Chapter 1

Introduction

1.1 Background

The current world population of 7.6 billion keeps increasing enormously; it is expected to reach 8.6 billion in 2030, 9.8 billion in 2050 and 11.2 billion in 2100.[1] The earth is gradually running out of (its limited amount of) natural resources while energy demand is still increasing. Between today and 2040, the global energy demand will expand by 30%[2], which directly affects greenhouse gas emissions and climate change. For instance, the global mean of surface temperature has increased by 0.6°C since the start of the 20th century. This asks for a move towards urban sustainability, and specifically a reduction in energy consumption [33].

Europe's 2030 Energy Strategy targets a 40% cut in greenhouse gas emissions compared to 1990 levels, at least a 27% share of renewable energy consumption, and at least 27% energy savings compared with the business-as-usual scenario.[3] In order to meet this targets, energy policies and programmes should be formed and individuals (i.e., citizens) should be motivated to change their energy consumption behavior [21], both in terms of energy conservation and energy efficiency. Energy conservation involves saving energy by reducing or omitting an activity; for instance, turning a light off or reducing the time one watches television. This calls upon a behavioral change by the individual. On the other hand, energy efficiency involves using less energy to provide the same service; for instance, replacing a single pane window in the house with an energy-efficient one. Energy consumption is significantly characterized by behavioral aspects [55], which is why understanding and changing the energy consumption behavior of individuals is considered as a powerful approach to improve energy conservation and stimulate energy efficiency at the individual level [42].

Multiple studies have examined how energy efficiency and conservation could be motivated among policy makers and citizens. In [19] the author explains how comparative feedback on energy usage with others can generate feelings of competition, social comparison or social pressure, which appears to be more effective in motivating energy conservation than temporal self-comparisons. The author of [30] endorses this

[1] https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html
[2] https://www.iea.org/weo2017
[3] https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2030-energy-strategy

in his Social Electricity case study, which “allows people to compare their energy footprint with other online peers or with the consumption at their neighbourhood, village or town, to perceive if their own consumption is low, average or high". Multiple energy saving applications ([21]) have been developed yet, using visualized consumption feedback and gamified social interactions to motivate people to adopt energy-efficient lifestyles.

Before we can motivate individuals to change their energy consumption behavior, we need a thorough understanding of why and how (through which activities) they consume energy. In order to thoroughly understand this energy consumption activity, insights into the individual’s activities behind the energy consumption should be gathered at a high-granular level.

1.1.1 Traditional Data Sources for Energy Consumption Activity

These days, multiple data sources are used to provide insights into energy consumption activity, including (governmental) energy consumption surveys, smart meters, and smart plugs. Nevertheless, these mainly focus on residential energy consumption activity (i.e., activities within the home) without taking external activity (activities outside the home) into account. Hence, most of the scientific research in energy consumption activity has studied consumption at the household ([42, 47, 55]) or building ([12, 45]) level.

Smart sensor data (derived from smart meters and smart plugs) can be used to provide insights into domestic energy consumption. It focuses on aggregate energy consumption; however, by using specific techniques insights into disaggregated end-use energy data (down to the individual appliance or device, which can be related to the individual’s domestic activities) can be provided, which helps individuals understand how energy is consumed in the home. These techniques involve discriminating between appliances based on the total power consumed by each device, or differentiating by different features (in current wave forms or transient voltage noise signatures) during the device start-up [23, 41, 51]. However, since the energy usage is affected by a lot of variables, such as the number of appliances that are active simultaneously, it is questionable how well this approach reflects the actual energy usage. [23]

In addition, survey or questionnaire research ([49, 50, 9]) is often used to gather insights into energy consumption activity. Similarly to smart sensors, these data sources merely focus on the domestic energy consumption. Through a variety of questions, the energy consumption activity is broken down into different end-uses. Yet, these data sources are not readily (i.e., real-time) available, since the surveys are conducted periodically (e.g., monthly or annually). This makes it very hard to model energy consumption activity at a high-granular level, which is needed for a thorough understanding of the individual’s energy consumption activity.

