required for further development and continuous training of forecast models, therefore an efficient labelling method is needed.

In this study, we propose a labelling tool that performs automatic labelling using machine learning for most of the cases but leaves only those with insufficient confidence to manual processing.

Methodology

・Random forests, gradient boosting decision trees, and logistic regression are employed for labelling models with Air Temperature, Wetness Level, and VPD (Vapor Pressure Deficit) as input features.

・The labelling models are first trained with a small set of manually labelled data, and then applied to unlabelled data to make prediction. The prediction results are then used as labels with the probabilities of prediction as confidence.

・**One-threshold method**: the threshold applies to confidence as cutting point for labelling Frost and Non-frost

・**Two-threshold method**: only those cases with sufficiently high (low) confidence are confirmed as Frost (Non-Frost), and the rest as unconfident for manual labelling.

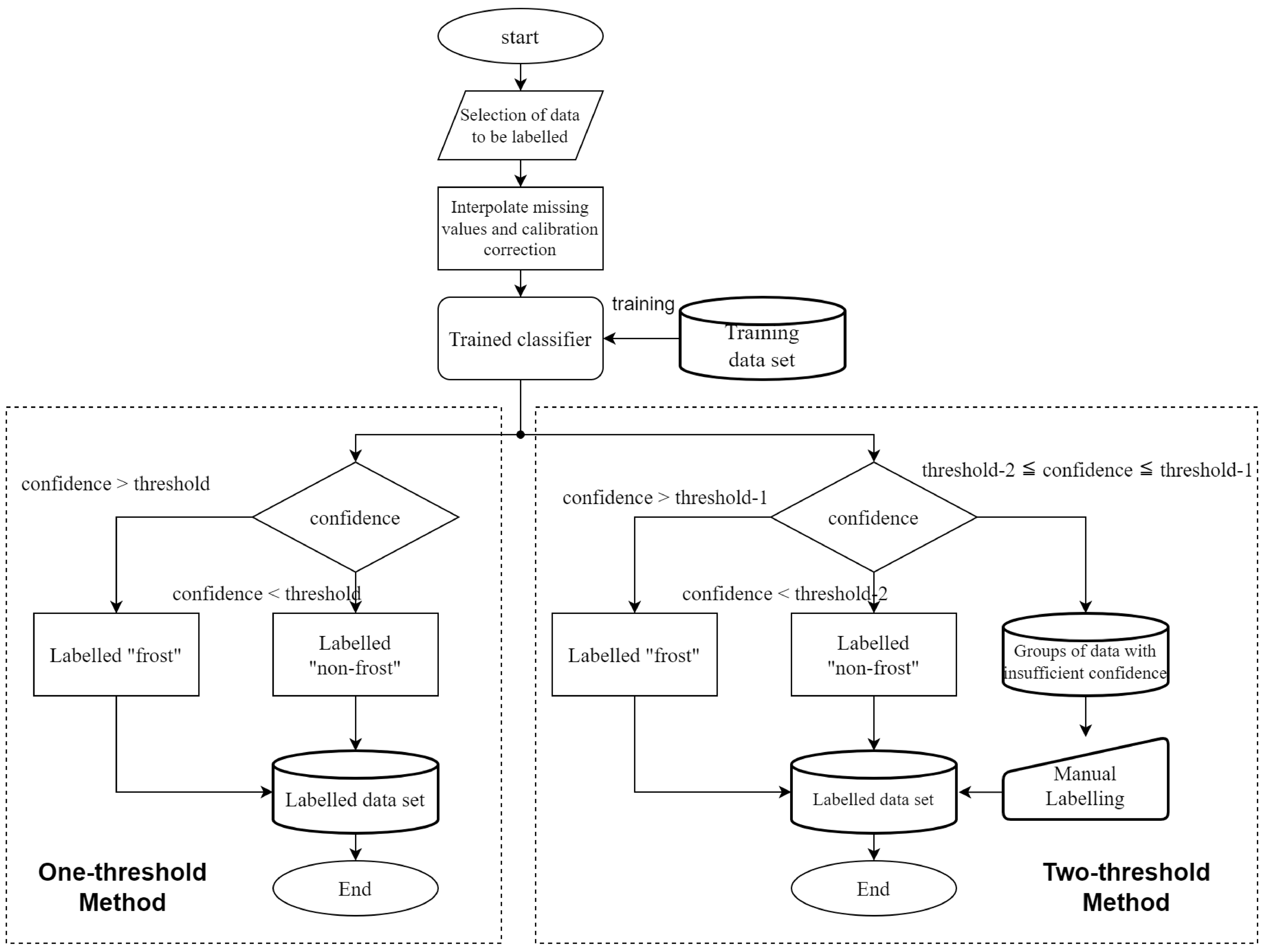
Datasets

Air Temperature and VPD measured by "ATMOS 41 All-in-one Weather Station" and "Wetness Level" measured by "PHYTOS 31/LWS Leaf Wetness" was used.

Table 1. List of datasets used

Name	Measurement place	Measurement start time	Measurement end time	Increment [min]	Number of samples	Dataset used for
Dataset_Ikeda_M19S	Ikeda-cho, Hokkaido, Japan	2019/4/25 0:00:00	2019/6/18 23:55:00	5	15840	training (labelled)
Dataset_Ikeda_M19A	Ikeda-cho, Hokkaido, Japan	2019/9/20 0:00:00	2019/10/19 23:55:00	5	8640	training (unlabelled)
Dataset_Ikeda_M20S	Ikeda-cho, Hokkaido, Japan	2020/4/26 0:00:00	2020/6/1 23:55:00	5	10656	training (unlabelled)
Dataset_Ikeda_M20A	Ikeda-cho, Hokkaido, Japan	2020/9/28 0:00:00	2020/10/20 23:55:00	5	6616	test
Dataset_Ikeda_M21S	Ikeda-cho, Hokkaido, Japan	2021/4/26 0:00:00	2021/6/2 23:55:00	5	10944	test

Processing Flow



Result

Table 2. shows case numbers in different groups after processing

Group-1: mislabelled cases by one-threshold method

Group-2: unconfident cases by two-threshold method

Group-3: cases in group-1∩group-2

Table 2. Processing Result

Dataset	Classification method	Group-1	Group-2	Group-3	Group-3/Group-1
Dataset_Ikeda_A20A	LogisticRegression	48	116	47	97.9%
	random_forest	48	103	34	70.8%
	lightGBM	57	27	11	19.3%
Dataset_Ikeda_A21S	LogisticRegression	92	394	92	100.0%
	random_forest	47	316	46	97.9%
	lightGBM	58	79	9	15.5%

Conclusion and Future Work

・Logistic regression appeared to offer best performance.

・From Table 2, by using two-threshold method, we could successfully capture most of mislabelling when using one-threshold method as unconfident cases.

・By manually labelling only the unconfident cases, high quality labelling data can be obtained with greatly reduced effort.

・We would consider unsupervised learning to extract useful patterns for supporting automatic labelling so that we can further reduce the amount of manual labelling required.

Reference

Y. Tamura, S. Yoshida, K. Owada, L. Ding, K. Noborio & K. Shibuya, “Method of Data Labelling Using Machine Learning for Frost Forecasting”, the 15th Inter. Conf. on Innovative Computing, Information and Control (ICICIC2021), A3-5: ICICIC2021-031, Sep 15-16th, 2021.