## Liverpool CDT Cohort

Melanie Green, Natalie Rose, Krasen Samardzhiev, Celine Chalupa, Nikos Patias, Susie Philp





CDT Data Analytics and Society Workshop 18<sup>th</sup> September 2018

Geographic Data Science Lab Department of Geography and Planning University of Liverpool



Data Analytics & Society

## Our projects

- Topological data analysis of big spatio-temporal urban data
- Weather and the impact on high street retail
- Comparing socio-economic characteristics to the built environment using high resolution aerial imagery
- Neighbourhood trajectories in the UK
- Sensing dynamic retail environments
- Neighborhood trajectories from planning data







Providing construction intelligence to UK businesses



## Modules

## **Compulsory Modules**

- Social Survey Analysis
  - Descriptive statistics
  - Inferential statistics
  - Correlation and regression
- Qualitative Research Methods
  - Research design
  - Various qualitative methodologies e.g. ethnography, interviews
  - Thematic coding

## Geographic Data Science

Deprivation in Manchester

- Visualisation
- Choropleth mapping
- Spatial weighting
- Clustering





## **Big Data Analysis**

- Probabilistic modelling
- Scalable algorithms
- Real-world applications of batch analytics
- Sequential Bayesian Inference
- Streaming analytics





## **Computational Intelligence**

- Structure and learning processes of neural networks
- Different types of neural networks
- Genetic algorithms



### **Spatial Analysis**

- Random intercept and random slope multilevel modelling
- Spatial autocorrelation
- Geographically weighted regression
- Spatial interpolation, regression and flows estimation





Random intercept



Random slope

## **Spatial Analysis**

- Random intercept and random slope multilevel modelling
- Spatial autocorrelation
- Geographically weighted regression
- Spatial interpolation, regression and flows estimation



Local I Cluster Map for Unemployment





## **Spatial Analysis**

- Random intercept and random slope multilevel modelling
- Spatial autocorrelation
- Geographically weighted regression
- Spatial interpolation, regression and flows estimation



Predictive check - Model 3



## Internships

## Exploring the weather dependency of different product categories

#### Aim

To determine what product categories are weather dependent and explore the nature of the relationships they present

#### **Objectives**

- 1) Identify the product category sales that are most dependent on each weather condition
- Determine the nature of 2) these relationships

#### Data

Daily sales data: in-store and online

Store characteristics

Daily weather data at store level

#### O1: Random forest regression models



#### For (i in 1:48) {

set.seed(1) results[[i]] <- ranger(categories[,i] - MAX\_TEMP</pre> STORE\_NUMBER + STORE\_TOWN + STORE\_COUNTY + TYPE\_OF\_STORE\_DESCRIPTION + day\_of\_week + day\_of\_year, data = store\_purchases.2, importance="impurity") #print(i)

### list\_max[i] <- results[[i]]\$r.squared

#### O2: Partial dependence plots for skincare products and sun preps











# Investigating the spatial and temporal trends in footfall data

 Visualization techniques and k-means clustering were used to develop understanding of the patterns within footfall.







## Topological data analysis of workplace area characteristics data





<matplotlib.collections.CircleCollection at 0x7ff27c072518>)





## Red Ninja – Viability of LiFE 1.0



- North West Ambulance Services & Liverpool's traffic control center
- Improve ambulance response times
- Solution: Al Algorithm
  - Real time city congestion data
  - Ambulance location data
- Junction Types Up to 40% reduction in journey times



Research on Ambulance accidents & response times

Ambulance Quality Indicators (NHS)

Availability of data sets

• Found ambulance crashes through attributes

Predicting accident severity in Ambulance crashes Possible predictors:

- Junction type
- Road conditions (wet, dry, ... )
- Light Conditions
- Urban vs. Rural area

Develop similar application for freight transport Save on time, fuel cost & cut down emissions

Map based on Longitude and Latitude. Color shows details about Junction Control. The data is filtered on Action (Ambulance, Junction Control), which keeps 10 members.

## Exploratory image analysis

#### Aim

One of the recent research interests of Ordnance Survey is to test methods for automatically finding windows (and doors) within façade imagery.

<u>Key question</u>: A representative sample of the windows is required (and doors) to be used as templates in order to perform further analysis (i.e. identify different styles)

#### <u>Tasks</u>:

- Process images through a neural network and produce txt file outputs by layer/image/filter
- Write code for clustering the layer file outputs
- Find optimal K for clustering
- Source central images to each cluster
- Manually check their viability to be used as templates for next method
- Dimensionality reduction to be used for visualisation

#### **Dimensionality reduction**

 t-Distributed Stochastic Neighbour Embedding (t-SNE) method applied to reduce dimensions from (25-2000) to 2 (X and Y)











Deeper layer of neural network output

## Classification of aerial images using features extracted from a neural network

- Related to Ordnance Survey ImageLearn project investigating the  $\bullet$ potential of machine learning to reduce manual survey costs and enhance data products
- Extracted features from images at various layers of a neural network  $\bullet$ and trained a Support Vector Machine to classify as inland water or not
- Examined the classification accuracy for each layer of the network  $\bullet$ and each set of weights





trained



Layer 94, ImageNet

Layer 104, scratch

Layer 94, fine-tuned

## PhD life

- Weekly lab meetings
- Monthly supervisor meetings
- Demonstrating undergraduate modules
- Training courses
  - Secure data training
  - R for spatial analysis
  - Tableau
- Events/Conferences
  - Liverpool School of Environmental Sciences PGR conference
  - Ordnance Survey PhD conference and neural network hack
  - Data science conferences
  - Summer schools