Moreover, these numbers derived from smart sensor and survey data do only take direct energy usage into account. Indirect energy usage (i.e., energy usage that is “related to the production, transportation and disposal of a variety of consumer goods and services", such as “the availability of meat or cheese" [2]) is neglected [11]. Yet, conservation and efficiency of indirect energy consumption may also lead to a significant reduce in energy demand and should thereby also be integrated in the energy consumption activity models. Analysis of ecological (and carbon) footprint does take indirect energy usage into account though. For instance, carbon footprint calculators (e.g., the ones by WWF[4] or The Nature Conservancy[5]) estimate an individual's environmental footprint by assessing living habits in the domains of home, food, travel, and shopping. However, these footprints are not standardized yet, and the corresponding calculators, which are the principle tools, lack consistency and calculate different results [16]. Besides the calculators, scientific models for ecological footprint analysis have been proposed. Yet, these only perform analyses at the city level and do not model individual footprints [24, 37].

Opposed to smart sensor and survey data sources, which focus on the domestic energy consumption, there are a lot of data sources that might provide relevant information about outdoor energy consumption activity (such as supermarket data on daily groceries, financial transactions, etc.). However, these are not publicly accessible.

1.1.2 Social Media Data Sources as Sensors for Energy Consumption Activity

Social media data sources (including online social networks) arise as an alternative and novel approach, and typically do not face the above-mentioned issues that traditional data sources are coping with. Hundreds of millions of people frequently use social media to share, communicate, connect, interact, and create user-generated data, which makes social media an extraordinary source of big data [52]. Social media data sources are scalable, publicly available, and have lower setup and maintenance costs compared to certain physical sensors such as smart meters. Furthermore, real-time data can be obtained (i.e., they have a frequent update rate), and most social media data is tagged to space and time (hence, they include spatial and temporal dimensions) and allows for analysis at the individual level. In general, social media data contains a lot of metadata (about the post, user, checked-in location, etc.) and is thereby (semantically) very rich, which aids in extracting meaningful information out of it.

Because of its rich and dynamic data, social media has proven to be a good source for human (daily) activity recognition [8, 57]. If we consider the user to be a sensor, his/her corresponding social media data (in the form of textual or visual content, geolocation, and time) are signals that can be utilized for recognizing main activities. Since social media allows users to create posts about both domestic and outdoor activities, we might be able to not only capture information about the individual's domestic energy consumption activity but also about his/her energy consumption activity outside the home. Hence, social media data sources seem to be a good complementary source of information for describing energy consumption activity.

However, social media data is generated by people with other intentions than creating information about their energy consumption activity. Thus, the purpose of social media data sources differs from the data sources that are designed to analyze energy consumption activity, such as smart meters and plugs, and energy surveys. Moreover, daily life activities are rarely posted online at such high frequency, which makes it hard to get a complete overview of the individual's daily activities. In addition,

[4] http://footprint.wwf.org.uk/

[5] https://www.nature.org/greenliving/carboncalculator/index.htm

users do not only create social media posts about their daily activities but also use social media to communicate with others, or express their feelings and interests. Thus, social media data often contains a lot of insignificant, irrelevant information, which makes it hard to separate signal (relevant content) from noise (irrelevant content) [10].

1.2 Problem Statement

A variety of data sources have been deployed to provide insights into individuals' energy consumption activity. The traditional ones (smart sensor devices and energy surveys) merely focus on the energy consumption activity at the domestic level; these only capture activity at the home. Outdoor activities related to energy consumption fall outside the scope of these data sources. The same applies for an individual's activity that involves indirect energy consumption (e.g., consuming a piece of meat during dinner); this activity can not be captured by traditional data sources either. In order to cope with those shortcomings, social media data sources emerge as a complementary source of information. Social media has yet been used to extract meaningful information about user behavioral patterns, such as food nutrition patterns ([1, 22, 44]), and user transport and activity patterns ([8, 43, 57]), which makes it a promising source to model outdoor energy consumption activity at the individual level. Ultimately, a framework is desired in which both traditional and social media data sources are integrated, which allows us to gather insights into the individual's entire energy consumption activity, at a high-granular level.

