# A Multi-Layer Feature-Assisted Approach in Crowd-Labelling

**Ping-Hao Chen[1], Meng-Ying Chan[2], Chi-Chia Huang[2], Yi-Ching Huang[2] and Jane Yung-jen Hsu[1,3]**
[1]Department of Computer Science and Information Engineering
[2]Graduate Institute of Networking and Multimedia
[3]Intel-NTU Connected Context Computing Center
National Taiwan University
{r03922079, r03944015, d01944003, d00944010, yjhsu}@csie.ntu.edu.tw

### Abstract

Consistent and correct labels help a good performance in machine learning. However obtaining high-quality label from people remains a challenge in crowdsourcing domain. To get accurate and consistent labels, we propose a multi-layer feature-assisted labelling approach to guide the crowds to give correct annotations which satisfy the requesters requirements. Through labelling guidance and re-labelling mechanism, we can reduce the effect of concept evolution and make the outcomes more consistent. We implemented the approach to collect activity labels in a seminar room. We hope the idea can contribute to various machine learning topics.

## Introduction

Labelling is significant for machine learning, and a good learning performance depends on high quality labels. However, (Kulesza et al. 2014) issues the phenomenon of concept evolution that annotators cannot follow a consistent criterion on labelling and thus would inevitably hurt label quality. The task requester was used to set up the annotation rules first and expected a high quality result in a crowd-labelling task, but in practice it is very difficult for the requester and the annotators to share a same mind without any cognition gap. Therefore, how to guide the crowd to label correctly and meet the requesters objectives becomes an important issue.

People would easily hold separate opinions and viewpoints on complicated questions, but consent with each other on simple and obvious ones. Thus dividing a challenging labelling job into some simple tasks might be a possible solution. (Farhadi et al. 2009) tells us that an object can be better described in its attributes. We consider using the features of the labelling target to assist on the problem would be a brilliant way.

Assume that the features could infer some hints or constraints on the label outcome, we propose a feature-assisted approach to eliminate cognition gap and increase label consistency. We make the crowd to note on some objective and unambiguous features first before giving a label, sequentially moving onto harder features. The previous annotated features would provide some messages to the later annotators. We can build up a model to automatically predict a potential outcome fed by previous features. We consider validation bring more consistency than labelling directly. Here we can use the prediction as a guide to the annotators to consent their opinions on their task.

To modelize the idea, we structured the features into layers. First we collect all the possible features of our labelling target, distribute the features into different layers by how it is easy to make the crowds disagree with each other. The features that make less diverse opinions would be processed earlier, vice versa. We construct the system that asks the crowd to label on easy features first, and then use the validated features to generate a prediction on other harder features, step by step finishing annotating all the features with the guidance from early stage annotators and finally the labelling task itself. The multi-layer structure gradually converge possible final outcomes layer by layer, lowering the risk caused by some "creative" labelling workers.

In this work, we applied the approach into activity recognition field as a use case. Many activity recognition researchers require high quality ground truth labels for training their model, some of them would commit the need to crowd-sourcing solutions but disappointed on getting a satisfiable result. Something must be done here.

## Use Case: Room Activity Labelling

We built a system which implements double-layer for collecting activities in seminar room. We used a sequence of surveillant pictures of a college seminar room as our labelling material, with the frequency of 5 minutes each and length of 6 months. We asked the participants to categorize the pictures into 4 classes: Empty, Meeting, Lecture and Study. New labels are also acceptable if he/she didnt find an appropriate one among the 4 given classes.

Figure 1 shows the flow of our structure. We first asked the crowd to generate the 10 most significant features of the pictures. As a double-layer approach, all the features here are to be annotated in the first layer. We allocate some crowd resource on annotating the features in this stage. After gathering the annotations, we then build up a model and feed them inside to predict second layer features. The feature here is the ultimate label we desire which variates from the 4 categories. The prediction would be a guide for the second layer annotators to make the labels more consistent. After finishing all the labelling process in each layer, we propose a re-labelling system which allows the last layer annotators to review and change their answers, mitigates the problem of in-person inconsistency.

