Big Trajectory Data Mining: A Survey of Methods, Applications, and Services

Di Wang [1,*], **Tomio Miwa** [2] and **Takayuki Morikawa** [3]

1   Department of Civil and Environmental Engineering, Nagoya University, Nagoya 464-8603, Japan
2   Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan; miwa@nagoya-u.jp
3   Institute of Innovation for Future Society, Nagoya University, Nagoya 464-8603, Japan; morikawa@nagoya-u.jp
*   Correspondence: wang.di@e.mbox.nagoya-u.ac.jp; Tel.: +81-070-4032-5426

Abstract: The increasingly wide usage of smart infrastructure and location-aware terminals has helped increase the availability of trajectory data with rich spatiotemporal information. The development of data mining and analysis methods has allowed researchers to use these trajectory datasets to identify urban reality (e.g., citizens' collective behavior) in order to solve urban problems in transportation, environment, public security, etc. However, existing studies in this field have been relatively isolated, and an integrated and comprehensive review is lacking the problems that have been tackled, methods that have been tested, and services that have been generated from existing research. In this paper, we first discuss the relationships among the prevailing trajectory mining methods and then, classify the applications of trajectory data into three major groups: social dynamics, traffic dynamics, and operational dynamics. Finally, we briefly discuss the services that can be developed from studies in this field. Practical implications are also delivered for participants in trajectory data mining. With a focus on relevance and association, our review is aimed at inspiring researchers to identify gaps among tested methods and guiding data analysts and planners to select the most suitable methods for specific problems.

Keywords: trajectory data; data mining; urban dynamics; human mobility; travel pattern

1. Introduction

The development of information and communications technology (ICT) and the proliferation of smart cities have generated tremendous volumes of data comprising specific geographic locations and corresponding time stamps [1]. The Internet of Things (IoT) comprising web-enabled smart devices using built-in sensors [2], radiofrequency identification (RFID), automated fare collection (AFC) systems, the Global Positioning System (GPS), Global System for Mobile Communications (GSM) beacons, and social networks provide abundant trajectory information for researchers to observe urban dynamics on a round-the-clock basis [3]. These trajectory datasets have demonstrated significant academic and practical value; they have been mined and analyzed by researchers to develop solutions for a wide range of emerging but important research questions in fields such as transportation, urban planning, abnormity and violation detection, and environmental protection [3–6].

Diverse methods have been utilized for analysis; they can be classified as statistical, visual, computational, or a combination of these [3], and they have been tested in many related planning, transportation or geographical studies. However, mining methods are continuously being developed as new and diverse application issues arise [3]. Hence, an integrated survey is urgently needed on the application issues regarding trajectory data and the corresponding mining methods applicable to these

issues. Such a survey will help other researchers identify problems, discover methodological gaps, and further develop new methodologies more rationally and efficiently.

The core objective of this paper is to review the methods and applications of trajectory data mining, as well as services that harness these methods to specific urban issues. The rest of the paper is structured as follows. In Section 2, a survey of similar literature reviews focusing on trajectory data mining is presented, which is used for developing research questions. Section 3 briefly elaborates the methodology applied for conducting this research. Section 4 offers an overview on the concept of trajectory data and its classifications. Section 5 discusses mining methods for trajectory data, while Section 6 presents potential applications of these methods as well as the problem-solution mapping relationship. Section 7 reviews the services that are supported by this mapping relationship. Section 8 presents a series of open discussions regarding the practical implications of trajectory data mining. Section 9 concludes this review and presents an outlook for future research.

2. Research Questions

Data mining, also popularly referred to as "knowledge discovery" [7,8], is an important process that extracts useful information from huge datasets. Since the emergence of data mining, its methods and applications have been widely investigated in the general data mining domain, as indicated in numerous literature reviews from early stages. For example, surveys of data mining methods for classic relational and transactional data can be found in Fayyad et al. [7] and Han et al. [9], which investigate the general concepts of data mining, and the fundamental techniques for preprocessing, clustering, classification, outlier identification, etc. Beyond this, some scholars (e.g., Mennis [10] and Miller [11]) reviewed theoretical and applied research in spatial and geographic data mining. Such research essentially derives from those in the general data mining domain, with methods specifically adapted to address spatial peculiarities, such as spatial correlation rules and spatial–non-spatial association. These mining tasks present certain rudiments for movement data research, but the reviews fail to consider the temporal dimension that is immanent in trajectory data.

Review papers that focus solely on trajectory data mining rarely follow a complete application-driven framework. Kong et al. [12] categorized trajectory data into explicit trajectory data and implicit trajectory data according to the degree of data structured, and introduced the "applications" of trajectory data from travel behavior, travel patterns, and other aspects. Their review contributed to the classification of multi-source heterogeneous trajectory data, but confused trajectory data mining methods with application issues, as well as practical services. Zheng [3] developed a profound survey on the techniques concerned with different stages of trajectory data mining, following a road map from the derivation of trajectories, to the preprocessing and management of trajectory data, and to the mining tasks such as trajectory pattern mining, trajectory classification, and abnormality detection. This review technically explored the approaches to adapt the existing methods in the general data mining domain to deal with emerging trajectory data, yet lacked the association between practical problems and methodological bases. For this reason, it contributes more to the community of data science than to a broad range of disciplines.

Andrienko et al. [13] developed a taxonomy describing the possible types of information that could be extracted from trajectory data and the respective types of analytical tasks in a systematic way. This taxonomy considers three fundamental sets, i.e., space, time, and objects, and distinguishes tasks according to the relations among the elements involved in each set. Andrienko et al. [13] also discriminated generic classes of analytical techniques, including visualization, data transformation, computational analysis methods, etc., and linked the types of tasks to the classes of techniques that could support fulfilling them. This work helps to match generic approaches with specific tasks in the field of trajectory data mining. However, it merely focuses on the methods developed from GIS-based visual analytics. By contrast, the contribution of our work breaks through this limitation, and moves into a broader field of computational analysis.

From a pure application perspective, Castro et al. [6] surveyed the existing research on mining taxi GPS traces, and grouped the surveyed work into three categories: social dynamics, which studies the collective behavior or movement patterns of a city's pollution; traffic dynamics, which studies the resulting flow of the population through the city's road network; operational dynamics, which learns from taxi drivers' knowledge of the city. This categorization method is more intuitionistic than Andrienko et al. [13] and has been widely referenced by researchers in the field of taxi trajectory mining, for example in [14–16]. However, Castro et al. [6] only considered the application of taxi trajectories. Besides, the matching relationship between fundamental mining methods and practical applications were not clearly elaborated. Our paper extends their work by considering beyond a specific kind of trajectory, meanwhile focusing more on the application issues in each category as well as their corresponding solutions.

Based on the review of similar research, the following gaps are identified. First, the exponential growth of ICT has enriched the connotation of trajectory data in recent decades, while some existing review papers are limited to specific kinds of trajectory data, e.g., taxi GPS traces in Castro et al. [6] or photo streams with spatiotemporal tags in Andrienko et al. [13]. In order to acquire a well-rounded understanding towards trajectory data mining, the first research question has been outlined as: What kinds of trajectory data can be utilized nowadays? Following this prerequisite, the major questions forming this application-driven research are delivered naturally: Which mining methods are applied or adapted to deal with trajectory data? What are the up-to-date application issues in trajectory data mining? What are the practical services that can be developed from studies in this field? Besides these obvious questions, there is a core thought that runs through our entire research and distinguishes our research from previous reviews: What is the matching relationship between the mining methods, the application issues that can be solved by these methods, and the practical services that can be derived from these applications?

3. Methodology

In order to answer the research questions proposed above, a systematic literature review (SLR) is performed "with respect to the planning for literature review, the design of search string, sources to be searched, publication inclusion and exclusion criteria, publication quality assessment and the data extraction process" [17]. Following the methodology indicated by Bach et al. [18] and Wahono [19], our research procedure is designed as shown in Figure 1. Since the major sections of our work follow a narrative workflow from trajectory mining methods to applications of these methods and to practical services derived from these applications, the literature for each section may be scattered and independent from each other. Therefore, unlike Bach et al. [18], which focuses on the applications of textual mining specifically in the financial sector, the literature reviewed in our research is evaluated by manual judgement, rather than the bibliometric software they indicated. The relevance between the literature in each section is also established through manual analysis.

Based on the analysis of existing literature reviews that are concerned with trajectory mining as presented above, the need for a systematic review and its research questions are identified (Step 1). Following the established review protocol in Moher et al. [20] (Step 2), the materials for this research are acquired by searching publications using the search string "trajectory data mining", for the period from 2004, in the Web of Science Core Collection database (Step 3), and then, are selected according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standard [20] for writing SLRs to form our review list (Step 4). Useful information is manually extracted from the selected literature (Step 5) and assigned to each major section that corresponds to research questions (Step 6). The relevance between the information assigned to each major section is established based on manual analysis (Step 7), so as to interpret the matching relationship between trajectory mining methods, application issues, and practical services (Step 8). During the reviewing process, additional literature is tracked and included by snowballing approaches [20], in order to fulfil our research needs. Thus, the reference list of this paper is longer than the original review list generated by PRISMA.
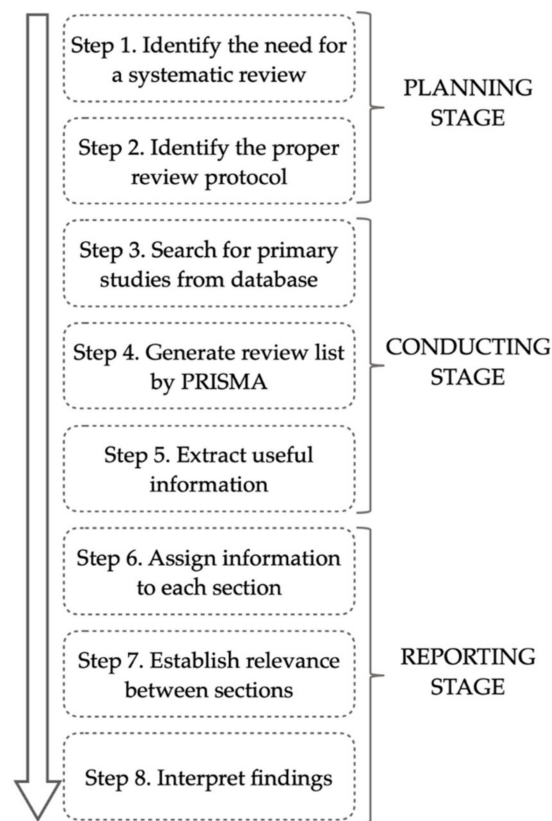
Figure 1. Research steps for the literature review.

4. Trajectory Data

What does the concept of a "trajectory" mean in the field of data mining? According to Zheng et al. [3,21], it is a trace generated by a moving object within a certain spatiotemporal context and is generally represented by a series of chronologically ordered points. In other words, trajectory data are essentially a sequence of spatial points ordered by timestamps and generally carry some descriptive information in addition to basic spatiotemporal messages. Therefore, a piece of trajectory data can be described as $TR = <P_1, P_2, \dots, P_n>$, where $P_n = (ID_n, X_n, Y_n, T_n, A_n)$ is the nth trajectory point; ID_n is the identifier; (X_n, Y_n) is the location of P_n in the specific coordinate system (i.e., natural geographic coordinate system or self-built coordinate system); T_n is the timestamp of the point (i.e., the moment when P_n is event-triggered [22] or regularly recorded); A_n is potentially a list of additional descriptive properties for P_n (e.g., instantaneous speed, running direction).

Different categories of trajectory data have been emerging and are being applied with the development of ICT. Renso et al. (2013) [23] distinguished GPS, GSM, and geosocial networks as significant carriers of trajectory data. Soon afterwards, Pelekis and Theodoris (2014) [24] enriched trajectory data sources with RFID and Wi-Fi. These data carriers represent currently existing types of trajectory data, which can be roughly categorized as explicit or implicit [12], as illustrated in Figure 2.
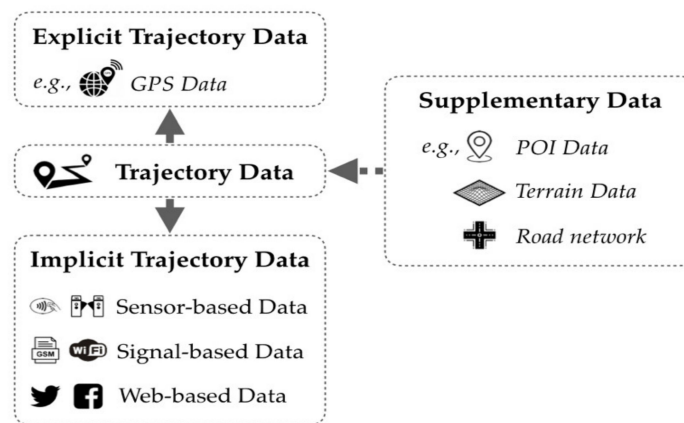
Figure 2. Trajectory data categories [12].

4.1. Explicit Trajectory Data

In this paper, explicit trajectory data are defined as a type of well-structured data which directly provide time and location information and have strong spatiotemporal continuity. They are regularly collected by terminal equipment at high (and usually fixed) frequencies, with no need to be triggered by any specific events. For example, trajectory data reported from GPS devices equipped in taxis are an uninterrupted series of spatiotemporal points recorded at fixed time intervals (e.g., 30 s intervals in most Chinese cities [25]). This kind of data is well-structured and contains relatively precise and direct spatiotemporal information along with other data fields, such as data fields indicating altitude, speed, direction, vehicle status, etc. [26].

In such cases, the above trajectory point model can be further evolved into a trajectory segment model: $TR = <SubTR_1, SubTR_2, \ldots, SubTR_n>$, where $SubTR = (ID_{sub}, trP_1, trP_2)$ is the sub-trajectory (i.e., segment) that forms the complete trajectory; ID_{sub} is the unique identity of the segment; trP_1 and trP_2 are the adjacent chronologically ordered trajectory points that delineate the segment. The trajectory segment can be obtained by finding the linear difference between trP_1 and trP_2, while the complete trajectory consists of n such segments connected end to end in chronological order.