In order to gather those insights, we need a better understanding of the domain of energy consumption activity, its main characteristics and all different instances (including the different types of activities) that are related to energy consumption activity. We need to comprehend the meaning of those instances, and how they are mutually related. In addition, a definition is needed of how social media (instance) data relates to the physical world, and how it may reflect the individual's energy consumption activity.

However, as mentioned before, multiple challenges will be faced when using social media as our data source for describing energy consumption activity. One of the greatest challenges is to extract meaningful information from social media data. Social media data is often noisy; users might include shorthand or slang in their messages, words or expressions might vary in meaning depending on the context, and content might contain information that stretches over multiple locations [28]. Since social media data contains a lot of irrelevant information, it is hard to separate the signal (relevant information) from the noise (irrelevant information). Moreover, social media data is often biased, which makes it very important to understand the degree and nature of this bias. For instance, social media platforms are particularly used by the younger generations (i.e., selection bias) [28]. Besides that, a user's social media post might not always be a reflection of his/her daily activities in the physical world, due to multiple semantic ambiguities and discrepancies that may be encountered. A user may create a post before or after an activity instead of during the activity itself; in figure 1.1 the user recalls an old memory of a time (one year ago) when she was at a concert (figure 1.1). In figure 1.2 the user wrote a message about the tramway in Amsterdam, along with a picture of a tram; however, this does not necessarily mean that the user indeed travelled by (this) tram. He might have wanted to capture a typical Amsterdam street view instead. Also, in the case of figure 1.2, it is clear that the mention of Amsterdam in the message refers to the capital of the Netherlands, since the user checked in there. Amsterdam is also a town in the state of New York (United States of America) though. Hence, it might also be the case that a user–when he/she has not checked into a location–refers to the town of Amsterdam in the state of New York instead of the Dutch capital (ambiguity in location). It is very important to be aware of these types of social media bias and semantic ambiguities and discrepancies when analyzing and evaluating the data.

Figure 1.1: Discrepancy in time

1.3 Research Aim, Objectives and Scope

Given the challenges mentioned in the previous section, the aim of this work is to design a framework for the analysis, integration, and visualization of social media data to facilitate the understanding of individual energy consumption activity.

1.3.1 Contributions

To meet this aim, four objectives (or contributions) for this research have been posed:

- A literature review will be conducted in order to create an overview of how the state of the art studies energy consumption activity to determine which methods and tools to include in our framework.

Figure 1.2: Ambiguity in content

- A conceptual data model of energy consumption activity will be designed to facilitate the definition of its main characteristics. Moreover, it will support in the integration of multiple social media data sources.

- A pipeline will be developed in order to extract characteristics from social media data; this pipeline will include (i) the data collection from the social media data sources, (ii) the pre-processing of the collected data, (iii) the Machine Learning algorithms that allow to classify the data regarding the defined characteristics of energy consumption activity, (iv) the (semantic) enrichment of the data, and (v) the data exploration and visualization methods that allow to extract meaningful information from the data, which will contribute to a more thorough understanding of the user's energy consumption activity. Multiple dimensions of analysis (among others, space and time) will be taken into account.

- Subsequently, the framework will be evaluated through an experimental case study performed for the social media activity in the city of Amsterdam. The results will be visualized and the performance will be evaluated through several metrics such as accuracy, and precision and recall.

1.3.2 Scope

The scope of this work will focus on four types of energy consumption activity; based on previous work ([6, 39, 25]), we define the following types of activity: dwelling, mobility, food consumption and leisure. As for the geographical boundaries, we will limit our study to the Netherlands, to keep the (cultural) circumstances adequately similar (and thereby, avoid possible significant differences between train and test data).

Furthermore, we will study energy consumption activity at the individual level. Since the overarching goal in sustainability studies is to move towards urban sustainability (as mentioned in the introductory background section), we will mainly focus on urban individuals (and thereby, urban areas) within the Netherlands. Information gathered on the individual level can eventually be aggregated to provide insights into particular neighborhoods (or other (urban) areas).

With respect to the data sources, this work will be limited to data collection through Twitter and Instagram, mainly due to their public APIs. A more thorough analysis of the different data sources can be found in appendix B.

1.4 Research Questions

Our main research question follows from the research aim and objectives:

How can we describe user energy consumption activity using social media data?

