



SOUTHERN UNIVERSITY OF SCIENCE AND
TECHNOLOGY

CS212 INTELLIGENT DATA ANALYSIS

Director: Prof. Peter Tiño

Investigating Properties of Houses from a New York Airbnb Dataset

Author:

Edward Yidong FANG

Student Number:

11510493

Email: 11510493@mail.sustc.edu.cn

Shenzhen, China June 29, 2017

Abstract

In these experiments, three basic methods of data analysis are implemented, which is Principle Components Analysis (PCA), Clustering and Self-organizing Map (SOM). By applying there methods, the relations among different attributes are discussed. After the dimension reduction, it was found that the attributes can be predicted by others to some extent, while some can not. The relations will be presented mostly as different figures.

Contents

1	Introduction	2
2	Data Preprocessing	2
3	Data Description	2

1 Introduction

Airbnb is an online marketplace and hospitality service, enabling people to lease or rent short-term lodging including vacation rentals, apartment rentals, homestays, hostel beds, or hotel rooms. The dataset analysed in there experiments is derived from Tom Slee's blog and it is crawled from the website of Airbnb. And only a small part of data for New York, crawled on 05/06/2017 are analysed.

2 Data Preprocessing

2.1 Data Description