AgriBase: Status Labeling for IoT Systems

Hiroyuki Uchiyama
NTT Software Innovation Center
uchiyama.hiroyuki@lab.ntt.co.jp

1. INTRODUCTION

Various IoT Systems have been proposed on many areas. Big data generated by IoT devices will make the world change. For example, in agriculture production, field sensors acquire temperature, humidity, illuminance, and so on. These values may be useful for farm work(e.g., efficient pesticide spraying against harmful insects or plant disease). However, general farmers don't know how to leverage the big data to predict harm and control environments. In this paper, we show the vision of the *AgriBase* system to manage sensor data and to support inheritance of professional skills user-friendly.

We introduce status labeling to make farmers utilize the big data by themselves. Status labeling is simple conception by which users can handle their data easily. When users hope to record status(e.g. incidence of vermin), they can label as vermin(e.g. cucurbit leaf beetle) into AgriBase with current timestamp. It can be said that AgriBase is spatio-temporal database which has machine learning functionalities. After users registered status with label, if the same status occurs, AgriBase informs users labeled alarm. Recognition method of similar status is implemented by comparing the current time series data and past labeled time series data.

In Japan, decreasing the nubmer of farmers by population aging(The farmers' average age is about 65 years old!) is an urgent issue because (1)inheritance of skills is insufficient due to lack of successors and (2)as a result, abandonment of cultivated land continues to grow. For the near future, Japanese agricultural skills gets lost when the current farmers retire, who have skills that has not documented. To prevent such a situation, AgriBase records and provides advanced famers' operations (e.g. preventive pesticide spraying, quantity and timing of fertilizing) with environmental time series sensor data. When new farms' environmental sensor data is registered to AgriBase, then AgriBase generates the filtering model based on labeled status. If new time series farm data insert into AgriBase and is similar to already stored labeled time series data, then AgriBase sends farmers alert/warn/info level information and suggests what they should do.

8th Biennial Conference on Innovative Data Systems Research (CIDR '17) January 8-11, 2017 , Chaminade, California, USA.

2. CHALLENGES

AgriBase must give farmers simple and valuable functionalities. In general, users who utilize machine learning and time series data matching must decide prediction target, pattern matching target which includes temporal interval, their parameter settings, and so on. These assumptions will impair usability. AgriBase accepts sensor data and labeled snapshot point and actions by user, then it aims to automatically suggest risks and manipulations for the current environments.

3. TEMPORAL INTERVAL SELECTION TO COMPARE

There are two important questions related to temporal interval selection to compare. What statistical methods can we use to assess the differences between the time series? How long temporal interval to compare time series should we select? For the former, various methods [1] are proposed. However, methods based on dimensionality reduction are difficult to understand generally for farmers. So L_p norm or time warping is useful to show users easily understandable visualization(Sec.4). For the latter, naive approach which compares all interval around labeled status timing and all inputs. However, this approach require high computational complexity. By introducing emergency levels which related to response time required by users. In agriculture context, temporal lack of water must water the field as soon as possible. On the other hand, prevention of disease based on AgriBase's prediction isn't needed momentarily. But to predict disease may need to compare time series data such as during 3 months. A farmer may spend time to decide whether do action or not. In this context, we need to research stratified stream processing which generates statistics(e.g. average) various granularity automatically. By using this conception, we think AgriBase will be able to reduce computational complexity.

4. VISUALIZATION

When AgriBase informs alert to a farmer, she checks the alert correctness totally and decides whether she does next actions or not. To decision making, AgriBase must describe "why the alert arise" by intuitive and comprehensible visualizations. AgriBase holds sensor data in multi-dimensional space. By generating visualizations automatically, AgriBase aims to give users intuitive understandable results.

5. REFERENCES

[1] Y. Sakurai, Y. Matsubara, and C. Faloutsos. Mining and forecasting of big time-series data. *ACM SIGMOD Conference, Tutorial*, pages 919–922, 2015.