To answer this overarching research question, four sub-questions have been posed:

1) **How is user energy consumption activity studied in the state of the art?**

   In order to determine the best methods and tools for describing and understanding user energy consumption activity, a literature review should be conducted to explore the state of the art in this field. Previous studies have yet explored several methods in user energy consumption activity, though very generic, or with a focus on a single domain of energy consumption activity [2, 11]. The strengths and weaknesses of the existing contributions should be identified to lay the groundwork for this research.

2) **What are the main characteristics of an individual's energy consumption activity?**

   For a better comprehension of the domain of energy consumption activity at the individual level, a conceptual data model will be developed, including all four types of activity (dwelling, mobility, food consumption, and leisure), that will facilitate the definition of the main characteristics of an individual's energy consumption activity. On the other hand, it will support the definition of how social media (instance) data relates to the physical world.

3) **How can we extract the energy consumption activity characteristics from social media data?**

   A pipeline containing multiple components (including data collection, (pre-)processing, enrichment, and classification) will be developed that allows to extract the energy consumption activity characteristics from the social media data. Here, we will also cope with many of the challenges that arise from social media data.

4) **To what extent can social media data sources serve as sensors to understand energy consumption activity?**

   In order to understand to what extent the extracted energy consumption activity characteristics provide insights into an individual's energy consumption activity, data exploration and visualization methods will be used to extract meaningful information from the data, e.g. in different dimensions such as space and time.

1.5 Research Design: Approach and Methods

In figure 1.3 our research design is illustrated using a schematic overview to visualize the research design structure and thesis outline. Each sub-question links to an action, which is represented by one of the numbered circles, comprising the approach and methods for that particular part. The number within the circle refers to the number of the corresponding chapter. The actions to be taken are (2) providing an overview of the *state of the art* in describing energy consumption activity, (3) defining the main *characteristics* of an individual's energy consumption activity, (4) *extraction* of the energy consumption activity characteristics from social media data, and (5) *understanding* the individual's energy consumption activity. The outcomes and findings of all parts together lead to the answer to the main research question.

In the first part of the research design a literature study (and review) will be conducted to explore the state of the art on energy consumption activity. Key findings, strengths and weaknesses of existing studies are compared and analyzed in order to identify which methods and tools are most relevant and promising and should be included in our framework.

The second part introduces a conceptual data model of an individual's energy consumption activity and its characteristics. At first, entity-relationship modeling has been the starting point for this conceptual data model. At a later stage (or in future work) this model can be transformed into an ontological representation in order to contribute to the Linked Data and the Semantic Web in the domain of energy consumption activity.

In the third part a variety of methods and tools will be used to design the different components of our pipeline, which is designed to extract the energy consumption characteristics from social media data. The "raw" data will be collected trough the public APIs of Twitter and Instagram. Regarding the pre-processing of the collected data, multiple techniques will be utilized that are applicable to either text (e.g., tokenization, stop word removal, stemming, etc.) or images (e.g., filtering, noise removal, thresholding, segmentation, etc.). Next, the data will be enriched and classified using state-of-the-art techniques and (machine learning) algorithms. The performance of each component will be determined based on several techniques, e.g. precision and recall. Then, meaningful information (including insights into the individual's energy consumption activity) is acquired through a rule-based approach which allows to reason about higher-order metadata. Different dimensions will be taken into account; temporal and spatial patterns will be examined along with a comparison between the different types (or domains) of energy consumption activity (dwelling, mobility, food consumption, and transport). Several data exploration and visualization techniques will be used to present the results.

In the fourth part the framework will be evaluated through an experimental case study. Test data from particular urban areas in the Netherlands will be acquired through several social media APIs. Then, our framework will be evaluated by testing our trained models on the acquired data, to find out to what extent social media data sources can serve as sensors to understand energy consumption activity.

1.6 Thesis Outline

Each of the six parts of the research design (figure 1.3) is affiliated with a chapter of the thesis. Chapters 2 through 5 address one of the four research sub-questions. The main research question is answered in chapter 6, as well as a discussion following from the answers to each sub-question.