4.2. Implicit Trajectory Data

Apart from explicit trajectory data, there exist such data carrying spatial and temporal information. Although they are not the trajectory data we usually think of, trajectory information can be extracted from them after basic data processing operations. This paper refers to such data that do not directly represent trajectory information as implicit trajectory data. In contrast to explicit trajectory data, implicit trajectory data have no definite continuity in time and space. In other words, they are triggered by an event rather than passively recorded. Such an event may refer to a bus or subway check-in, social network sign-in, sensor activation, signal tower reception, etc.; points with spatiotemporal information will not be recorded unless the corresponding event happens.

Unlike explicit trajectory data, which are usually recorded in a structured database format (e.g., Oracle DMP), the storage formats of implicit trajectory data are diverse and unstructured (e.g., text, image, audio, video) because of the variety in data sources and data collectors [21]. Although these forms of data present different properties, they have been applied to deal with similar or correlative issues among common mining methodologies [23,24].

4.2.1. Sensor-Based Trajectory Data

Sensor-based trajectory data (e.g., active fiber composite (AFC) and transit smartcard data [27]) are recorded when an object passes through a sensor. These sensors are mounted at a series of fixed positions and can only be activated at very close distances. Thus, sensor-based data have

high spatiotemporal accuracy but weak spatiotemporal continuity, which is limited by the number of sensors.

4.2.2. Signal-Based Trajectory Data

The collection of signal-based trajectory data requires multiple signal projectors (e.g., cell towers, Wi-Fi transmitters, Bluetooth connectors) to be distributed in advance. GSM-based data consist of chronologically ordered sequences of cell identifiers along which the moving object passes. Wi-Fi and Bluetooth-based data comprise temporal sequences of identifiers of access points that have communicated with the moving object [28]. This type of data is more complex than the previous ones and generally contains the device ID, connection/disconnection timestamp, signal strength, etc. Preprocessing is needed to extract trajectory information [29].

4.2.3. Web-Based Trajectory Data

Web-based trajectory data are contained in a geolocalized social network. Recent years have witnessed the rise of social sites/apps (e.g., Twitter, Facebook, Weibo) equipped with geotag functions. In addition to spatiotemporal information, such social network services also carry semantic information regarding specific events, human activities, emergencies, etc. [30]. To some extent, web-based trajectory data are more informative than any other categories because useful knowledge may be extracted from the additional semantic information they are carrying, but they are also more implicit due to the noise they contain and semantic messages that are difficult to process.

4.3. Supplementary Data

In most studies, trajectory data do not function alone; they are projected into built environments for better analysis. Such built environment information (e.g., points of interest (POIs), road network, terrain distribution, urban structure), which is usually integrated with trajectory data in data mining applications, is regarded as supplementary data. This kind of data generally exists on or can be extracted [31] from comprehensive digital map platforms.

POI is a major data category that represents the reality of built environments. It literally refers to a specific point that someone finds useful or interesting [32]; POI data contain information for almost all key nodes within an urban area (e.g., locations of buildings concerned with retail, catering, education; administration and locations of facilities related to transportation, communication, and security) [32,33]. The supplementary information from POI data can be used to develop a more reasonable and rational explanation for patterns and behavior detected in trajectories.

Other supplementary data are mainly related to the geographical context of trajectories [34] (e.g., road network, elevation system). In fact, it is quite a natural perception and practice to connect trajectory data with such geographic frameworks because geography is one of the two most remarkable attributes of trajectories (the other is time) [35]. Existing online map platforms (e.g., open-source platforms like OpenStreetMap (OSM) and commercial platforms like Google Map and Baidu Map) can serve as sources for these supplementary datasets [36–38].

5. Trajectory Data Mining Methods

What does trajectory data mining mean? Similar to the common understanding of general data mining, trajectory data mining means to discover interesting knowledge (e.g., movement patterns, travel behavior, traffic abnormality) from trajectory datasets. Generally, trajectory data mining has two major tasks: description and prediction [3,21]. Description is to interpret human-readable information from massive volumes of trajectories, while prediction is to discover uncharted or prospective values by analyzing existing variables in datasets. These two basic tasks are performed for all applications and services that are related to trajectory data.

With regard to methods for trajectory data mining, this paper focuses more on the methodology or principles rather than listing all specific technical procedures that have been adopted in previous

research. We concentrated only on the prevalent methods and tried to ascertain the connections between them. We divided these methods into two categories: first-tier and second-tier methods. The former sorts trajectories directly based on their attributes, while the latter usually contains sequences of first-tier methods, sometimes together with non-mining methods (e.g., statistical or topological), to study the spatiotemporal permutation of trajectories.

5.1. First-Tier Trajectory Data Mining Methods

As noted above, first-tier methods classify trajectories directly from cleansed datasets based on their inherent properties. These methods are basic, yet most important in the field of trajectory data mining. The application of first-tier methods is usually followed by a descriptive interpretation of the results, and on many occasions, functions as the preparation for subsequent extended analysis, such as with second-tier methods. Data cleansing is a preprocessing task before first-tier data mining, but we will not discuss it in detail here. Detailed information regarding data cleansing can be found in [39–41]. In this article, all of the methods discussed are assumed to be based on cleansed datasets.

5.1.1. Clusterings

Clustering is a first-tier trajectory data mining method. It is an unsupervised learning process that reveals similarities within a trajectory dataset by dividing trajectories into categories (i.e., clusters) according to their properties to indicate homogeneity and heterogeneity [42]. In other words, the movement characteristics of trajectories should be similar within a cluster, while different between clusters. A general clustering approach is to represent each trajectory with a feature vector, and then, measure the similarity between trajectories by calculating the distance between their feature vectors [3]. However, it is not easy to generate feature vectors with a uniform length for different trajectories, since trajectories may vary significantly in terms of length, shape, sampling frequency, point quantity, point order, and many other properties. Besides, it is also difficult to encode the sequential properties of points in a trajectory into its feature vector.

Considering the challenges mentioned above, a series of technical explorations have been done. On the one hand, there are widely accepted clustering algorithms for trajectories which are essentially extensions of classical clustering algorithms [43–45] with specific customization on the similarity (or distance) functions to determine cluster membership. A detailed discussion on how these functions are applied can be found in Rokach [45]. Generally, depending on the goal of analysis, similarity (or distance) functions such as similar destination, similar origin, similar direction, or others are utilized to determine which trajectories belong to the same cluster.

On the other hand, there have been efforts to develop trajectory-specific clustering approaches. Many of them are accommodating statistical or probabilistic models for measuring the characteristics of trajectories. For instance, Gaffney and Smyth [46] and Cadez et al. [47] proposed mixed regression model-based approaches to aggregate trajectories likely to be generated by a common representative trajectory with Gaussian noise. The Expectation Maximization (EM) algorithm they proposed clusters trajectories with respect to the overall distance between two entire trajectories. Similarly, Alon et al. [48] abstracted trajectories as sequences of position transitions and utilized a Hidden Markov model (HMM) that best fit the trajectories to select cluster members.

The approaches proposed by Gaffney and Smyth [46], Cadez et al. [47], and Alon et al. [48] are applicable to entire trajectories. In other words, they group similar trajectories as a whole. However, in reality, moving objects rarely move together for an entire path. Besides, discovering common sub-trajectories is also useful in many applications, especially when there are regions of special interest for analysis. To this end, Lee et al. [49] proposed a partition-and-group framework, which partitions an entire trajectory into a set of line segments, and groups similar line segments into a cluster using the Trajectory Hausdorff Distance [50]. A representative trajectory describing the overall movement of the trajectory partitions that belong to a cluster is identified by sweeping a vertical line across the line segments in the direction of the major axis of a cluster.

The clustering approaches mentioned previously are developed for static datasets. They are not suitable for incremental clustering, when trajectory data are received incrementally, e.g., continuous new points reported by a GPS system. Li et al. [51] proposed an incremental clustering framework for trajectories to deal with this situation. The framework has two components: online micro-clustering maintenance and offline macro-clustering creation. For the online part, micro-clusters are incrementally updated when new data are added; for the offline part, when the user requests current clustering results, macro-clustering is performed on the sets of micro-clusters rather than all trajectories over the entire time span. This approach is able to save the computational cost and the storage of received trajectories when processing trajectory data streams.

All the methods discussed in this section are oriented towards trajectories in a free spatial context, i.e., with no road network constraints. Kharrat et al. [52] proposed the NETSCAN algorithm that applies specifically to trajectories that lie on a predefined network. NETSCAN is essentially an extension of classic DBSCAN that first computes dense paths in the network and then, clusters the sub-trajectories similar to the dense paths. Apart from this, there are few studies focused on trajectory clustering in a road network setting, because this task can be easily solved by the combination of map matching and regular trajectory clustering algorithms. Map matching is the process to project trajectories onto a corresponding road network, and meanwhile, attaching road network information to the trajectories. Map matching approaches can be found, for example, in Miwa et al. [53] and Quddus et al. [54].

5.1.2. Classification

Classification differs from clustering because it is a supervised or partially supervised learning process [55]. The classification classes need to be predefined, and a training set of objects needs to be prelabeled with the class that they belong to. For example, a typical case of trajectory classification may be to label each trajectory from a large set with its means of transportation based on a small set of trajectories that have already been labeled. This small set is the training set. Thus, the labeling process (i.e., assigning objects to predefined classes based on the means of transportation) is classification.

A typical trajectory classification algorithm contains two steps. First, it needs to extract a set of discriminative features that can be used to train an existing standard classification model (e.g., logistic regression [56], support vector machine (SVM) [57], decision trees [58], nearest neighbors [59]). This step is to find the trajectory properties that are best suited to defining the various classes of trajectories. Trajectories have many potential useful properties (e.g., transportation means, average speed, time duration, trajectory length), but their discriminative power depends on the type of classes expected. For example, if the taxi fare is a class type, the trajectory length has a stronger discriminative power than the time duration because taxis charge according to mileage rather than time. The second step is to select a proper standard classification model, and then, apply it to the extracted discriminative features.

Several comparative studies have been performed on standard classification models and their corresponding classical classification algorithms [60]. Most of these classical methods can be directly applied to trajectory classification. For example, Bolbol et al. [61] utilized SVMs for transportation mode classification. They first evaluated the discriminative power of several features for six transportation modes (i.e., bus, subway, train, private car, bicycle, and walking) through statistical methods, and identified speed and acceleration as the most discriminative. Second, they applied a standard SVM algorithm to these features to classify trajectory segments. Zheng et al. [62] did similar work, except that that they applied a decision tree-based inference model to the discriminative features for transportation mode classification.

In many situations, trajectories are classified following some preprocessing (e.g., segmentation, clustering, statistical analysis) that prepares the features needed for classification [62–64]. For example, Zheng et al. [62] proposed a change point-based segmentation method to partition each complete trajectory into separate segments of different transportation modes; they identified a set of features not affected by differing traffic conditions that could be fed to the inference model. Lee et al. [63]

performed trajectory clustering to extract regional and sub-trajectory features for an SVM-based classification model.

5.2. Second-Tier Trajectory Data Mining Methods

The first-tier methods presented above are generally used to categorize trajectories. In many cases, they are then followed by second-tier trajectory mining methods, which are used to analyze the spatiotemporal characteristics of the individual trajectories within or between categories that were identified by the first-tier methods. In other words, this is a subsequent processing of the results from the first-tier processing. Many types of methods are available for this stage but the three most versatile are pattern mining, outlier identification, and prediction.

5.2.1. Pattern Mining

Pattern mining concentrates on discovering interesting, significant, or unexpected patterns that exist in databases. It is one of the most fundamental tasks of data mining [65]. Various patterns can be mined (e.g., frequent items, sequential rules, periodic patterns, subgraphs, associations) corresponding to various algorithms for pattern mining (e.g., frequent pattern (FP) growth, a priori, ECLAT) [66–68]. These algorithms can be categorized into three types of trajectory pattern mining: periodic, frequent, and collective.

A periodic pattern refers to trajectories periodically executed by a moving object [69–71]. For example, it may reflect the regular movement patterns from office staff, which are rather similar each working day. In contrast, a frequent pattern is not focused on such temporally repetitive phenomena of individuals but refers to a specific sequence of places that have been visited by a certain number of moving objects with no specific temporal constraints [66]. A typical example of a frequent pattern is a park itinerary, which is followed by most tourists. A collective pattern is a combination of these two and is performed by groups sharing similar mobility interests both temporally and spatially. In other words, these moving objects travel together [70]. Periodic pattern mining utilizes location sequences as mining criteria. Early-stage approaches [71,72] require the time period to be a specific input in the mining algorithm. They cluster the sequences of locations in each preset time branch and then, iteratively connect the detected frequent sequences to obtain the integral pattern. However, such work to preset the time period involves many uncertainties. For example, the division of time intervals will definitely affect the clustering output, but these effects are difficult to measure. Meanwhile, different time periods may occur as the discovery progresses, but these algorithms are not equipped with dynamic adjustment capabilities. Li et al. [73] developed the Periodica algorithm to overcome these problems by bypassing the time period presetting. Their algorithm selects regions where more trajectory points exist as the reference spots and then, automatically detects the periods in each spot through a combination of Fourier transformation and autocorrelation. These periods are used to discover periodic patterns from location sequences between reference spots. Hierarchy-based clustering with a probability-based distance measurement model is performed on these location sequences. Compared to previous algorithms, Periodica better matches realworld scenarios because the period-setting criteria are unpredictable in principle until the real movement sequences are considered [74].