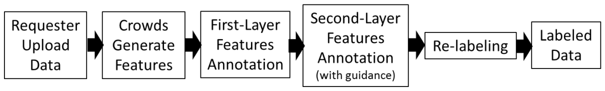


Figure 1: Double-layer labelling flow.

## Crowds generate features

To extract the general rules and the most considered features while people labelling, we recruited some feature generators in advance. They were asked to label through a set of 200 pictures and leave some words as an explanation after labelling each picture. We then collect these reason-carrying words and aggregate them into 20 clusters, and later rephrase them into 20 picture features which indicate the 20 most considered aspect that helps people making determination. We chose 10 unambiguous ones like the number of people, the projector on/off, the lights on/off and so on, to be our first layer features.

## Labelling guidance

In our system, the annotated lower layer features of the labelling target would help to train a prediction model to guide the upper layer annotators. Given the annotations on the 10 special features from the first layer, we are now focusing on training a good model to predict on second layer features, providing a likelihood values for all the categories the picture should be assigned to. Figure 2 is our sample guidance interface. The main part is our target picture which calls for labelling, and the lower right shows the 10 features annotated by the first layer participants. The upper right is the likelihood level if the picture belongs to each category, predicted automatically by the first layer features. At the "Others" part, we also show some most common used new tag below for picking. The lower left displays some reference labelled pictures which have similar characteristics to our target observed from the 10 features. We hope these guidance mechanisms would help annotators to keep in a simple and identical mind on their labelling job.

## Re-labelling

To strengthen the overall consistency and reduce the side effect of concept evolution, we added the re-labeling system in the last stage. In second layer, after the annotators finish all labelling task with the assistance of our guidance mechanism, we would check through all his/her labels and automatically detect if there are any abnormal or improper labels. The picked out ones would then be sent to the annotator again and asked him/her to re-label. As what we have done in the guidance mechanism, we would show the features, the likelihood prediction on each category and some similar pictures aside as a reference.

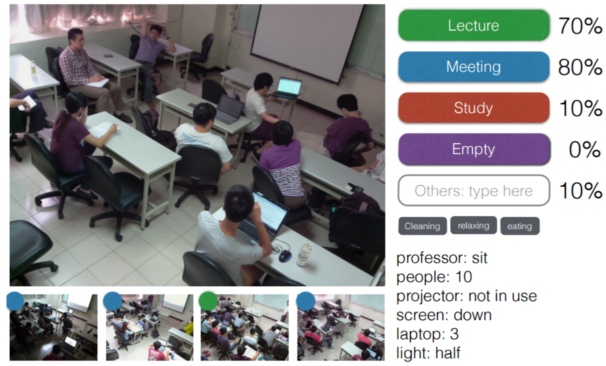


Figure 2: Guidance interface.

## Conclusion and Future Work

We propose a multi-layer feature-assisted approach in crowd-sourcing labelling task. With the assistance of label guidance and the re-labelling mechanism, we expect the effect of concept evolution and label inconsistency would decline. The labels would become more convergent and meet the requesters needs more. We implemented a double-layer version on activity recognition labelling task like categorizing pictures, used the crowd to generate features first, and recruited some workers to annotate the features on the first layer. In the second layer, we proposed a guiding way of proving annotated features, likelihood in each category and similar pictures, and now training a model for guidance. In the future, we will keep going on this use case, building models and paying effort on evaluations.


## Acknowledgments

This work was supported in part by the Ministry of Science and Technology, National Taiwan University, and Intel Corporation under Grants MOST 103-2911-I-002-001, NTU-ICRP-104R7501, and NTU-ICRP-104R7501-1


## References


Farhadi, A.; Endres, I.; Hoiem, D.; and Forsyth, D. A. 2009. Describing objects by their attributes. In *CVPR*, 1778–1785. IEEE Computer Society.

Kulesza, T.; Amershi, S.; Caruana, R.; Fisher, D.; and Charles, D. 2014. Structured labeling for facilitating concept evolution in machine learning. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, 3075–3084. New York, NY, USA: ACM.