- Chapter 1 discusses the background (regarding energy consumption) of the issue we aim to solve along with the aim, research questions (following up on the objectives), and scope of the research.
- Chapter 2 provides an overview of how the state of the art studies energy consumption activity. Existing methods and tools, including its strengths and weaknesses, will be reviewed.
- Chapter 3 introduces a conceptual data model of an individual's energy consumption activity and its characteristics. It will also shed light on the ties between energy consumption activity and social media activity.
- Chapter 4 proposes the design of a data processing pipeline, developed for the extraction of the energy consumption characteristics from social media data. Each of its components will be discussed along with its performance.
- Chapter 5 explores and visualizes the extracted information for social media data from urban areas in the Netherlands in order to acquire meaningful information and insights into the individual's energy consumption activity.
- Chapter 6 summarizes and discusses the findings for each sub-question, resulting in the explanation of our main research question. Subsequently, multiple possibilities for future research are suggested.

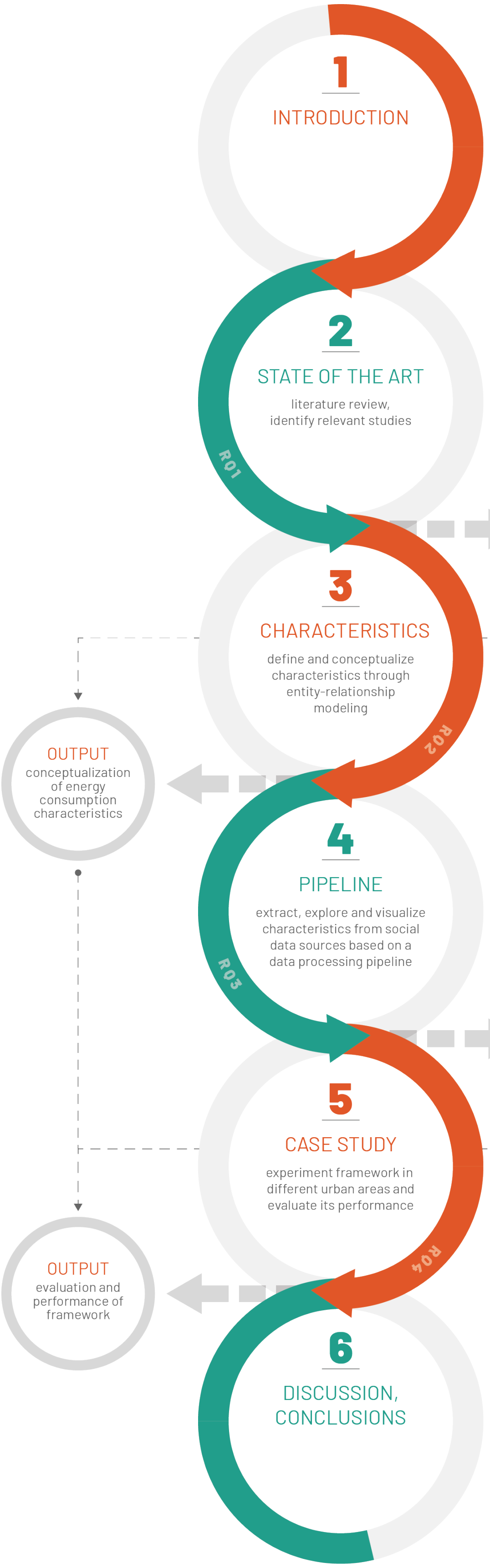

Figure 1.3: Schematic overview of the research design structure and thesis outline

Chapter 2

Related Work

In order to model citizens' energy consumption behavior, we should have a clear understanding of what this actually entails. Energy consumption can be separated into direct and indirect consumption. Direct energy consumption is related to the use of gas, electricity and fuel, whereas indirect energy consumption is related to the production, transportation and disposal of a diversity of consumer goods and services [2]. Moreover, four types of energy consumption activity can be distinguished based on previous work ([6, 39, 25]):

Dwelling the energy consumption necessary for all activities performed within the home, peculiarly by using appliances available in the house.

Food the energy consumption necessary for all stages of the food chain (production, processing, distribution, consumption and waste).

Leisure the energy consumption necessary for citizens to achieve the "freedom provided by the cessation of activities"[1].

Mobility the energy necessary to bring one person from one place of activity to another place of activity.