Frequent pattern mining focuses on the collective routes or paths that have been frequently traveled by multiple moving objects [66]. Thus, such patterns can be discovered simply by using the spatial features of trajectories [75] (i.e., only the sequences of spatial locations need to be considered). Typical examples include frequent spatiotemporal sequential patterns (FSSP) mining [76] and generalized sequential patterns (GSP) mining [77]. However, some have considered frequent patterns not only as spatial elements but also as temporal elements along spatial trajectories. For example, Giannotti et al. [78,79] defined the T-pattern as an assemblage of individual trajectories sharing the common attribute of visiting the same sequence of locations with similar transition times. There are roughly two scenarios for frequent patterns, as illustrated in Figure 3. For Figure 3a, frequent pattern mining can be based on the clustering methods discussed in Section 5.1.1 or simply by applying

statistical analysis [76,77]. For Figure 3b, however, frequent patterns cannot be discovered within one step. Thus, a two-step approach can be used [79,80], which consists of detecting significant regions outside the trajectories and then, performing sequence mining in these regions as a temporally annotated sequence.
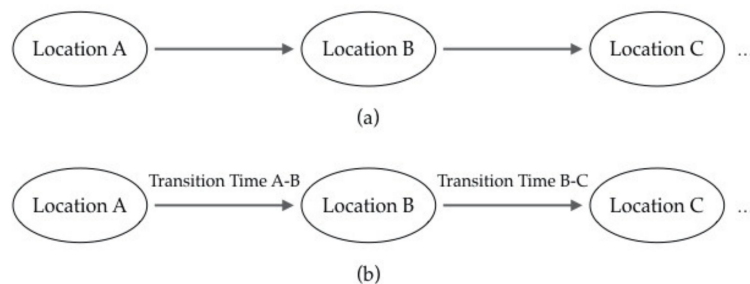

Figure 3. Frequent patterns based on (**a**) spatial sequences and (**b**) spatiotemporal sequences.

Collective pattern mining (i.e., group pattern mining) essentially finds movement patterns that have been performed by groups with similar mobility interests [70]; such interests require not only spatial proximity but also time coordination. Considering the spatiotemporal closeness, internal structures, and external performances, group patterns can be roughly categorized into three types, as illustrated in Figure 4: flock, convoy, and swarm [3]. A flock [81,82] refers to a group of at least o objects that move together for at least t successive timestamps; the positions where these moving objects remain on each time slice can be observed in a disk with a radius r. Thus, such patterns can be described with three parameters: o, t, and r. A convoy [83,84] is similar to a flock except for its relaxed requirements for the disk shape. A convoy pattern allows its moving objects to form any disk shape on each time slice as long as the positions can be clustered, usually by density-based clustering with a maximum neighborhood distance d and minimum object number o [83]. A swarm further relaxes the requirements for a convoy. The timestamps do not need to be successive in this situation; in other words, there does not need to be at least o positions that can be clustered on every time slice [85]. A swarm shares similar parameters to the other two, including o, t, and d. However, it also includes the swarm parameter k, which indicates the minimum number of time slices on which the collective patterns can be detected. For example, Figure 4c shows a swarm situation where $o = 4$, $t = 3$, and $k = 2$. It is neither a flock nor a convoy because object O_4 breaks away from the group at timestamp t_2. It can be considered a swarm pattern because all four objects can be clustered into one group at timestamps t_1 and t_3. Density-based clustering is the most common method for collective pattern mining, whether it is for flocks [86], convoys [83,87], or swarms [88,89]. In most cases, clustering is the first-tier step, while parameters are checked next to determine which category a collective pattern belongs to.
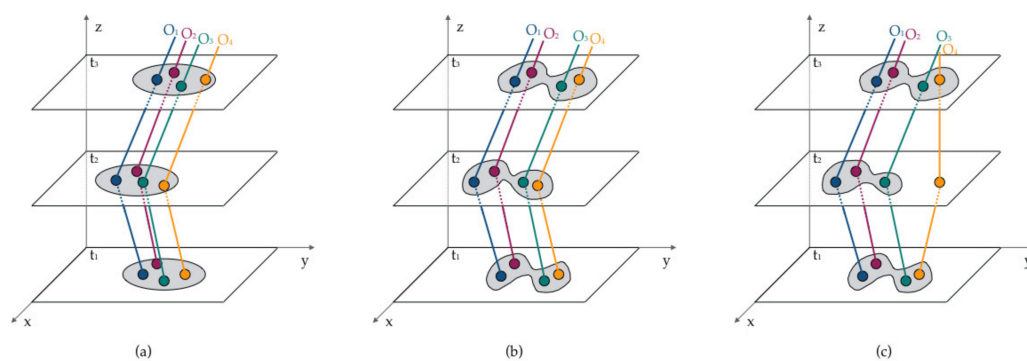

Figure 4. Collective pattern categories: (**a**) flock, (**b**) convoy, and (**c**) swarm [3]. Each image contains three timestamps (i.e., t_1, t_2, t_3) and four moving objects (i.e., O_1, O_2, O_3, O_4).

5.2.2. Outlier Identification

In data mining, outlier identification (i.e., outlier or anomaly detection) involves the detection of rare items, events, or observations that arouse suspicion by differing significantly from the majority of the dataset. Outliers can also be referred to as anomalies, novelties, noise, deviations, and exceptions [90]. For trajectory data, outlier detection involves discriminating trajectories that are barely consistent with the common characteristics of the majority of trajectories [91]. To some extent, it is complementary to the above trajectory mining methods; these methods focus on the homogeneity of the data, while outlier identification is more concerned with the heterogeneity.

Thanks to this complementary relationship, a major methodology for outlier detection is to concentrate on the byproducts of trajectory clustering. Theoretically, trajectories that do not belong to any cluster should be outliers. However, a significant disadvantage of this indirect approach is that it cannot guarantee sufficient differentiation between byproducts. In other words, these byproducts may also show some similarities, although these similarities may not satisfy the clustering criteria of the previous step. Thus, further detection should be performed to distinguish real outliers among these byproducts. This kind of work may fall into a trap of continuous looping.

There have been some attempts at outlier identification from a more direct perspective. One approach is to mine all trajectories for outliers [92,93]. In such cases, each complete trajectory is abstracted to a set of key features (e.g., the spatial coordinates of the start and end points, the values (minimum, maximum, mean) of the directional vectors and velocities). Then, distance-based algorithms, which are usually equipped with a distance function defined as the weighted sum of the differences of the abstracted features [93], are applied to outlier detection. In this situation, the basic unit for mining is the complete trajectory. However, such methodology may not be able to find outlying trajectory sections. For example, Figure 5 clearly shows that section A–B of trajectory TR_3 is different from its neighboring trajectories (TR_1, TR_2, TR_4), but it may not be distinguished as unusual because the overall behavior of TR_3 is similar to that of its neighbors. From a mathematical perspective, the significant differences in sub-trajectories may be averaged out over the complete trajectory.
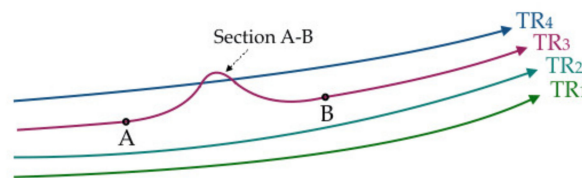

Figure 5. Example of an outlying sub-trajectory: section A–B of trajectory TR_3 [91].

Another approach is to focus on the decomposed trajectories [91,94,95]. These methods generally partition each complete trajectory into a set of sub-trajectories and then, detect the outlying sub-trajectories by applying a distance function or clustering approach. Eventually, the complete trajectories that contain outlying sub-trajectories are discriminated as outliers. Lee et al. [91] defined this as a partition-and-detect framework, as demonstrated in Figure 6, and proposed the TRAOD algorithm, which utilizes a hybrid of distance-based and density-based approaches for the second step of outlier detection.

These methods are generally based on clustering and its extensions; in other words, they are unsupervised or semi-supervised learning processes. However, supervised learning approaches are also available for outlier identification; these are typically based on classification. For example, Yuan et al. [95] extracted a set of pre-identified features (i.e., direction, speed, angle, and location) from trajectories to which they then applied distance measures to discriminate anomalies. Li et al. [96] utilized trajectory features to train a two-label classifier model: one label classified normal trajectories while the other classified abnormal ones.
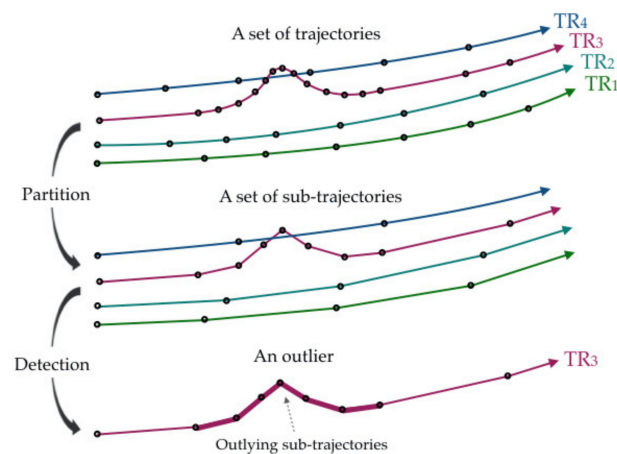
A set of trajectories

TR₄
TR₃
TR₂
TR₁

Partition

A set of sub-trajectories

Detection

An outlier

TR₃

Outlying sub-trajectories

Figure 6. Partition-and-detect framework [91].

5.2.3. Prediction

In data mining, prediction involves assuming that certain turns of events will occur based on the description of other related data. The prediction itself is calculated from the available data and modeled in accordance with the existing dynamics [97]. There are two approaches to predictions using trajectory data: predicting the future location of a moving object, and predicting its entire route within a road network context. There are three categories of location prediction: (1) based on the dynamics of the moving objects of concern [98–101], (2) based on the dynamics of other objects (e.g., a set of located users with a social-spatial performance that exceeds IP-based geolocation) [102], and (3) based on both the objects of concern and other objects [103–105].

Two major approaches have been applied to these three categories of location predictions: the Markov chain model [103–105] and the trajectory pattern-based method which relies on frequent pattern mining (see Section 5.2.1) and the association rules among the patterns and corresponding influencing factors. For example, Ying et al. [106] first extracted trajectory patterns to identify the mobility behavior motivated by geographic, temporal, and semantic factors; they then matched the current movements of the objects of concern to the extracted patterns. Monreale et al. [107] introduced a decision tree called the T-pattern tree after extracting trajectory patterns as predictive rules. The tree was built and evaluated with a formal training and test process and eventually, shows a certain level of accuracy for next-location prediction.

In contrast, route prediction speculates a sequence of paths starting from a certain location of a certain moving object. This is generally conducted under strict built environment constraints (e.g., a road network) [108–110]. Currently, there are three approaches to route prediction: trip observation-based, Markov model-based, and turning behavior-based. Trip observation-based prediction is based on the fact that a large portion of a typical driver's trips are repeated. Thus, this type of prediction utilizes observed locations of the object's past trips to develop algorithms for end-to-end route prediction [110]. These algorithms essentially match the first part of an object's current trip with its set of previously observed trips to determine the most likely following part. Markov model-based prediction is applicable to both location prediction and short-term route prediction. For route prediction, a simple Markov model is first trained from the object's long-term trip history and then, applied to making a probabilistic prediction for the next road segment considering the path that the object just followed [111]. Turning behavior-based prediction is focused on the object's turning choices at intersections. When strung together, these choices form a route in the road network. In other words, if the object's aggregate turning behavior (including the choice to go straight ahead) can be predicted, its future route can be identified [112,113]. The Markov model is a typical method for turning behavior prediction, but other ways include pure statistical methods. For example, Krumm [113] proposed an

algorithm and variations to infer the proportion of drivers that take each turning option at intersections based on the assumption that drivers are more likely to choose a turning option that offers more destination options.

5.3. Relationships between Trajectory Data Mining Methods

As discussed previously, first-tier methods are the foundation of second-tier methods. For most tasks, we need the former to categorize trajectories according to their homogeneity or heterogeneity, while the latter is used for deeper or more synthetic analysis on the already clustered or classified trajectories. Table 1 describes such relationship for specific tasks. Each task is concerned with a specific second-tier mining method, which naturally requires a corresponding first-tier method. For example, frequent pattern mining (second-tier) applies clustering (first-tier) to find places of significance [79,80]. There exists some overlapping between second-tier methods. For example, a prediction method (second-tier) can use pattern mining (also second-tier) to obtain a concise representation of the object's moving behavior, which is essential for future location prediction [106,107].

Table 1. Relationships between trajectory data mining methods.

Categories		First-Tier Mining Methods	
Categories	**Methods**	**Clustering**	**Classification**
Second-tier Mining Methods	Pattern Mining	Grouping spatially close trajectories [86,87]; Grouping temporally related trajectories for periodic pattern mining [69,71,73]; Extracting places of significance for frequent pattern mining [79,80]; Detecting similar mobility interests for collective pattern mining [83,86–89]; Aggregating close locations for sequence analysis [114];	No classification-related tasks have been identified for pattern mining.
	Outlier Identification	Grouping trajectories or sub-trajectories with homogeneity [91,94];	Sorting out trajectories based on pre-identified features [95,96];
	Prediction	Grouping multiple users with similar mobility intentions [102,115]; Grouping similar trips of one specific object [116]; Mining trajectory patterns for location prediction [100,101,106,107];	Matching one object's current movement with its movement patterns for location prediction [107,116]; Matching one object's ongoing trajectory with its previous trajectories for route prediction [110];

Note: Not all cases are listed. Information in this table is summarized based on our literature survey.

6. Application Issues with Trajectory Data Mining

In the previous sections, we discussed three categories of emerging trajectory-related data and two classes of data mining methods for extracting information from trajectories. In this section, we look at the application issues that can be addressed with mining trajectory data. Current application issues can be sorted into three categories: social dynamics, traffic dynamics, and operational dynamics [6]. These issues can be matched with the trajectory data mining methods that were presented in Section 5. Note that there is no firm one-to-one mapping relationship between an application issue and mining method. A single application issue may require several mining methods, or different issues can be tackled with the same method. The matching relationships between issues and methods depend on the specific tasks involved with an issue. These relationships and their corresponding references are presented in Table 2, which may help guide other researchers to select the most suitable methods for specific application issues.

Table 2. Relationships between trajectory-related application issues and trajectory mining methods.

