Active Learning of Urban Functional Clusters

Hangxin Lu

Prof. Gerhard Tröster
Wearable Computing Group, ETH Zürich
hangxin.lu@epfl.ch

Introduction
City is a dynamic and complex system with aggregated and interconnected regions where people interact with different purposes. To analyze urban model, we seek to assess the functional aspects of urban spaces where functional clustering[3] is regarded as an effective tool. Our objective is to effectively perform functional clustering with active learning and crowdsourcing techniques.

Theory/Method/Hypothesis
Our project contains two parts: construction of crowdsourcing platform, design & evaluation of the proposed active learning algorithm.

1. Crowdsourcing Platform
We design an explicit crowdsourcing platform to collect functional data from human’s perspectives. Our web interface effectively visualizes the functional aspects of urban spaces and it is implemented with HTML, CSS, javascript, jquery axj and google map API for web. We incorporate our web interface with crowdflower, a crowdsourcing platform that recruits on-demand workers to perform data related works. Users are asked about how much they agree on the statement “the two regions are functional similar”, where they could choose an answer from Likert scale strongly disagree, disagree, neutral, agree and strongly agree. In this manner, we collect the functional pairwise similarities.

2. Active Learning Model
In order to facilitate functional clustering with crowd provided information, we propose an active learning algorithm. It consists of two phases: Fringed Based Active Query Selection and Iterative Feature Update. The fringe based labeling selection is developed from pool-based uncertainty sampling setting in classic active learning literature. Iterative Feature update approach is an intuitive method based on Rocchio’s algorithm[1] of relevance feedback.

We modify the method by taking into account the relative similarity measurements. In other words, features of fringe elements are updated according to query results, meaning fringe element would be moved closer to similar centroids and further away from dissimilar centroids.

Results
We present our results from two perspectives, firstly, how well could the crowd compare functional aspects, secondly how well is our proposed active learning algorithm performs.

1. Crowdsourcing Platform
Applying the interactive website with explicit crowdsourcing, we have collected judgements from both 2500 crowd workers and 35 domain experts, labelling 780 and 204 region pairs respectively. Domain experts could perceive more relevant functional aspects of urban spaces, with an positive correlation of 0.636 with ground truth, while crowd workers has a less significant correlation of 0.369.

2. Algorithm Performance
We apply our model to 40 most frequent visited regions in Seattle and conduct functional clustering. Our active learning model outperforms baselines including random policy, active learning with binary similarity[2], etc. Our model is robust even given incomplete users’ judgements.

Conclusion/Perspectives
1. Our active learning algorithm has succesfully exploited the functional labels from our crowdsourcing platform. Our clustering has a good performance even given incomplete or noisy human oracles
2. Further work can be done to perform functional clustering on the entire dataset or decrease noisy human oracles.

Acknowledgments
The work is conducted at Wearable Computing Group, ETH Zurich.

References