2.1 Energy Consumption Activity Modeling

Up to now, many studies have aimed to describe or model energy consumption activity. Many rely on survey data, which is the basic collection method for information on energy consumption [48]. In [4] the authors proposes a human-activity based residential energy modeling framework that can create power demand profiles considering the characteristics of household members, based on survey research data. The authors of [42] use home appliance energy use data for performing agent-based modelling (ABM) to represent the complexities of energy demand, such as social interactions and spatial constraints. In [9] the individual's environmental impact (assessed by different behavioral determinants of the ecological footprint) is compared with self-assessments

[1] www.merriam-webster.com

of their own environmental impact, retrieved from Belgian survey data. Furthermore, the authors of [49] analyze national time use survey data to assess how dependent energy-related social practices in the household (preparing food, washing, cleaning, washing clothes, watching TV and using a computer) are in relation to the time of the day.

Besides surveys, smart sensor data (derived from smart devices such as smart meters and plugs) have also emerged as a valuable source of information. In [47] the authors demonstrate how household characteristics related to energy efficiency can be extracted from smart electricity meter data by using supervised-machine-learning-based techniques. The authors of [21] also use algorithms to extract activity from sensor data to profile different types of user behavior and infer activity context.

Nevertheless, these studies all rely on traditional data sources such as surveys and (smart) sensors and solely focus on the activity at the domestic level. Recently, some researchers have proposed the integration of social media data sources as an alternative to the traditional ones. However, these studies took only terms directly related to energy into account ([46]) or examined the correlation between external events (such as Christmas) discovered through social media and the actual energy consumption. Thus, both do not examine citizens' energy consumption *activity* discovered through social media. Furthermore, most studies perform top-down modeling of energy consumption: based on the total numbers, the energy consumption signals are disaggregated into different end uses ([26, 34]), whereas we aim for bottom-up modeling (i.e., aggregating the energy consumption based on the end uses discovered through social media posts by citizens).

2.2 User Activity Recognition from Social Media Data

Up to now, there have been rarely any studies that have focused on recognizing the actual user energy consumption activity from social media data. In [10] Bodnar et al. propose a social media network-driven model that aims to approximate electricity utilization patterns from large-scale textual and geo-spatial social media data. Through a Bayesian process, topics are modeled for user posts that are compared to events and phenomena in the physical world; physical hardware systems are excluded here. Since we envision traditional and social media data sources as complementary sources of information, a framework that integrates both is desired. Furthermore, "real world energy utilization" is not defined prior to the case study and is only analyzed at the household level.

Moreover, there have been many studies that have examined and modeled user activity patterns in different fields than energy consumption though. For instance, previous studies have looked into nutrition patterns ([1, 3, 22]) and activity and mobility behavior ([17, 35, 43, 54, 56]). These topics, nutrition, mobility and activities, can all be related to energy consumption activity in some way. Nutrition is affiliated with food consumption; thereby, it is indirectly also associated with energy consumption. The same reasoning holds for mobility and activities; it is not possible to travel or to perform another activity without consuming energy. For that reason–despite the fact that the purpose of these studies is different than our aim to describe user energy consumption activity–the findings of these studies are relevant for this research as well.

For each type of energy consumption activity (dwelling, mobility, food consumption, and leisure) the most important and relevant findings are discussed below and summarized in table 2.1.

2.2.1 Dwelling

Many traditional data sources focus on residential energy consumption, which is affiliated to dwelling. Nevertheless, many residential activities are not exclusively related to dwelling but could also be associated with leisure or food consumption. For instance, cooking (at home) is an activity that is related to both dwelling and food consumption. Existing literature ([4, 10, 21]) does not distinguish these different types of energy consumption. Moreover, many studies examine the total residential energy consumption and do not disaggregate the consumption into different end uses, which makes it hard to obtain insights into the corresponding energy consumption activity. Additionally, existing work barely incorporates social media data in their models and does not focus on the individual level.