Application Categories	Application Issues	Description of Issues	Major Tasks Involved	Mining Methods Involved
Social Dynamics	Discovery of Social Relationships	Discovery of social ties between individuals and communities	Grouping individuals' stay locations	Clustering [117]
			Extracting chronologically ordered sequences of stay locations	Frequent pattern mining [117]
		Discovery of interaction between animals	Detecting groups of moving animals, describing groups' features	Collective pattern mining [88]
	Detection of Social Events	Detection of event occurrence	Grouping based on spatiotemporal properties	Clustering [118]
		Profiling of discovered events	Extracting features and categorizing events	Classification [119]
	Characterization of Connection between Places	Detection of hotspots	Grouping according to spatiotemporal properties	Clustering [120,121]
		Description of land uses and regional functions	Discovering regions with similar functions	Clustering [120,121]
			Extracting features and categorizing regions	Classification [120,121]
		Description of connection between places	Extracting origin/destination links	Clustering [122,123]
			Discriminating abnormal links	Outlier identification [124]
Traffic Dynamics	Profiling of Moving Objects	Inferring mobility activities and modes	Extracting features and categorizing activities	Classification [61,125,126]
		Profiling movement patterns	Extracting sequences of visited places	Frequent pattern mining [127,128]; Clustering [129]
	Trajectory-based Prediction	Predicting an object's future location/route	Establishing probabilistic model for prediction	Statistical methods (e.g., Markov Chain) [104,105]
			Comparing current trajectory with extracted historical trajectories	Frequent pattern mining [106,107]
		Predicting traffic jams	Inferring traffic density and comparing it with road capacity	Frequent pattern mining [130]; Statistical methods [131]
Operational Dynamics	Interest Recommendation	Friend–place recommendation	Extracting shared movement patterns and ranking similarities	Frequent pattern mining [132–134]
	Trip Recommendation	Suggesting order of visiting locations	Predicting routes based on user preferences	Frequent pattern mining [135,136]

Note: Information in this table is summarized based on our literature survey. Application categories are recommended by Castro et al. [6].

6.1. Social Dynamics Issues

To some extent, social dynamics can also be considered as community dynamics. A community is a group of entities that share some common interests. In the context of trajectory data mining, such interests are represented as common mobility behavior based on the observed trajectories [6]. In most cases, the application issues for social dynamics are not constrained by the built environment (e.g., road network). Rather, they are concerned with collective movement trends at the community, city, or even region level. This is in contrast to detailed network-based paths, which are motivated by various internal demands (e.g., work, shop, school) and affected by various external factors (e.g., weather, traffic, policy) [6,15]. Previous research on social dynamics issues has utilized trajectory data to tackle questions such as where people go during the day [15,137], the locations of hotspots (i.e., where traveling origins and destinations accumulate) around the city [138,139], the functions of these spots in an urban context [120,121], and the strength of connections between different parts of the city [140]. These studies have used diverse categories of trajectory data to reveal a well-rounded understanding of the urban reality. Zheng et al. [21] defined such data as digital footprints and the framework for mining trajectory data as urban computing, which they have detailed in several papers [4,21,22].

6.1.1. Discovery of Social Relationships

Theoretically, trajectory data can be used to extract existing interactions between moving objects and discover more about the properties of these interactions. Such interactions and their properties are defined here as social relationships; they include those between individuals, communities, or even animals (e.g., predator-prey interactions [141]).

To what extent can interpersonal relationships be inferred from spatiotemporal trajectories? The wealth of geographic information in social media has provided an opportunity for researchers to explore this question in detail. For example, Crandall et al. [142] proposed two sub-questions: (1) Provided that, on multiple occasions, two individuals are in roughly the same geographic location at nearly the same time, how likely are they to know each other? (2) How does this likelihood depend on the proximity of the co-occurrences in time and space? They then established a framework for quantifying answers from a social media website and found that a high likelihood of interpersonal ties can be triggered, even from a small chance of co-occurrences. They also built a probabilistic model to show how such large probabilities of social ties arise from co-occurrences. Meanwhile, some researchers have attempted to answer this question by examining the relationship between individuals' social ties and their visits to the same places [117,143]. For example, Wang et al. [143] tracked the trajectories and communication records of millions of cellphone users and discovered that the similarity between two individuals' movements strongly correlates with their proximity in a social network.

In terms of community-level social relationships, many studies have followed classical clustering and pattern mining methodologies, as discussed in Section 5 [6,117,144]. For example, Gaito et al. [117] proposed the concept of a geocommunity, which combines the geolocations of individuals and social communities with common mobility interests. They extracted geolocations by clustering the stay-locations of individuals and then, utilized density-based clustering to discover their communities. They then adopted sequential pattern analysis methods to detect the social relationships between communities.

6.1.2. Detection of Social Events

From the perspective of trajectory mining, a social event refers to a gathering of long-term but temporary stay-locations. Thus, detecting social events involves recognizing the existence of such gatherings. When combined with the properties of the gatherings (e.g., semantic information), the type of social event can also hopefully be identified [145]. Typically, there are three methods for recognizing

social events from trajectories: statistics-based, classification-based, and clustering-based. In many cases, multiple methods are applied.

As a typical example of statistics-based detection, Giannotti et al. [130] discovered social events by identifying a high concentration of stationary objects that were previously moving within a specific spatiotemporal constraint. This method can be followed by a classification procedure to identify the event type. For example, Calabrese et al. [119] classified the feature vectors of attendees' origins detected from cellphone data to estimate the type of event.

As a more realistic scenario, Zheng et al. [118] proposed the snapshot, which indicates a social event that satisfies the following conditions: (1) the groups of individuals are dense, (2) the shape and location of the groups generally do not change, and (3) the group members can enter and leave at any time as long as there are a certain number of members in this group for a certain period of time. They then proposed a density-based clustering method to detect snapshot clusters, from which gathering (i.e., social event) patterns can be extracted.

6.1.3. Characterization of Connections between Places

The characterization of places and profiling of connections between places are closely related application issues for trajectory data mining. Both utilize origin–destination (OD) information as the key to uncovering urban realities at a relatively macro scale, and they share similar mining methods (e.g., hierarchy-based clustering, density-based clustering, classification).

Detecting hotspots within a city is a first-tier task for characterizing a place. A hotspot refers to a region where urban activities regularly accumulate [138,139] and is usually mined through clustering methods. For example, Chang et al. [138] considered areas with a high intensity of taxi requests to be a typical kind of urban hotspot and clustered passenger pick-up points (which can be distinguished in taxi GPS datasets based on certain field values) to discover such hotspots. They even built a hotness index based on the properties of these clusters. Liu et al. [146] also used a clustering-based approach to represent urban hotspots by certain crowdedness dynamics considering the real clustering properties of objects. They proposed a non-density-based approach called mobility-based clustering, where each sample object is utilized as a sensor to perceive the crowdedness around it by using its instant mobility properties (e.g., a taxi's instant speed).

Another type of application is identifying the land use types and regional functions within a city. Such work is generally conducted in two steps: clustering to extract regions and classification to assign function-related properties to the extracted regions [120,121,147]. For example, Pan et al. [121] tried using taxi GPS trajectories to classify urban land use. They applied a modified density-based clustering method called iterative DBSCAN to extract regions. They then classified regions into different social functions based on the taxis' pick-up and drop-off dynamics.

Similarly, characterizing connections between places also relies on the OD mechanism within the trajectories. For example, Liu et al. [122] utilized taxis' pick-up and drop-off records (PDRs) and the check-in and check-out records generated by smartcards to study the connected regions and corresponding connection strength. They applied clustering techniques to analyze the trip relationship between different zones with an OD matrix. As a complement to such research, abnormal connections can be detected with outlier identification methods as discussed in Section 5.2.2, which discriminate abnormal connections from normal ones [124]. Another issue is identifying the properties of the detected anomalies, which, in turn, comes back to the above characterization of connections [148].

6.2. Traffic Dynamics Issues

Traffic dynamics specifically refers to how people carry out their mobility intentions depending on the road network or other built environments and governed by their underlying travel demands [6]. In this paper, we broaden the understanding of traffic dynamics issues by referring to tasks that are directly related to the movement or moving object, as well as predictions directly based on existing

movements. Unlike research on social dynamics, which principally utilizes OD-related information, research on traffic dynamics usually makes full use of trajectories.

6.2.1. Profiling of Moving Objects

Starting from the trajectory, the most direct type of research would be to profile the moving object that generates the trajectory. Such research includes but is not limited to deducing the activity types of humans [125], profiling the mobility routine of humans [127,128], inferring transportation modes [61,62], understanding the moving behavior of animals [126,149], and describing the movement patterns of animals [88,150]. These issues seem scattered but are essentially concerned with the inherent properties of the trajectories and have the common ambition of understanding the behavior of the moving object.

Research in this category has mostly been focused on inferring the activity types of travelers and identifying their traffic mode. These two issues can be addressed through classification methods based on some preset features, as done by Zheng et al. [62]. Clustering methods are used in frequent pattern mining to deal with profiling issues for human mobility routines and animal movement patterns. Such approaches essentially extract the sequences of places that moving objects have frequently visited, as discussed in Section 5.2.1.

6.2.2. Trajectory-Based Prediction

In Section 5.2.3, we discussed prediction in detail as a major trajectory mining method. This involves two major issues: predicting the next position (or destination) of the moving object and inferring the route that the moving object will follow. Here, we introduce another major issue: forecasting the occurrence of traffic-related incidents such as traffic congestion. Specific solutions for the first two issues were presented in Section 5.2.3. Thus, here we only discuss the methods for predicting traffic congestion.

Areas with traffic congestion essentially have high traffic density. Thus, the problem of predicting congestion can be transformed into the problem of inferring traffic density. Giannotti et al. [130] established a tree structure formed by T-patterns to predict the locations of areas where large amounts of trajectories accumulate. Each T-pattern represents a sequence of visited positions and the corresponding transition time, and each tree node carries a support value indicating the number of T-patterns that connect the tree root to the current node. Another approach is extending the classical Markov-based route prediction method. The predicted routes will eventually constitute a certain level of traffic density. By comparing the predicted traffic density with the capacity of the corresponding road segments, we can theoretically find areas at risk of congestion. Castro et al. [131] used this approach to build a prediction model based on the probabilities of switching between road segments and determined the capacity by referencing the historical traffic density.

6.3. Operational Dynamics Issues

In this paper, operational dynamics refers to the information that can be extracted from trajectory data, which can potentially be applied to social, economic, commercial, or other operations. Compared with the above two categories, issues with operational dynamics require deeper mining to reveal the properties hidden within the trajectories.

6.3.1. Interest Recommendation

Interest recommendation is based on the hypothesis that people who share similar mobility profiles are likely to share similar interests and preferences [132]. There also exists a two-way positive interaction mechanism between potential friendship and shared movement patterns. In other words, if two people consistently display similar mobility profiles, they can be recommended to become friends. In turn, if a person's friends frequently visit certain places or follow certain routes, these places and routes can be recommended to the person. This has become one of the most fundamental hypotheses in current social network operations.

A common method for interest recommendation is to mine the mobility history for frequent patterns. For example, Li et al. [132] established a framework for a friend–place recommender system with three internal modules: mobility history representation, user similarity evaluation, and friend-place recommendation. In the first module, hierarchy-based clustering is performed on the stay-locations of each user to obtain their mobility history, which is then visualized in a hierarchical graph. In the second module, similar sequences of shared graph nodes are retrieved from all users' graphs, which are then used to generate similarity scores for each user pair. In the third module, users are ranked according to their scores in relation to a given user. Those ranking relatively high can then be recommended as potential friends to the given user. Places can be recommended to the user by integrating supplementary data (e.g., POIs, semantic tags) with the mobility histories of these potential friends. Li et al.'s framework [132] has been widely accepted by researchers and followed in several studies [133,134]. For example, Zheng et al. [134] established a fully executable friend–place recommender system based on this framework. Additionally, there exists interest recommendation research using implicit data, which explores beyond friend–place recommendations. For instance, Amato el al. [151,152] described a recommender system on the basis of the interactions among users and generated multimedia content, which can support different social applications using proper customizations (e.g., recommendation of news, photos, art pieces, etc.).

6.3.2. Trip Recommendation

Another issue in this category is trip recommendation. This differs from the route prediction discussed in Sections 5.2.3 and 6.2.2; although trip recommendation also produces a sequence of places to be visited like route prediction, it does not specify the visiting order within the sequence. Similar to interest recommendation, trip recommendation also mines mobility history for frequent patterns with some customization to deal with users' specific trip preferences. For example, Brilhante et al. [135] formulated TripBuilder, which abstracts each user's mobility history as a chronologically ordered (annotated with the start and end times) sequence of POIs and then, profiles each user's trip preferences based on the functional classification of the POIs. This framework utilizes the wisdom of the crowds to find personalized itineraries for a user, given their trip preferences and visiting time budget. Zheng et al. [136] built a complete system that can mine GPS traces to perform two kinds of trip recommendations: a generic one indicating the most interesting places and routes of a given region, and a personalized one that provides the user with places matching their personal preferences.

7. Trajectory Data-Based Services

The information and knowledge obtained from trajectory data mining are applicable to a wide range of services. Services are rooted in real life while being based on the solutions to the application issues discussed in Section 6. Note that there is no specific one-to-one mapping relationship between application issues and trajectory data-based services. In fact, their links are rather flexible depending on specific scenarios. Table 3 lists some examples of trajectory data-based services and their relation to the corresponding application issues.

7.1. Transportation and Urban Planning

Transportation services generally require the characterization of regions (e.g., OD distribution, hotspot distribution) and commuters (e.g., preferred means of transportation, spatiotemporal law of commuting). Occasionally, realworld transportation services also require prediction and trip recommendations. Transportation services include but are not limited to improving the driving experience [153–155], augmenting public transit services [156–158,183], and transportation planning and management [14,160,161].

Table 3. Relationships between trajectory data-based services and application issues.

