Fine-Grained Population Estimation

Simon Weidner

Chair of Algorithms and Data Structures

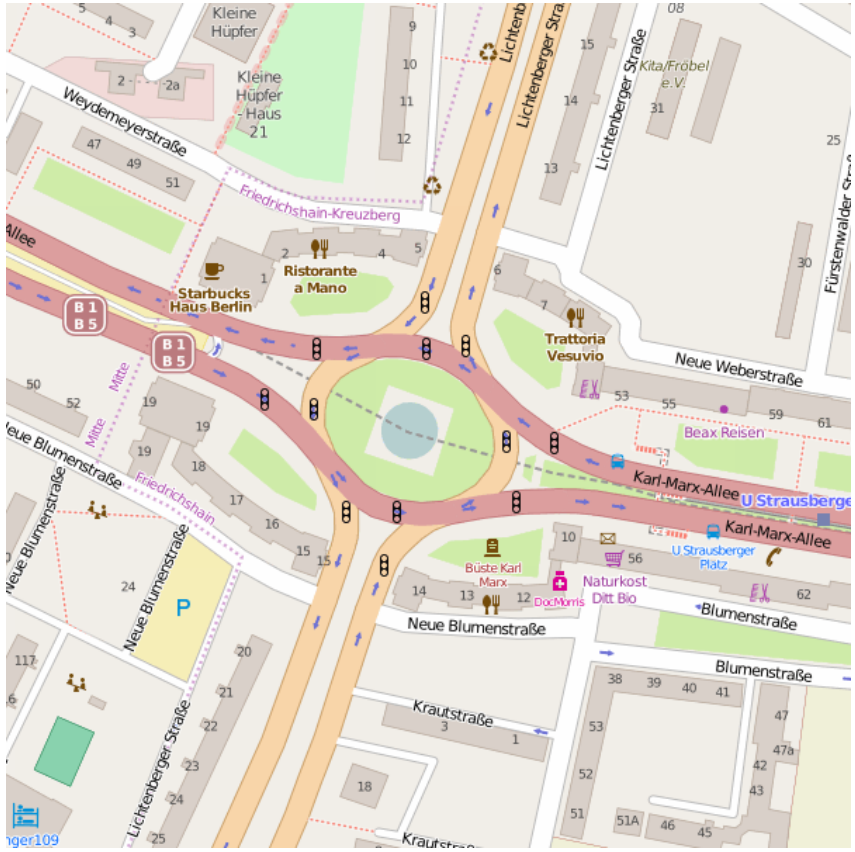
Albert-Ludwigs-University, 2015

Introduction

- Estimation of population numbers
- Using environmental information
- OpenStreetMap as data set

OpenStreetMap

- Free editable map of the world
- Three structures: Nodes, Ways, Relations
- Map Features represented by key value pairs



Problems with the OpenStreetMaps data

- Varying quality
- Misuse of Map Features
- Incomplete data

Our Approach

- Extracting the OpenStreetMap data
  - Extracting information
  - Preparing information
- Improving the underlying data
  - Area classification
  - Building classification
- Predicting population numbers
  - Distribute census data among building
  - Use machine learning

Data Extraction

- Fix misshaped buildings
- Connect information from Nodes to buildings
- Find informations about the location of a building

Area Classification

Idea

Distinguish between different types of areas, especially between residential and non residential

- Area describing Map Feature: Landuse
- Not all buildings are within an area with a specified Landuse
- Use Machine Learning to compensate

Area Classification

- Classify with logistic regression
- Two groups of buildings
- Find properties describing a certain type of area
  - Map Features in residential areas: Schools, playgrounds and parks
  - Map Features in commercial areas: Craft producers, shops

Building Classification

Idea

Distinguish between different types of buildings by population density

- Three categories: Non residential buildings, single family houses and apartment buildings
- Separate residential from non residential buildings
- Further split residential buildings into single family and apartment buildings

Building Classification - Residential or not

- Classify with logistic regression
- Use Map Features and residential areas to generate the samples
- Use properties describing certain type of area a building resides
  - Map Features as: Schools, playgrounds, parks, leisure facilities, craft producers, shops and more
- Use the type of area where the building resides

Building Classification - Single Family or Apartment Building

- Classify with logistic regression
- Use Map Features to generate samples
- Use the same Map Features as in the last step
- Search for Map Features in multiple ranges
- Use the buildings size

Population Estimation

Goal

Find population values for particular buildings

- OpenGeoDB provides population values
- Distribute the population among buildings
- Estimate the remainder with logistic regression

Population Distribution

- Distribute population values among all buildings of a certain area

$$Population(i) = \frac{weight(i) \cdot area(i)}{\sum_{b \in buildings} weight(b) \cdot area(b)} \cdot totalPopulation$$

- Weight is determined by the buildings location, its type and its purpose

Population Estimation

- Predict with linear regression
- Map Features as shops, supermarkets, parks, leisure facilities
- Features of the building as its size and if there is some facility or a shop within
- Previously obtained information as type of the area and type of the building

Evaluation

- Area Classification
  - 87% coverage in the OpenStreetMap
  - 83% precision
- Building Classification - Residential or non Residential
  - 78% of all buildings are classified before learning
  - 90% precision
- Building Classification - Single Family or Apartment Building
  - Only 6% of all buildings are classified before learning
  - 91% of all residential buildings are single family houses
  - 66% of all residential buildings should be single family houses

Evaluation

- Population value fits for Germany as a whole
- In most instances too high population numbers
- Results for regions of different sizes
  - Average error of 27% for large cities
  - 31% for medium sized and small cities
  - 29% for villages and urban districts
- All results are within the factor two of the optimum

Evaluation

- Hamlets and villages are populated
- Buildings within industrial and commercial areas are almost never populated
- Multi-part buildings are often partly populated
- Lack of apartment buildings, especially in villages
- High-rise buildings have too low population numbers
- Population numbers for single family houses fit