2.2.2 Mobility

Nowadays, location-based social networks (LBSNs) allow users to share geo-tagged information along with the content (text, images, videos, etc.) they post. A check-in is one of the ways a user can share such geo-tagged information. In previous studies users' check-in behavior (and corresponding geolocated tweets) are analyzed for different purposes: to model travel demand and behavior ([43, 54]), to detect traffic oddities [40], to predict turning points in migration trends ([53]), or to predict user activities ([35, 56]). In [32] all algorithms related to LBSNs are elucidated by means of a survey.

For this research, we are interested in the purposes of both modeling travel demand and behavior, and predicting user activities. The latter one will be discussed in more detail in the paragraph on leisure below. Regarding the travel behavior modeling, not only spatial-temporal information but also semantic information is utilized. The check-in data is often enriched with activity information, such as the category of the location or point of interest (POI), as seen in the approaches of [17, 35, 56].

In [43] the author describes that the mode of transport should be determined using text mining and natural language processing approaches (by constructing a dictionary). As for duration of the activity, either text mining or natural language processing approaches should be utilized, or information for this travel attribute should be extracted by considering multiple tweets about the same trip joined together as a chain of tweets. Information for these two travel attributes, duration and mode of transport, have not been extracted and considered yet in the previously mentioned work. In the case that the consumed energy per mobility activity should be determined, these attributes do actually have to be taken into account.

2.2.3 Food consumption

Recently, a lot of progress has been made regarding food detection and recognition in text and images. Food recognition in text is somewhat less common nowadays;

natural language processing in combination with topic modeling ([22]) or Naive Bayes classification in combination with n-gram matching ([1]) approaches have been used in previous work. With regard to food recognition in images, the use of convolutional neural networks (CNNs) occurs to be one of the most frequently used approaches ([3, 13, 14, 31, 36, 44]).

In [13] the author describes an approach in which neural networks are utilized to jointly consider food recognition, ingredient recognition, and cooking method recognition. Verb-noun pairs are generated that indicate both the type of food and how it was cooked. Considering all those recognition factors could ease the process of determining the amount of energy consumed for preparing such a dish that is shared by a user through an image in its social network.

Furthermore, the author of [27] shows how the performance of food recognition can be improved by integrating multiple evidences (visual, location, and external knowledge, such as prior knowledge about the restaurant's menu). Besides improving the performance of the model, these context factors could also contribute to a better understanding of the user's food consumption related activity and provide more insights about the corresponding amount of consumed energy.

In [1] food-related tweets were determined using a unigram Naive Bayes classifier, after which the most popular food-related terms were enriched with nutritional information in terms of the amount of calories per serving. Longest n-gram matching was performed to detect the foods in the tweet text and aggregate their caloric content. Since the amount of calories is related to the amount of consumed energy, this enrichment and matching approach could also be applied in this research.

2.2.4 Leisure

In existing work, user activities are not necessarily separated into leisure activities. Nonetheless, most activities that are studied are implicitly affiliated with leisure. There have been multiple studies ([17, 35, 56]) that look into activity prediction based on travel behavior and patterns, as mentioned before. Check-in records contain semantic information (e.g. category of the POI) along with the spatial-temporal information, which makes a user's check-in behavior convenient input to infer activities. Opposed to the previous studies, [8] also takes the activity duration into account, which might be relevant when the amount of consumed energy will be linked to the activity. The authors follow up on the approach of [57], in which an SVM classification model is used to assign each tweet to an activity.

Type	**Papers**	**Achievements**	**Limitations**
Dwelling	[4, 10, 21]	Analysis at the residential level	No social media data incorporated in the models; no disaggregation of energy consumption activity; no analysis at the individual level
Mobility	[17, 32, 35, 40, 43, 53, 54, 56]	Analysis of check-in data, semantic enrichment of data	Mode of transport and travel duration not taken into account; not aimed at identifying the corresponding energy consumption
Food	[1, 3, 13, 14, 22, 27, 31, 36, 44]	Food recognition in text and images, ingredient and cooking method recognition in images, nutritional enrichment of data, enrichment of information about food-related venues	Not aimed at identifying the corresponding energy consumption
Leisure	[8, 17, 35, 56, 57]	Activity prediction based on travel behavior and patterns, enrichment of venue information (e.g., category)	Not aimed at identifying the corresponding energy consumption

Table 2.1: General overview of the state of the art in recognizing user behavior and activity from social media data