Services	Service Contents	Application Issues Involved		
		Social Dynamics	**Traffic Dynamics**	**Operational Dynamics**
Transportation	Improving driving experience			Trip recommendation [153–155]
	Augmenting public transit services	Characterization of connections between places [96,156]	Trajectory-based prediction [157–159]	
	Enhancing transportation planning and management	Characterization of connections between places [14,160]	Trajectory-based prediction [161]	
Urban Planning	Understanding urban land use and urban evolution	Characterization of connections between places [120,162]		
	Facilitating urban infrastructure planning	Characterization of connections between places [163,164]		
	Evaluating transportation system	Characterization of connections between places [165,166]		
Environment	Assessing air pollution	Characterization of connections between places [167]		
	Assessing noise pollution	Characterization of connections between places [168,169]		
Energy	Inferring energy consumption	Characterization of connections between places [170]		
	Eco-car infrastructure planning	Characterization of connections between places [171]	Profiling of moving objects [171]	
Social Services	Supporting friend-searching	Discovery of social relationships [133,142]	Profiling of moving objects [133,142]	Interest recommendation [134,142]
	Suggesting routes and places		Profiling of moving objects [172–174]	Trip recommendation [172–174]
	Understanding communities	Discovery of social relationships [175]	Profiling of moving objects [175,176]	

Table 3. *Cont.*

Services	Service Contents	Application Issues Involved		
		Social Dynamics	**Traffic Dynamics**	**Operational Dynamics**
Commercial Services	Optimizing commercial localization	Characterization of connections between places [177]	Profiling of moving objects [177]	Trip recommendation [177]
	Guiding advertising allocation	Characterization of connections between places [178]	Profiling of moving objects [178]	Trip recommendation [178]
	Optimizing department layout	Characterization of connections between places [179]	Profiling of moving objects [179]	
Public Administration	Detecting abnormal behavior		Profiling of moving objects [180]	
	Monitoring public gathering	Detection of social events [181]	Profiling of moving objects [181]	
	Predicting natural disasters		Trajectory-based prediction [182]	

Note: Information in this table is summarized based on our literature survey. Not all cases are listed.

In urban planning, trajectory data mining has two main functions: characterizing locations and characterizing the connections between locations. These two categories can help urban planners understand urban boundary evolution [162], plan urban infrastructure [163,164], assess the transportation system [165,166], etc.

7.2. Environment and Energy

Evaluating the pollution at different locations is a prerequisite for pollution mitigation. To this end, trajectory data mining can be used to characterize places and integrate their properties with supplementary data, e.g., air pollution data [21,167] and noise pollution data [168,169], in order to describe the pollution situations in different regions of the city.

With regard to energy, trajectory data mining is concerned with discovering energy-consuming patterns from a regional or individual perspective. This is related to characterizing places and profiling commuters. Researchers have utilized trajectory data to mine the movement patterns of energy-wasting vehicles [170], establish eco-driving feedback platforms [184], select locations for eco-car charging infrastructure [171], etc.

7.3. Social and Commercial Services and Public Administration

Social services are mainly related to profiling individual movement patterns, discovering social relationships, and recommending interests. Addressing these application issues can help facilitate social services, recommend potential friends [133,134,142], suggest places and routes [172–174], understand community life [175,176], etc.

Commercial services need information regarding the visiting potential of commercial places based on the mobility routines of consumers. Characterizing places and profiling individuals can help improve commercial services, such as optimizing commercial siting [177], guiding advertising allocation [178], and improving department layouts [179].

Public administration is often related to identifying places or individuals that are likely to trigger public incidents. Characterizing places and moving objects are significant in this domain. In addition, public administration requires a certain foresight regarding mobility dynamics; thus, trajectory-based prediction is also important. Researchers have improved public administration by detecting abnormal behavior [180], monitoring public events [181], monitoring and predicting hurricane movement [182,185], etc.

8. Practical Implications

In previous sections, we reviewed the concepts, methodologies, application issues, and resulting services of trajectory data mining from a rigorous academic perspective. In this section, we present an open discussion on the practical implications. For potential participants in the domain of trajectory data mining, a series of commonly used practical tools are recommended. For the most concerned privacy protection problem involved in this domain, a brief survey on current situations is conducted and potential solutions are proposed. There is also a future outlook based on our surveyed literature, which may indicate some directions for authors about trajectory data mining.

8.1. Practical Tools in Trajectory Data Mining

The practice of trajectory data mining methodologies requires the use of certain software tools. As the number of available tools continues to grow, it is increasing difficult to define a most suitable tool, or even to determine the most widely accepted tools nowadays. The typical life cycle of new tools generally begins with theoretical papers as methodological prototypes, followed by demand-responsive software distribution of successful algorithms [186]. These algorithms are either included as a family in new commercial or open-source packages, or are being integrated into existing commercial or open-source packages afterwards.

In fact, from the very beginning, programming languages provide researchers with the initial tools to conduct trajectory data mining. Python, for example, is a universal computer language that is widely applied in this domain [187]. If the users have become familiar with basic programming concepts such as variables, data types, functions, conditions, loops, etc., and are good at learning from online technical communities, such as GitHub, it is not very difficult to build up a basic data mining program. We always encourage researchers to be proficient in one of the computer languages (e.g., Python, C, Fortran), since this will enable them to easily convert algorithms or even algorithm thoughts into practice, without being restricted by the software platform.

Apart from pure programming languages, open-source software can be another good choice for researchers and learners in trajectory data mining. R is both a computer language and software that is powerful in statistical analysis [188]. Although its core computing modules are written in C, C++, and Fortran, it also provides a scripting language, i.e., R language, for customized programming. A series of analysis techniques, including statistical testing, predictive modeling, and data visualization, is supported by R. WEKA is another famous and powerful open-source software for data mining [189], which supports data preprocessing, data collection, classification, regression analysis, visualization, feature selection and many other machine learning functions. Advanced users can call its components through Java programming and command lines, while it also provides a graphical inference for basic users. KNIME is a platform that can be extended to use the mining algorithms in WEKA. Besides, it integrates many other data science tools covering data management, modeling, deploying, reporting, etc. [190]. KNIME uses a data flow-like approach to establish the mining process, which is composed of a series of functional nodes. Each node has an input port for receiving data and models, and an output port for exporting results, thus, users can easily connect to the nodes for process management. RapidMiner is also an extendable platform that can be applied in this domain. It owns a specific advantage in machine learning, by providing support for any third-party machine learning libraries [191].

Although the open-source software mentioned above can assist us in trajectory data mining, the power of commercial software can hardly be ignored. In order to obtain greater profits in a limited product life cycle, these tools are made more attractive in terms of user-friendliness and strong service support. IBM SPSS Modeler is a representative commercial tool for mining tasks. It is equipped with an intuitive user interface and allows users to create various algorithms without programming [192]. Oracle Data Mining (ODM) is a component of the Oracle platform, which is a world-famous database management tool. It enables users to build and apply models directly inside their Oracle Database [193]. SAS Data Mining is another commercial option with user-friendly GUI and specific strength in predictive modeling and prescriptive modeling [194].

Advances in computing power have enabled us to move beyond manual and time-consuming mining practice to quick and automated data analysis, meanwhile bringing about many powerful tools for trajectory data mining. Each tool has its own strengths. For scholars who need to dig deeply into the philosophy and methodology of data mining, we recommend programming languages and open-source tools with more customization possibilities. For business users who pursue practical efficiency and stable output, highly integrated commercial software may be a better choice.

8.2. Privacy Protection in Trajectory Data Mining

With the ubiquity of smart devices and the improvement of powerful data mining techniques, there are increasing concerns that trajectory data mining may pose a threat to our privacy and information security. However, we need to notify that the majority of applications in this domain are not deeply concerned with private information.

Consider a most extreme case that is happening at this moment: scientists are utilizing multi-source trajectory data to trace the transmission chain of COVID-19 (coronavirus disease). The major objective is to act quickly, when a person is diagnosed with COVID-19, to find all the people this person was in close proximity with [195]. One popular approach is contact tracing based on "check points" as

suggested by Yasaka et al. [196], which uses an anonymized graph of interpersonal interactions to report risk levels to users. This process does not technically need any location information or personal data. Another approach is using Bluetooth-based smartphone apps, e.g., the TraceTogether app from the Singaporean government. Such apps cryptographically create a new temporary ID periodically and utilize Bluetooth's near-field communication function to record IDs of close contact. If any user is diagnosed with COVID-19, the doctor will instruct them to share locally stored data with the central server. The server will obtain all the temporary IDs the "infected phone" has been in contact with, and then, inform them with a push token technique.

As indicated above, even under the conditions of a pandemic, we are still able to avoid privacy offences when taking advantage of trajectory data, typically with two ideas: the first is to represent personal information with virtual IDs during the process of tracing and publicity; the second is to replace absolute geographic coordinates with relative position information when unnecessary, e.g., when no mass infection is detected. In fact, apart from the emergent situations concerning individuals, the major focus of trajectory data mining is on the discovery of general or significant patterns, not on the specific information regarding individuals. For this reason, we believe that the real concerns are with unconstrained access to individual records, especially privacy-sensitive information such as religious, financial, or healthcare records that usually come along with implicit trajectory data. For applications that do involve such information, simple desensitization approaches, such as removing sensitive IDs from data, or to a more advanced degree, such as randomization methods [197] and encryption methods [198], are sufficient enough to protect the privacy of most individuals.

Nevertheless, privacy protection discussed here is from a pure technical perspective. In the real world, concerns cannot be completely eliminated whenever and wherever sensitive information is collected and stored in a digital form. Like any other technology, trajectory data mining is possible to be misused. Thus, not only the researchers in the general data mining domain, but also those in the fields of database encryption, counterterrorism, and social sciences, are expected to work with lawyers, politicians, entrepreneurs, and consumers to take responsibility in establishing solutions to protect personal privacy and data security.

8.3. Future Prospects for Trajectory Data Mining

The diversity of trajectory data, mining methods, and mining applications has brought about a series of challenging research issues. From the surveyed literature in this paper, we can glimpse some development trends in trajectory data mining and provide suggestions for authors in this filed.

The first significant trend is to combine trajectory data with other data sources to fulfil a mining task. The rationality of integrating multi-source heterogeneous data first lies in avoiding information bias, or in other words, enriching trajectory information with other sources. An example can be found in Wang et al. [199], which leveraged POIs and road network data to fill in the missing information in sparse trajectories, in order to better estimate the travel time of a path in a city's road network. On the other hand, such combination may unlock the potential power of knowledge that can hardly be discovered from a single data source. For instance, Zheng et al. [200] inferred the fine-grained noise situation of different times of day for each region of New York City by using the 311 complaint data together with social media check-in data, road network data, and POIs.

The second prospect in this field is the development of scalable and interactive mining methods. Unlike traditional data analysis, data mining must be able to process huge amounts of data effectively, and if possible, interactively. As the amount of data increases rapidly, it is essential to develop more scalable algorithms for mining tasks. The incremental trajectory clustering algorithms in Li et al. [51], for example, are an early-stage attempt to deal with such dynamic data growth. To this end, Ding et al. [201] established a united platform named Ultraman to achieve scalability and efficiency when dealing with big trajectory data, by extending Apache Spark with an integrated key-value store and enhancing the MapReduce paradigm to allow flexible optimizations based on random data access. Practical direction to improve the overall efficiency and user interaction is constraint-based

mining. As integrated in many practical tools such as KNIME [190], it endows users with added control by allowing specifications and constraints to guide the mining workflows in their search for interesting knowledge.

Beyond trajectory data mining itself, privacy handling is another task to be tackled in the future. There exists an underlying balance between a sufficient degree of useful knowledge and the ethics of tracking activities. Although there have been many technical approaches to protect personal privacy as mentioned previously (e.g., randomization [197] and encryption [198]), no promising methods to achieve or even measure this balance have yet been developed. In addition, the best results with real-life relevance of trajectory data mining can be achieved by interdisciplinary efforts. This community is expecting more research with extensive practical significance and interdisciplinary influence.

8.4. Trajectory Data Mining in Industry 4.0

Recent decades have witnessed that people and things are becoming increasingly interconnected. Smartphones, vehicles, devices, built environments, and natural environments have been filled with digital sensors, all of which are generating unprecedented big data, including trajectory data as we have discussed. Urban realities mined from such data make it possible that demand-responsive urban services be highly realized. For instance, the digitization journey in the transportation and logistics sector is already well underway and is expected to accelerate in the immediate future. Companies are using IoT solutions, such as big data analytics, for demand forecasting and, in turn, optimizing inventory planning, warehousing fulfillment, and distribution [202]. The consequence of developing such IOT solutions and Big Data science is the conception of Industry 4.0 [203].

Industry 4.0 refers to the fourth industrial revolution. Its initiative places significant emphasis on the utilization of data to form intelligent systems and processes, so that in this context, manufacturing and service-providing will largely have the ability to self-plan and self-adapt [203]. To accomplish the vision of Industry 4.0, or even to survive in the "Digital Darwinism", digital transformation is a common task faced by individuals, enterprises, and countries [204]. Yet, even in highly integrated Europe, a digital divide does exist among the member states [205]. Northern Europe takes the lead in terms of the utilization of innovative industries, such as the application of big data, while there also exists a worrisome trend of the European countries lagging behind other global leaders of Industry 4.0, such as the USA and China [205]. Our literature survey on trajectory data mining, to some extent, has also confirmed this trend, as the majority of research cases in this specific domain are contributed by countries with Industry 4.0 advantages.

In the new digital divide, Industry 4.0 will play an important role for individuals, in that routine jobs are likely to be replaced by those requiring analytical skills, flexibility in decision making, and training in certain topics, such as trajectory data mining, text mining, and machine learning, as proved by many recent surveys on Industry 4.0 employment demands, for example, in Bach et al. [206] and Fareri et al. [207]. Regardless of the social skill requirements of Industry 4.0 job positions, trajectory data mining is of educational advancement from a technical perspective. Not only can it address the specific application issues mentioned previously, it also provides a methodology for data evidence-based decision making, which will largely benefit future participants in Industry 4.0. Therefore, the practical implications here indicate the need for interventions of education, such as curriculum with a focus on big data acquisition, management, mining, and analysis.

9. Conclusions

Advances in smart infrastructure and location-acquisition terminals have contributed to the increasing availability of massive trajectory data with rich spatiotemporal information on the mobility of a wide range of moving objects, including humans, vehicles, and animals. By developing data mining and analysis methods, researchers have revealed abundant urban realities from trajectory datasets, such as the movement patterns of humans, the inter-place relationships within cities, and the dynamics of social events to solve complex urban problems in transportation, environment, public security, etc.

However, despite the wealth of information in this field, existing studies have been relatively isolated and lacking an integrated and systematic survey to the issues that have been addressed, solutions that have been tested, and services that have been developed. This paper was an attempt at conducting such a survey. We started with classifying diverse trajectory data and reviewed the prevailing trajectory mining methods in two classes. We classified the application issues of trajectory data mining into three major groups based on how they are related: social dynamics, traffic dynamics, and operational dynamics. We then built up matching relationships between data mining methods and application issues, and briefly presented the prospects of services that have been established based on the methods and techniques in this field. A series of open discussions was also conducted regarding the practical tools, ethics, and future directions of trajectory data mining.

The major contribution of this paper is in providing a systematic and integrated view on the emerging issues, methodologies, and services in trajectory data mining while identifying the inherent relevance and associations that have previously been unrepresented in this domain. The classification of application issues can help readers to identify new issues where trajectory data mining could be applied. The relevance between mining methods and the matching relationship between mining methods and application issues can contribute to identifying method gaps and inspire researchers to develop new methods. The consistent association linking methods, application issues, and services can provide a reference for data analysts and experts to select the most suitable solutions for specific problems. This paper can also provide new researchers with a quick understanding of trajectory data mining.

The main limitation of our work lies in the manual selection of surveyed literature, which may lead to certain information bias. Besides, our work is based on the review of the articles, reports, and book sections that can be accessed through Web of Science at present. Due to the time difference between literature creation and publication, as well as the imperfect search function, specific research, e.g., the latest ones or those Web of Science has no authorization to disclose, might be invisible in our review.

The smallest unit of analysis in our work is the application-oriented method. In fact, within each method, there correspondingly exist a series of algorithms. Therefore, our work can be extended by conducting a comparative analysis on the algorithms that implement a specific mining method on specific application issues, so that the review will be able to provide a reference for the choice between algorithms. Additionally, research work concerning the future prospects we have discussed in the section of practical implications may also be promising under the current context.

References

1. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [CrossRef]
2. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Gener. Comput. Syst.* **2013**, *29*, 1645–1660. [CrossRef]
3. Zheng, Y. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 29. [CrossRef]
4. Gudmundsson, J.; Laube, P.; Wolle, T. Computational movement analysis. In *Springer Handbook of Geographic Information*; Kresse, W., Danko, D.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 423–438.
5. Feng, Z.; Zhu, Y. A Survey on Trajectory Data Mining: Techniques and Applications. *IEEE Access* **2016**, *4*, 2056–2067. [CrossRef]

6.　Castro, P.S.; Zhang, D.; Chen, C.; Li, S.; Pan, G. From taxi GPS traces to social and community dynamics: A survey. *ACM Comput. Surv.* **2013**, *46*, 17. [CrossRef]

7.　Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*; Fayyad, U., Piatetsky-Shapiro, G., Amith, P., Smyth, R.U., Eds.; AAAI Press: Palo Alto, CA, USA, 1996; pp. 1–34.

8.　Maimon, O.; Rokach, L. Introduction to knowledge discovery and data mining. In *Data Mining and Knowledge Discovery Handbook*; Rokach, O.M.L., Ed.; Springer: Boston, MA, USA, 2010; pp. 1–15.

9.　Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2011.

10.　Mennis, J.; Guo, D. Spatial data mining and geographic knowledge discovery—An introduction. *Comput. Environ. Urban Syst.* **2009**, *33*, 403–408. [CrossRef]

11.　Miller, H.J.; Han, J. Geographic data mining and knowledge discovery. In *The Handbook of Geographic Information Science*; Wilson, J.P., Fotheringham, A.S., Eds.; Wiley-Blackwell: Hoboken, NJ, USA, 2007; pp. 352–366.

12.　Kong, X.; Li, M.; Ma, K.; Tian, K.; Wang, M.; Ning, Z.; Xia, F. Big trajectory data: A survey of applications and services. *IEEE Access* **2018**, *6*, 58295–58306. [CrossRef]

13.　Andrienko, G.; Andrienko, N.; Bak, P.; Keim, D.; Kisilevich, S.; Wrobel, S. A conceptual framework and taxonomy of techniques for analyzing movement. *J. Vis. Lang. Comput.* **2011**, *22*, 213–232. [CrossRef]

14.　Chen, C.; Zhang, D.; Zhou, Z.-H.; Li, N.; Atmaca, T.; Li, S. B-Planner: Night Bus Route Planning Using Large-Scale Taxi GPS Traces. In Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications, San Diego, CA, USA, 18–22 March 2013; pp. 225–233.

15.　Tang, J.; Liu, F.; Wang, Y.; Wang, H. Uncovering urban human mobility from large scale taxi GPS data. *Phys. A Stat. Mech. Its Appl.* **2015**, *438*, 140–153. [CrossRef]

16.　Zhang, D.; Sun, L.; Li, B.; Chen, C.; Pan, G.; Li, S.; Wu, Z. Understanding taxi service strategies from taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 123–135. [CrossRef]

17.　Niazi, M. Do systematic literature reviews outperform informal literature reviews in the software engineering domain? An initial case study. *Arab. J. Sci. Eng.* **2015**, *40*, 845–855. [CrossRef]

18.　Pejić Bach, M.; Krstić, Ž.; Seljan, S.; Turulja, L. Text mining for big data analysis in financial sector: A literature review. *Sustainability* **2019**, *11*, 1277. [CrossRef]

19.　Wahono, R.S. A systematic literature review of software defect prediction. *J. Softw. Eng.* **2015**, *1*, 1–16.

20.　Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Altman, D.; Antes, G.; Atkins, D.; Barbour, V.; Barrowman, N.; Berlin, J.A.; et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]

21.　Zheng, Y.; Capra, L.; Wolfson, O.; Yang, H. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.* **2014**, *5*, 38. [CrossRef]

22.　Zheng, Y.; Liu, Y.; Yuan, J.; Xie, X. Urban computing with taxicabs. In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 89–98.

23.　Renso, C.; Spaccapietra, S.; Zimányi, E. *Mobility Data: Modeling, Management, and Understanding*; Cambridge University Press: New York, NY, USA, 2013.

24.　Pelekis, N.; Theodoridis, Y. *Mobility Data Management and Exploration*; Springer: New York, NY, USA, 2014.

25.　Cartlidge, J.; Gong, S.; Bai, R.; Yue, Y.; Li, Q.; Qiu, G. Spatio-temporal prediction of shopping behaviours using taxi trajectory data. In Proceedings of the 2018 IEEE 3rd International Conference on Big Data Analysis, Shanghai, China, 9–12 March 2018; pp. 112–116.

26.　Hao, J.; Zhu, J.; Zhong, R. The rise of big data on urban studies and planning practices in China: Review and open research issues. *J. Urban Manag.* **2015**, *4*, 92–124. [CrossRef]

27.　Xu, X.; Xie, L.; Li, H.; Qin, L. Learning the route choice behavior of subway passengers from AFC data. *Expert Syst. Appl.* **2018**, *95*, 324–332. [CrossRef]

28.　Chen, Z.; Zou, H.; Jiang, H.; Zhu, Q.; Soh, Y.C.; Xie, L. Fusion of WiFi, smartphone sensors and landmarks using the kalman filter for indoor localization. *Sensors* **2015**, *15*, 715–732. [CrossRef]

29.　Werner, M.; Schauer, L.; Scharf, A. Reliable trajectory classification using Wi-Fi signal strength in indoor scenarios. In Proceedings of the 2014 IEEE/ION Position, Location and Navigation Symposium, Monterey, CA, USA, 5–8 May 2014; pp. 663–670.

30. Barbier, G.; Liu, H. Data mining in social media. In *Social Network Data Analytics*; Aggarwal, C., Ed.; Springer: Boston, MA, USA, 2011; pp. 327–352.

31. Brook, A.; Ben-Dor, E.; Richter, R. Modelling and monitoring urban built environment via multi-source integrated and fused remote sensing data. *Int. J. Image Data Fusion* **2013**, *4*, 2–32. [CrossRef]

32. Liu, B.; Fu, Y.; Yao, Z.; Xiong, H. Learning geographical preferences for point-of-interest recommendation. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 1043–1051.

33. Ye, M.; Yin, P.; Lee, W.-C.; Lee, D.-L. Exploiting geographical influence for collaborative point-of-interest recommendation. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, 24–28 July 2011; pp. 325–334.

34. Lee, J.-G.; Han, J.; Li, X.; Cheng, H. Mining discriminative patterns for classifying trajectories on road networks. *IEEE Trans. Knowl. Data Eng.* **2010**, *23*, 713–726. [CrossRef]

35. Demšar, U.; Virrantaus, K. Space–time density of trajectories: Exploring spatio-temporal patterns in movement data. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1527–1542. [CrossRef]

36. Alarabi, L.; Eldawy, A.; Alghamdi, R.; Mokbel, M.F. TAREEG: A MapReduce-based system for extracting spatial data from OpenStreetMap. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas, TX, USA, 4–7 November 2014; pp. 83–92.

37. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *Int. J. Remote Sens.* **2015**, *36*, 3144–3169. [CrossRef]

38. Arsanjani, J.J.; Zipf, A.; Mooney, P.; Helbich, M. An introduction to OpenStreetMap in Geographic Information Science: Experiences, research, and applications. In *OpenStreetMap in GIScience*; Arsanjani, J.J., Zipf, A., Mooney, P., Helbich, M., Eds.; Springer: Cham, Switzerland, 2015; pp. 1–15.

39. Müller, H.; Freytag, J.-C. *Problems, Methods, and Challenges in Comprehensive Data Cleansing*; Humboldt-Universität zu Berlin: Berlin, Germany, 2005.

40. Hernández, M.A.; Stolfo, S.J. Real-world data is dirty: Data cleansing and the merge/purge problem. *Data Min. Knowl. Discov.* **1998**, *2*, 9–37. [CrossRef]

41. Dasu, T.; Johnson, T. *Exploratory Data Mining and Data Cleaning*; John Wiley & Sons: Hoboken, NJ, USA, 2003; Volume 479.

42. Xu, R.; Wunsch, D. *Clustering*; John Wiley & Sons: Hoboken, NJ, USA, 2008; Volume 10.

43. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678. [CrossRef] [PubMed]

44. Yuan, G.; Sun, P.H.; Zhao, J.; Li, D.X.; Wang, C.W. A review of moving object trajectory clustering algorithms. *Artif. Intell. Rev.* **2017**, *47*, 123–144.

45. Rokach, L. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2009; pp. 269–298.

46. Gaffney, S.; Smyth, P. Trajectory clustering with mixtures of regression models. In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 15–18 August 1999; pp. 63–72.

47. Cadez, I.V.; Gaffney, S.; Smyth, P. A general probabilistic framework for clustering individuals and objects. In Proceedings of the 6th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, 23–27 August 2000; pp. 140–149.

48. Alon, J.; Sclaroff, S.; Kollios, G.; Pavlovic, V. Discovering clusters in motion time-series data. In Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 16–22 June 2003.

49. Lee, J.-G.; Han, J.; Whang, K.-Y. Trajectory clustering: A partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD Conference on Management of Data, Beijing, China, 12–14 June 2007; pp. 593–604.

50. Chen, J.; Leung, M.K.; Gao, Y. Noisy logo recognition using line segment Hausdorff distance. *Pattern Recognit.* **2003**, *36*, 943–955. [CrossRef]

51. Li, Z.; Lee, J.-G.; Li, X.; Han, J. Incremental clustering for trajectories. In *International Conference on Database Systems for Advanced Applications*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 32–46.

52. Kharrat, A.; Popa, I.S.; Zeitouni, K.; Faiz, S. Clustering algorithm for network constraint trajectories. In *Headway in Spatial Data Handling*; Ruas, A.G.C., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 631–647.

53. Miwa, T.; Kiuchi, D.; Yamamoto, T.; Morikawa, T. Development of map matching algorithm for low frequency probe data. *Transp. Res. Part C Emerg. Technol.* **2012**, *22*, 132–145. [CrossRef]

54. Quddus, M.A.; Ochieng, W.Y.; Noland, R.B. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp. Res. Part C Emerg. Technol.* **2007**, *15*, 312–328. [CrossRef]

55. Schwenker, F.; Trentin, E. Pattern classification and clustering: A review of partially supervised learning approaches. *Pattern Recognit. Lett.* **2014**, *37*, 4–14. [CrossRef]

56. Press, S.J.; Wilson, S. Choosing between logistic regression and discriminant analysis. *J. Am. Stat. Assoc.* **1978**, *73*, 699–705. [CrossRef]

57. Gold, C.; Sollich, P. Model selection for support vector machine classification. *Neurocomputing* **2003**, *55*, 221–249. [CrossRef]

58. Xie, F.; Quinlan, J.R. Induction on decision tree. *Mach. Learn.* **1986**, *1*, 81–106.

59. Coomans, D.; Massart, D.L. Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules. *Anal. Chim. Acta* **1982**, *136*, 15–27. [CrossRef]

60. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised Machine Learning: A Review of Classification Techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* **2007**, *160*, 3–24.

61. Bolbol, A.; Cheng, T.; Tsapakis, I.; Haworth, J. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Comput. Environ. Urban Syst.* **2012**, *36*, 526–537. [CrossRef]

62. Zheng, Y.; Chen, Y.K.; Li, Q.N.; Xie, X.; Ma, W.Y. Understanding Transportation Modes Based on GPS Data for Web Applications. *ACM Trans. Web* **2010**, *4*, 1–36. [CrossRef]

63. Lee, J.-G.; Han, J.; Li, X.; Gonzalez, H. TraClass: Trajectory classification using hierarchical region-based and trajectory-based clustering. *Proc. VLDB Endow.* **2008**, *1*, 1081–1094. [CrossRef]

64. Nascimento, J.C.; Figueiredo, M.A.T.; Marques, J.S. Trajectory classification using switched dynamical hidden Markov models. *IEEE Trans. Image Process.* **2009**, *19*, 1338–1348. [CrossRef]

65. Fan, W.; Bifet, A. Mining big data: Current status, and forecast to the future. *ACM SIGKDD Explor. Newsl.* **2013**, *14*, 2. [CrossRef]

66. Nasreen, S.; Azam, M.A.; Shehzad, K.; Naeem, U.; Ghazanfar, M.A. Frequent pattern mining algorithms for finding associated frequent patterns for data streams: A survey. *Procedia Comput. Sci.* **2014**, *37*, 109–116. [CrossRef]

67. Wu, J.; Zhu, X.; Zhang, C.; Philip, S.Y. Bag constrained structure pattern mining for multi-graph classification. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 2382–2396. [CrossRef]

68. Febrer-Hernández, J.K.; Hernández-Palancar, J. Sequential pattern mining algorithms review. *Intell. Data Anal.* **2012**, *16*, 451–466. [CrossRef]

69. Zhang, D.; Lee, K.; Lee, I. Periodic pattern mining for spatio-temporal trajectories: A survey. In Proceedings of the 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Taipei, Taiwan, 24–27 November 2015; pp. 306–313.

70. Jeung, H.; Yiu, M.L.; Jensen, C.S. Trajectory pattern mining. In *Computing with Spatial Trajectories*; Zheng, Y., Zhou, X., Eds.; Springer: New York, NY, USA, 2011; pp. 143–177.

71. Cao, H.; Mamoulis, N.; Cheung, D.W. Discovery of periodic patterns in spatiotemporal sequences. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 453–467. [CrossRef]

72. Mamoulis, N.; Cao, H.; Kollios, G.; Hadjieleftheriou, M.; Tao, Y.; Cheung, D.W. Mining, indexing, and querying historical spatiotemporal data. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 236–245.

73. Li, Z.; Ding, B.; Han, J.; Kays, R.; Nye, P. Mining periodic behaviors for moving objects. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington DC, USA, 24–28 July 2010; pp. 1099–1108.

74. Li, Z.; Han, J.; Ding, B.; Kays, R. Mining periodic behaviors of object movements for animal and biological sustainability studies. *Data Min. Knowl. Discov.* **2012**, *24*, 355–386. [CrossRef]

75. Körner, C.; May, M.; Wrobel, S. Spatiotemporal modeling and analysis—Introduction and overview. *Künstliche Intelligenz* **2012**, *26*, 215–221. [CrossRef]

76. Cao, H.; Mamoulis, N.; Cheung, D.W. Mining frequent spatio-temporal sequential patterns. In Proceedings of the 5th IEEE International Conference on Data Mining, Houston, TX, USA, 27–30 November 2005; pp. 82–89.

77. Orellana, D.; Bregt, A.K.; Ligtenberg, A.; Wachowicz, M. Exploring visitor movement patterns in natural recreational areas. *Tour. Manag.* **2012**, *33*, 672–682. [CrossRef]

78. Giannotti, F.; Nanni, M.; Pedreschi, D. Efficient mining of temporally annotated sequences. In Proceedings of the 6th SIAM International Conference on Data Mining, Bethesda, MD, USA, 20–22 April 2006; pp. 348–359.

79. Giannotti, F.; Nanni, M.; Pinelli, F.; Pedreschi, D. Trajectory pattern mining. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 330–339.

80. Kang, J.; Yong, H.-S. Mining spatio-temporal patterns in trajectory data. *J. Inf. Process. Syst.* **2010**, *6*, 521–536. [CrossRef]

81. Benkert, M.; Gudmundsson, J.; Hübner, F.; Wolle, T. Reporting flock patterns. *Comput. Geom.* **2008**, *41*, 111–125. [CrossRef]

82. Wachowicz, M.; Ong, R.; Renso, C.; Nanni, M. Finding moving flock patterns among pedestrians through collective coherence. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1849–1864. [CrossRef]

83. Jeung, H.; Yiu, M.L.; Zhou, X.; Jensen, C.S.; Shen, H.T. Discovery of convoys in trajectory databases. *Proc. VLDB Endow.* **2008**, *1*, 1068–1080. [CrossRef]

84. Yoon, H.; Shahabi, C. Accurate discovery of valid convoys from moving object trajectories. In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, Miami, FL, USA, 6–9 December 2009; pp. 636–643.

85. Li, Z.; Ding, B.; Han, J.; Kays, R. Swarm: Mining relaxed temporal moving object clusters. *Proc. VLDB Endow.* **2010**, *3*, 723–734. [CrossRef]

86. Vieira, M.R.; Bakalov, P.; Tsotras, V.J. On-line discovery of flock patterns in spatio-temporal data. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 286–295.

87. Jeung, H.; Shen, H.T.; Zhou, X. Convoy queries in spatio-temporal databases. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 1457–1459.

88. Li, Z.H.; Han, J.W.; Ji, M.; Tang, L.A.; Yu, Y.T.; Ding, B.L.; Lee, J.G.; Kays, R. MoveMine: Mining Moving Object Data for Discovery of Animal Movement Patterns. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 4. [CrossRef]

89. Yu, Y.; Wang, Q.; Kuang, J.; He, J. TGCR: An efficient algorithm for mining swarm in trajectory databases. In Proceedings of the 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services, Fuzhou, China, 29 June–1 July 2011; pp. 90–95.

90. Hodge, V.; Austin, J. A survey of outlier detection methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [CrossRef]

91. Lee, J.-G.; Han, J.; Li, X. Trajectory outlier detection: A partition-and-detect framework. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 140–149.

92. Zhang, D.; Li, N.; Zhou, Z.-H.; Chen, C.; Sun, L.; Li, S. iBAT: Detecting anomalous taxi trajectories from GPS traces. In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September 2011; pp. 99–108.

93. Knorr, E.M.; Ng, R.T.; Tucakov, V. Distance-based outliers: Algorithms and applications. *VLDB J.* **2000**, *8*, 237–253. [CrossRef]

94. Liu, L.; Qiao, S.; Zhang, Y.; Hu, J. An efficient outlying trajectories mining approach based on relative distance. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1789–1810. [CrossRef]

95. Yuan, G.; Xia, S.; Zhang, L.; Zhou, Y.; Ji, C. Trajectory outlier detection algorithm based on structural features. *J. Comput. Inf. Syst.* **2011**, *7*, 4137–4144.

96. Li, X.; Han, J.; Kim, S.; Gonzalez, H. Roam: Rule-and motif-based anomaly detection in massive moving object data sets. In *Proceedings of the 7th SIAM International Conference on Data Mining*; SIAM: Minneapolis, MN, USA, 2007; pp. 273–284.

97. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.

98. Gidófalvi, G.; Dong, F. When and where next: Individual mobility prediction. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems, Redondo Beach, CA, USA, 6 November 2012; pp. 57–64.

99.　Trasarti, R.; Guidotti, R.; Monreale, A.; Giannotti, F. Myway: Location prediction via mobility profiling. *Inf. Syst.* **2017**, *64*, 350–367. [CrossRef]

100.　Jeung, H.; Liu, Q.; Shen, H.T.; Zhou, X. A hybrid prediction model for moving objects. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 70–79.

101.　Krumm, J.; Horvitz, E. Predestination: Inferring destinations from partial trajectories. In *Proceedings of the 8th International Conference on Ubiquitous Computing*; Springer: Orange County, CA, USA, 2006; pp. 243–260.

102.　Backstrom, L.; Sun, E.; Marlow, C. Find me if you can: Improving geographical prediction with social and spatial proximity. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 61–70.

103.　Gambs, S.; Killijian, M.-O.; del Prado Cortez, M.N. Next place prediction using mobility markov chains. In Proceedings of the 1st Workshop on Measurement, Privacy, and Mobility, Bern, Switzerland, 10 April 2012; pp. 1–6.

104.　Asahara, A.; Maruyama, K.; Sato, A.; Seto, K. Pedestrian-movement prediction based on mixed Markov-chain model. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 25–33.

105.　Mathew, W.; Raposo, R.; Martins, B. Predicting future locations with hidden Markov models. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 911–918.

106.　Ying, J.J.-C.; Lee, W.-C.; Tseng, V.S. Mining geographic-temporal-semantic patterns in trajectories for location prediction. *ACM Trans. Intell. Syst. Technol.* **2014**, *5*, 2. [CrossRef]

107.　Monreale, A.; Pinelli, F.; Trasarti, R.; Giannotti, F. Wherenext: A location predictor on trajectory pattern mining. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 637–646.

108.　Chen, L.; Lv, M.Q.; Ye, Q.A.; Chen, G.C.; Woodward, J. A personal route prediction system based on trajectory data mining. *Inf. Sci.* **2011**, *181*, 1264–1284. [CrossRef]

109.　Krumm, J.; Gruen, R.; Delling, D. From destination prediction to route prediction. *J. Locat. Based Serv.* **2013**, *7*, 98–120. [CrossRef]

110.　Tiwari, V.S.; Chaturvedi, S.; Arya, A. Route prediction using trip observations and map matching. In Proceedings of the 2013 3rd IEEE International Advance Computing Conference (IACC), Ghaziabad, India, 22–23 February 2013; pp. 583–587.

111.　Simmons, R.; Browning, B.; Zhang, Y.; Sadekar, V. Learning to predict driver route and destination intent. In Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, Toronto, ON, Canada, 17–20 September 2006; pp. 127–132.

112.　Jeung, H.; Yiu, M.L.; Zhou, X.; Jensen, C.S. Path prediction and predictive range querying in road network databases. *VLDB J.* **2010**, *19*, 585–602. [CrossRef]

113.　Krumm, J. Where will they turn: Predicting turn proportions at intersections. *Pers. Ubiquitous Comput.* **2010**, *14*, 591–599. [CrossRef]

114.　Gao, P.; Kupfer, J.A.; Zhu, X.; Guo, D. Quantifying animal trajectories using spatial aggregation and sequence analysis: A case study of differentiating trajectories of multiple species. *Geogr. Anal.* **2016**, *48*, 275–291. [CrossRef]

115.　Ying, J.J.-C.; Lee, W.-C.; Weng, T.-C.; Tseng, V.S. Semantic trajectory mining for location prediction. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 34–43.

116.　Anagnostopoulos, T.; Anagnostopoulos, C.; Hadjiefthymiades, S. An online adaptive model for location prediction. In *Proceedings of the International Conference on Autonomic Computing and Communications Systems*; Springer: Limassol, Cyprus, 2009; pp. 64–78.

117.　Gaito, S.; Rossi, G.P.; Zignani, M. From mobility data to social attitudes: A complex network approach. In Proceedings of the Workshop on Finding Patterns of Human Behaviors in Networks and Mobility Data, Athens, Greece, 5–9 September 2011; pp. 52–65.

118.　Zheng, K.; Zheng, Y.; Yuan, N.J.; Shang, S.; Zhou, X. Online discovery of gathering patterns over trajectories. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1974–1988. [CrossRef]

119.　Calabrese, F.; Pereira, F.C.; Di Lorenzo, G.; Liu, L.; Ratti, C. The geography of taste: Analyzing cell-phone mobility and social events. In *Proceedings of the International Conference on Pervasive Computing*; Springer: Helsinki, Finland, 2010; pp. 22–37.

120. Liu, Y.; Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landsc. Urban Plan.* **2012**, *106*, 73–87. [CrossRef]

121. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-use classification using taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 113–123. [CrossRef]

122. Liu, L.; Biderman, A.; Ratti, C. Urban mobility landscape: Real time monitoring of urban mobility patterns. In *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*; Citeseer: Hong Kong, China, 2009; pp. 1–16.

123. Rinzivillo, S.; Mainardi, S.; Pezzoni, F.; Coscia, M.; Pedreschi, D.; Giannotti, F. Discovering the geographical borders of human mobility. *Künstliche Intelligenz* **2012**, *26*, 253–260. [CrossRef]

124. Fontes, V.C.; de Alencar, L.A.; Renso, C.; Bogorny, V.; Pisa, I. Discovering Trajectory Outliers between Regions of Interest. In *Proceedings of the Brazilian Symposium on GeoInformatics*; Citeseer: São Paulo, Brazil, 2013; pp. 49–60.

125. Reumers, S.; Liu, F.; Janssens, D.; Cools, M.; Wets, G. Semantic annotation of global positioning system traces: Activity type inference. *Transp. Res. Rec.* **2013**, *2383*, 35–43. [CrossRef]

126. Shamoun-Baranes, J.; Bom, R.; van Loon, E.E.; Ens, B.J.; Oosterbeek, K.; Bouten, W. From sensor data to animal behaviour: An oystercatcher example. *PLoS ONE* **2012**, *7*, e37997. [CrossRef]

127. Chen, X.; Pang, J.; Xue, R. Constructing and comparing user mobility profiles for location-based services. In Proceedings of the 28th Annual ACM Symposium on Applied Computing, Coimbra, Portugal, 18–22 March 2013; pp. 261–266.

128. Chen, X.; Pang, J.; Xue, R. Constructing and comparing user mobility profiles. *ACM Trans. Web* **2014**, *8*, 1–25. [CrossRef]

129. Carneiro, C.; Alp, A.; Macedo, J.; Spaccapietra, S. Advanced Data Mining Method for Discovering Regions and Trajectories of Moving Objects:"Ciconia Ciconia" Scenario. In *The European Information Society*; Bernard, L., Friis-Cristensen, A., Pundt, H., Eds.; Springer: Berlin, Germany, 2008; pp. 201–224.

130. Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F.; Renso, C.; Rinzivillo, S.; Trasarti, R. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* **2011**, *20*, 695–719. [CrossRef]

131. Castro, P.S.; Zhang, D.; Li, S. Urban traffic modelling and prediction using large scale taxi gps traces. In *Proceedings of the 10th International Conference on Pervasive Computing*; Springer: Newcastle, UK, 2012; pp. 57–72.

132. Li, Q.; Zheng, Y.; Xie, X.; Chen, Y.; Liu, W.; Ma, W.-Y. Mining user similarity based on location history. In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Irvine, CA, USA, 5–7 November 2008; pp. 1–10.

133. Xiao, X.; Zheng, Y.; Luo, Q.; Xie, X. Finding similar users using category-based location history. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 442–445.

134. Zheng, Y.; Zhang, L.Z.; Ma, Z.X.; Xie, X.; Ma, W.Y. Recommending Friends and Locations Based on Individual Location History. *ACM Trans. Web* **2011**, *5*, 1–44. [CrossRef]

135. Brilhante, I.; Macedo, J.A.; Nardini, F.M.; Perego, R.; Renso, C. Where shall we go today? Planning touristic tours with TripBuilder. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 757–762.

136. Zheng, Y.; Xie, X. Learning travel recommendations from user-generated GPS traces. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 2. [CrossRef]

137. Peng, C.; Jin, X.; Wong, K.-C.; Shi, M.; Liò, P. Collective human mobility pattern from taxi trips in urban area. *PLoS ONE* **2012**, *7*, e34487.

138. Chang, H.W.; Tai, Y.C.; Hsu, Y.; Jen, J.; Hsu, J.Y. Context-aware taxi demand hotspots prediction. *Int. J. Bus. Intell. Data Min.* **2010**, *5*, 3–18. [CrossRef]

139. Moreira-Matias, L.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; Damas, L. Predicting taxi–passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1393–1402. [CrossRef]

140. Liu, X.; Gong, L.; Gong, Y.; Liu, Y. Revealing travel patterns and city structure with taxi trip data. *J. Transp. Geogr.* **2015**, *43*, 78–90. [CrossRef]

141. Gilpin, M.E. Spiral chaos in a predator-prey model. *Am. Nat.* **1979**, *113*, 306–308. [CrossRef]

142. Crandall, D.J.; Backstrom, L.; Cosley, D.; Suri, S.; Huttenlocher, D.; Kleinberg, J. Inferring social ties from geographic coincidences. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 22436–22441. [CrossRef]

143. Wang, D.; Pedreschi, D.; Song, C.; Giannotti, F.; Barabasi, A.-L. Human mobility, social ties, and link prediction. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1100–1108.

144. Liu, S.; Wang, S.; Jayarajah, K.; Misra, A.; Krishnan, R. TODMIS: Mining communities from trajectories. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 2109–2118.

145. Zhou, S.; Shen, W.; Zeng, D.; Zhang, Z. Unusual event detection in crowded scenes by trajectory analysis. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 1300–1304.

146. Liu, S.; Liu, Y.; Ni, L.M.; Fan, J.; Li, M. Towards mobility-based clustering. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 919–928.

147. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194.

148. Pan, B.; Zheng, Y.; Wilkie, D.; Shahabi, C. Crowd sensing of traffic anomalies based on human mobility and social media. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5–8 November 2013; pp. 344–353.

149. Nadimi, E.S.; Jørgensen, R.N.; Blanes-Vidal, V.; Christensen, S. Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Comput. Electron. Agric.* **2012**, *82*, 44–54. [CrossRef]

150. De Groeve, J.; Van de Weghe, N.; Ranc, N.; Neutens, T.; Ometto, L.; Rota-Stabelli, O.; Cagnacci, F. Extracting spatio-temporal patterns in animal trajectories: An ecological application of sequence analysis methods. *Methods Ecol. Evol.* **2016**, *7*, 369–379. [CrossRef]

151. Amato, F.; Moscato, V.; Picariello, A.; Sperlí, G. Recommendation in social media networks. In Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), Laguna Hills, CA, USA, 17–21 April 2017; pp. 213–216.

152. Amato, F.; Moscato, V.; Picariello, A.; Sperlí, G. Kira: A system for knowledge-based access to multimedia art collections. In Proceedings of the 2017 IEEE 11th international conference on semantic computing (ICSC), San Diego, CA, USA, 30 January–1 February 2017; pp. 338–343.

153. Yuan, J.; Zheng, Y.; Zhang, C.; Xie, W.; Xie, X.; Sun, G.; Huang, Y. T-drive: Driving directions based on taxi trajectories. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 99–108.

154. Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. T-drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 220–232. [CrossRef]

155. Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. Driving with knowledge from the physical world. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 316–324.

156. Phithakkitnukoon, S.; Veloso, M.; Bento, C.; Biderman, A.; Ratti, C. Taxi-aware map: Identifying and predicting vacant taxis in the city. In *Proceedings of the European Conference on Ambient Intelligence*; Springer: Málaga, Spain, 2010; pp. 86–95.

157. Ge, Y.; Xiong, H.; Tuzhilin, A.; Xiao, K.; Gruteser, M.; Pazzani, M.J. An energy-efficient mobile recommender system. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: Washington, DC, USA, 2010; pp. 899–908.

158. Li, B.; Zhang, D.; Sun, L.; Chen, C.; Li, S.; Qi, G.; Yang, Q. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops, Seattle, WA, USA, 21–25 March 2011; pp. 63–68.

159. Zimmerman, J.; Tomasic, A.; Garrod, C.; Yoo, D.; Hiruncharoenvate, C.; Aziz, R.; Thiruvengadam, N.R.; Huang, Y.; Steinfeld, A. Field trial of tiramisu: Crowd-sourcing bus arrival times to spur co-design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 1677–1686.

160.　Bastani, F.; Huang, Y.; Xie, X.; Powell, J.W. A greener transportation mode: Flexible routes discovery from GPS trajectory data. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 405–408.

161.　Trasarti, R.; Pinelli, F.; Nanni, M.; Giannotti, F. Individual Mobility Profiles: Methods and Application on Vehicle Sharing. In *Proceedings of the 20th Italian Symposium on Advanced Database Systems*; Citeseer: Venezia, Italy, 2012; pp. 35–42.

162.　Ratti, C.; Sobolevsky, S.; Calabrese, F.; Andris, C.; Reades, J.; Martino, M.; Claxton, R.; Strogatz, S.H. Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE* **2010**, *5*, e14248. [CrossRef] [PubMed]

163.　Shahraki, N.; Cai, H.; Turkay, M.; Xu, M. Optimal locations of electric public charging stations using real world vehicle travel patterns. *Transp. Res. Part D Transp. Environ.* **2015**, *41*, 165–176. [CrossRef]

164.　Shreenath, V.M.; Meijer, S. Spatial big data for designing large scale infrastructure: A case-study of electrical road systems. In Proceedings of the 2016 IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies, Shanghai, China, 6–9 December 2016; pp. 143–148.

165.　Chawla, S.; Zheng, Y.; Hu, J. Inferring the root cause in road traffic anomalies. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 141–150.

166.　Liu, W.; Zheng, Y.; Chawla, S.; Yuan, J.; Xing, X. Discovering spatio-temporal causal interactions in traffic data streams. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1010–1018.

167.　De Lucca Siqueira, F.; Bogorny, V. Discovering chasing behavior in moving object trajectories. *Trans. GIS* **2011**, *15*, 667–688. [CrossRef]

168.　Maisonneuve, N.; Stevens, M.; Niessen, M.E.; Steels, L. NoiseTube: Measuring and mapping noise pollution with mobile phones. In *Proceedings of the 4th International ICSC Symposium*; Springer: Thessaloniki, Greece, 2009; pp. 215–228.

169.　Rana, R.; Chou, C.T.; Bulusu, N.; Kanhere, S.; Hu, W. Ear-Phone: A context-aware noise mapping using smart phones. *Pervasive Mob. Comput.* **2015**, *17*, 1–22. [CrossRef]

170.　Shang, J.; Zheng, Y.; Tong, W.; Chang, E.; Yu, Y. Inferring gas consumption and pollution emission of vehicles throughout a city. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 1027–1036.

171.　Momtazpour, M.; Butler, P.; Hossain, M.S.; Bozchalui, M.C.; Ramakrishnan, N.; Sharma, R. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 126–133.

172.　Bao, J.; Zheng, Y.; Mokbel, M.F. Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 6–9 November 2012; pp. 199–208.

173.　Yoon, H.; Zheng, Y.; Xie, X.; Woo, W. Social itinerary recommendation from user-generated digital trails. *Pers. Ubiquitous Comput.* **2012**, *16*, 469–484. [CrossRef]

174.　Zheng, V.W.; Cao, B.; Zheng, Y.; Xie, X.; Yang, Q. Collaborative Filtering Meets Mobile Recommendation: A User-Centered Approach. In Proceedings of the 24th AAAI Conference on Artificial Intelligence, Atlanta, GA, USA, 11–15 July 2010; pp. 236–241.

175.　Yuan, N.J.; Zhang, F.; Lian, D.; Zheng, K.; Yu, S.; Xie, X. We know how you live: Exploring the spectrum of urban lifestyles. In Proceedings of the 1st ACM Conference on Online Social Networks, Boston, MA, USA, 7–8 August 2013; pp. 3–14.

176.　Filho, R.M.; Borges, G.R.; Almeida, J.M.; Pappa, G.L. Inferring user social class in online social networks. In Proceedings of the 8th Workshop on Social Network Mining and Analysis, New York, NY, USA, 24–27 August 2014; pp. 1–5.

177.　Karamshuk, D.; Noulas, A.; Scellato, S.; Nicosia, V.; Mascolo, C. Geo-spotting: Mining online location-based services for optimal retail store placement. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013; pp. 793–801.

178.　Körner, C.; Hecker, D.; May, M.; Wrobel, S. Visit potential: A common vocabulary for the analysis of entity-location interactions in mobility applications. In *Geospatial Thinking*; Painho, M., Santos, M., Pundt, H., Eds.; Springer: Berlin, Germany, 2010; pp. 79–95.

179. Gil, J.; Tobari, E.; Lemlij, M.; Rose, A.; Penn, A.R. The differentiating behaviour of shoppers: Clustering of individual movement traces in a supermarket. In *Proceedings of the 7th International Space Syntax Symposium*; Royal Institute of Technology (KTH): Stockholm, Sweden, 2009.

180. Da Silva, F.P.; Fileto, R. A method to detect and classify inconsistencies of moving objects' stops with requested and reported tasks. *J. Inf. Data Manag.* **2015**, *6*, 71–80.

181. Versichele, M.; Neutens, T.; Delafontaine, M.; Van de Weghe, N. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Appl. Geogr.* **2012**, *32*, 208–220. [CrossRef]

182. Dodge, S.; Laube, P.; Weibel, R. Movement similarity assessment using symbolic representation of trajectories. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1563–1588. [CrossRef]

183. Li, X.; Pan, G.; Wu, Z.; Qi, G.; Li, S.; Zhang, D.; Zhang, W.; Wang, Z. Prediction of urban human mobility using large-scale taxi traces and its applications. *Front. Comput. Sci.* **2012**, *6*, 111–121.

184. Tulusan, J.; Staake, T.; Fleisch, E. Providing eco-driving feedback to corporate car drivers: What impact does a smartphone application have on their fuel efficiency? In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 212–215.

185. Buchin, M.; Dodge, S.; Speckmann, B. Similarity of trajectories taking into account geographic context. *J. Spat. Inf. Sci.* **2014**, *2014*, 101–124. [CrossRef]

186. Mikut, R.; Reischl, M. Data mining tools. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2011**, *1*, 431–443.

187. Stančin, I.; Jović, A. An overview and comparison of free Python libraries for data mining and big data analysis. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 20–24 May 2019; pp. 977–982.

188. Torgo, L. *Data Mining with R: Learning with Case Studies*; CRC Press: Boca Raton, FL, USA, 2016.

189. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]

190. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME-the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 26–31. [CrossRef]

191. Hofmann, M.; Klinkenberg, R. *RapidMiner: Data Mining Use Cases and Business Analytics Applications*; Chapman & Hall/CRC Data Mining and Knowledge Discovery Series; CRC Press: Boca Raton, FL, USA, 2016; ISBN 9781498759861.

192. Wendler, T.; Gröttrup, S. *Data Mining with SPSS Modeler: Theory, Exercises and Solutions*; Springer International Publishing: Cham, Switzerland, 2016.

193. Tamayo, P.; Berger, C.; Campos, M.; Yarmus, J.; Milenova, B.; Mozes, A.; Taft, M.; Hornick, M.; Krishnan, R.; Thomas, S. Oracle Data Mining. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O.R.L., Ed.; Springer: Boston, MA, USA, 2005; pp. 1315–1329.

194. Fernandez, G. *Statistical Data Mining Using SAS Applications*; CRC Press: Boca Raton, FL, USA, 2010.

195. Abeler, J.; Bäcker, M.; Buermeyer, U.; Zillessen, H. COVID-19 contact tracing and data protection can go together. *JMIR mHealth uHealth* **2020**, *8*, e19359. [CrossRef]

196. Yasaka, T.M.; Lehrich, B.M.; Sahyouni, R. Peer-to-Peer contact tracing: Development of a privacy-preserving smartphone app. *JMIR mHealth uHealth* **2020**, *8*, e18936. [CrossRef]

197. Aggarwal, C.C.; Philip, S.Y. A survey of randomization methods for privacy-preserving data mining. In *Privacy-Preserving Data Mining: Advances in Database Systems*; Aggarwal, C.C., Yu, P.S., Eds.; Springer: Boston, MA, USA, 2008; pp. 137–156.

198. Gai, K.; Qiu, M.; Zhao, H.; Xiong, J. Privacy-Aware Adaptive Data Encryption Strategy of Big Data in Cloud Computing. In Proceedings of the 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), Beijing, China, 25–27 June 2016; pp. 273–278.

199. Wang, Y.; Zheng, Y.; Xue, Y. Travel time estimation of a path using sparse trajectories. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 25–34.

200. Zheng, Y.; Liu, T.; Wang, Y.; Zhu, Y.; Liu, Y.; Chang, E. Diagnosing New York city's noises with ubiquitous data. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 715–725.

201. Ding, X.; Chen, L.; Gao, Y.; Jensen, C.S.; Bao, H. Ultraman: A unified platform for big trajectory data management and analytics. *Proc. VLDB Endow.* **2018**, *11*, 787–799. [CrossRef]

202. Witkowski, K. Internet of things, big data, industry 4.0–innovative solutions in logistics and supply chains management. *Procedia Eng.* **2017**, *182*, 763–769. [CrossRef]

203. Roblek, V.; Meško, M.; Krapež, A. A complex view of industry 4.0. *Sage Open* **2016**, *6*. [CrossRef]

204. Tomičić Furjan, M.; Tomičić-Pupek, K.; Pihir, I. Understanding Digital Transformation Initiatives: Case Studies Analysis. *Bus. Syst. Res. Int. J. