# Literature Review

* 2000 words

*Introduction:* This section reviews the current understanding in the literature surrounding mode classification and transport purpose. It then examines volunteered geographic information and how it has been harnessed in research with additional focus on the city of Montreal as a study area. Finally, this section provides an evaluation of current methods, including machine learning, employed in the space-time realm.

*1. Mode-Classification and Purpose*

Although there is a wealth of literature on both mode-classification and . There exists a lack of pertinent literature to the study of how people travel for around a city for certain activities. Arguably this owes to a lack of data set available to examine this topic.

Space-time investigations in general

individuals’ mobility is found to be highly regular (Lin & Hsu, 2014)

*Transport*

Transportation is a classic social dilemma where individually rational behaviour (being mobile) leads to collectively irrational outcomes such as congestion (Miller, 2013)

*Volunteered Geographic Information (VGI)*

‘People as sensors’. Li *et al.* (2016) distinguish between participatory (conscious inclusion of their data i.e. like MTL Trajet) and opportunistic (unconscious) forms of VGI.

Can see space-time events (i.e. Traffic Jam, Crowds, THINK OF MORE etc.).

Using twitter and sentiment analysis (similar to purpose)

Governments pushing for greater capacity of transport networks rather than efficiency (Attard *et al.*, 2016).

(Attard et al., 2016 referencing Lyons & Harman, 2002). → 1. There are issues with trust over the information provided, 2. Lifestyle changes are opportunities for travel behaviour change, 3a. People have very poor judgement of cost and time when travelling by car with control over their journeys being seen as important. 3b. Public transport in contrast, is seen as difficult as information is sought from unfamiliar and uncertain sources.

VGI requires rethinking of geographical concepts (Elwood *et al.*, 2012).

Many studies have shown that voluntary contributions by individuals follow a frequency distribution with a long tail, with a few individuals making large numbers of contributions (Goodchild & Li, 2012)

Liu & Cheng (2018) conjoins socio-economic data to transit patterns to interpret behaviour

using temporal clusters from LDA and temporal words (Liu & Cheng, 2018)

Liu & Cheng (2018) Looks at who constitutes each temporal cluster (which socio-economic groups)

*Transport mode detection*

A significant amount of literature exists for transport-mode detection. Something which is of prime concern to companies utilising spatial information derived from GPS data. Determining transport mode through the use of deep-neural networks such as those with convolutional layers. These networks .

Dabiri & Heaslip (2018) raw GPS to mode

Bantis & Haworth (2017) socio-demographics and how you travel. Environmental and social factors affect the way you travel. Although data used in this study is not

Striving to include meaning to space (POI)

Bantis & Haworth (2017) overlay of GPS tracks and underlying LSOA socio-demographic information

Inferring employment status (Zhang & Cheng, n.d.)

high quality traveller information for public transport is undoubtedly crucial for Government’s transport policies (Lyons & Harman, 2002)

*Network Research*

The understanding of human mobility in an urban space has become crucial to optimize the network management (Jahromi *et al.*, 2016)

Size of space-time neighbourhoods change (Cheng *et al.*, 2014)

*2. (Smart) City Research, VGI and Montreal*

*City research*

Both fast and slow dynamics/flows have an impact on changes in a variable across space, accounting for these is essential in modelling [i.e. differencing to remove long-term trend] (Batty, 2013).

Understanding the interactions within complex system such as a city is a prerequisite for predicting changes within it… big data offers us an opportunity to study this (Cheng *et al.*, 2017)

Flows generate change immediately whereas the ultimate locational redistribution takes longer to work itself out. [In reality] this process of working out is implicit and the ultimate equilibrium that occurs is a product of both fast and slow processes with no explicit time scale. (Batty, 2013)

\* Jahromi *et al.* (2016) try to simulate GPS movement/mobility with purpose so that infer about interactions of people with a city and its services

*On Neighbourhood effect ->* People often traverse neighbourhoods and boundaries throughout one day (Kwan, 2018)

*Volunteered Geographic Information (VGI)*

‘People as sensors’. Li *et al.* (2016) distinguish between participatory (conscious inclusion of their data i.e. like MTL Trajet) and opportunistic (unconscious) forms of VGI.

Can see space-time events (i.e. Traffic Jam, Crowds, THINK OF MORE etc.).

Using twitter and sentiment analysis (similar to purpose)

problems of representativeness in VGI (Li *et al.*, 2016; ref; ref)

*Montreal:*

The city of Montreal, located within the Greater Montreal region is the largest city in Quebec with a population size of … . It is of particular interest to city transport research due to its unique road and public transport networks. There are a total of

[Map of the Island Montreal within the Greater Montreal Region and then Quebec]

Eluru *et al.* (2012) survey to look at public transport and how people move around Montreal. Trying to encourage it

*MTL Trajet & Similar Survey Projects*

Data collection [or survey data] driven by data availability or convenience of data collection rather than by domain knowledge, theory, or insight into the process(es) of interest. (An *et al.*, 2015). Many previous studies have proposed and validated in limited geographical areas, such as campuses (Jahromi *et al.*, 2016)

Fallah-Shorshani (2018) uses MTL to look at pollution exposure

DataMobile Smartphone Travel Survey example from 2014 carried out by Concordia University in Montreal (Patterson & Fitzsimmons) 10 November - 5 December 2014. Close to 900 people participated in the survey

Iternerium Platform

*3. Metrics and Space-time Investigation*

*Lead onto class identification in transport*

Significant class-imbalance exists in the MTL Trajet data

*Space-Time*

ST-KDE and space-time decomposition to compute space-time methodologies (Hohl *et al.*, 2016).

Spatio-temporal modeling has always been of interest to researchers in geographical information science (GIS) (Ren *et al.*, 2019).

CNN was not optimal for modeling the spatial patterns of the road because the authors did not consider the road network topology (i.e. direction) (Ren *et al.*, 2019 in their literature review)

STARIMA and ANN (Cheng & Wang, 2011)

Transport Forecasting and modelling (Yue & Yeh, 2008; Cheng & Wang, 2011)

network and topological models not well suited for handling geospatial big data (Li *et al.*, 2016)

Paradigm shift (Brunsdon, 2015)

Space–time analysis seeks to understand when and where (and sometimes why) things occur. (An *et al.*, 2015).

there is not a close coupling between big data and space-time methods used to analyse them (An *et al.*, 2015)

Spatial Weights (Anselin & Rey, 2014)

Determining threshold of time and space (Adepeju & Evans, 2018)

MTUP (Cheng & Adepeju, 2014)

PYSAL

*Big Data*

[On big data] However, they are often collected on a fairly ad-hoc basis when compared with traditional data sources, and usually must be repurposed to fulfil research objectives (Li *et al.*, 2016)

[On visualising big data maps] New computational and technical paradigms for cartography are accompanying the rise of geospatial big data (Robinson *et al.*, 2017)

*ABM:*

ABM a great tool but ABMs could be very data demanding and sometimes too complex without offering much additional insight (An *et al.*, 2015).

*Machine Learning:*

methodologies originally developed to analyse small data and are not necessarily equipped to tackle some of the distinctive features of newer sources (Gorman, 2013; Arribas-Bel, 2017).

However, methods to assess quality assurance of the geospatial data still relevant on big geospatial data as they still describe the same processes (Li *et al.*, 2016)

Big geospatial data visualisations do not always scale well – because they can become messy (Li *et al.*, 2016)

Li *et al.* (2016) challenges in dealing with big geospatial data, reviews if those traditional methods still useful for data

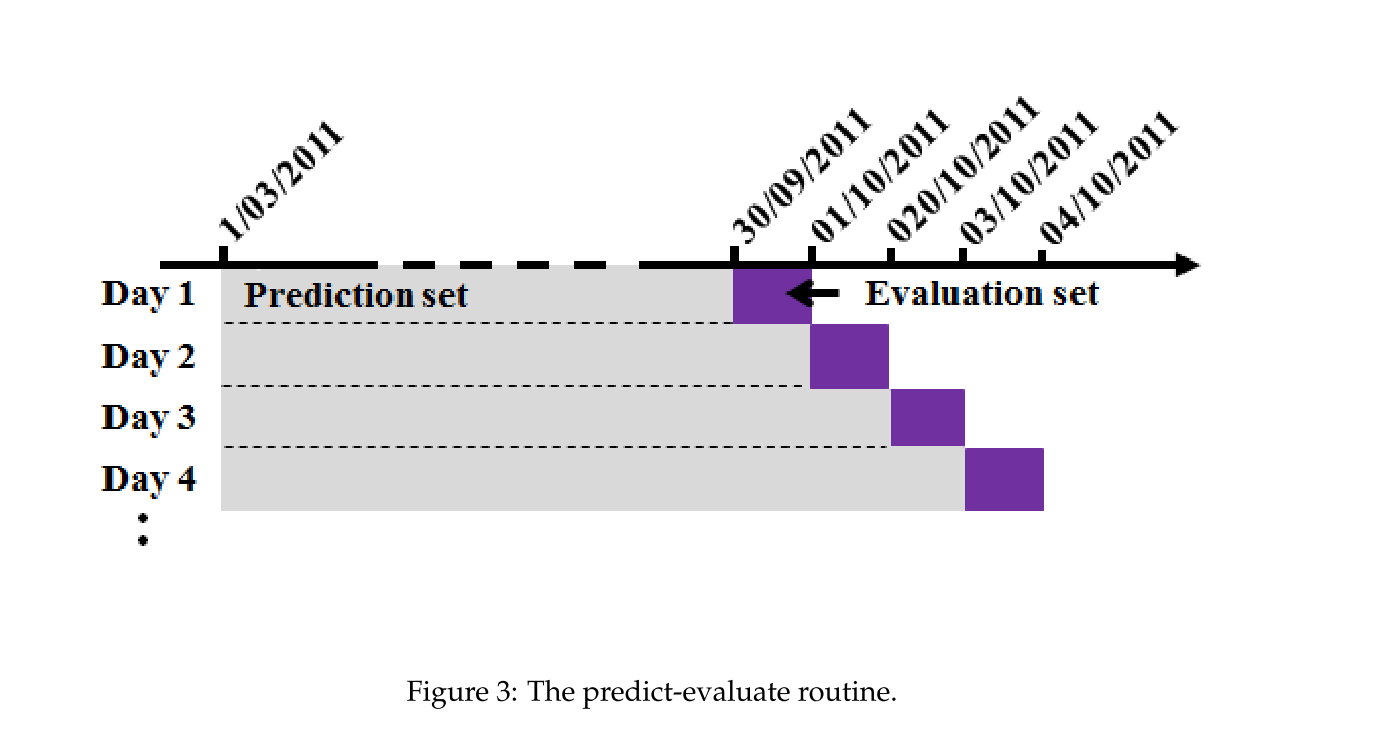
﻿ML methods are generally effective in tackling nonlinearity in spatial data (Li *et al.,* 2016)

Spatial statistics is well suited to handle big data. It offers capabilities to summarize the data, and express measures of variation and uncertainty. (Li *et al.*, 2016)

Machine Learning are often applied in detecting transportation mode (Gong *et al.*, 2014)

[With big geospatial data] validity or trust is traded for the velocity of information production (Li *et al.*, 2016)

Space-Time CNN -> CNN-based models employ a grid map to represent spatial data which is unsuitable for road- network-based data. (Ren *et al.*, 2019)\*



Source: Adepeju & Evans (2018)

*Similar Studies:*

*Similar datasets:*

* NY Taxi-cab movements
* Uber-movements

*Similar Methods:*

* Maybe classify mode and where people are going? (after Bantis & Haworth, 2017)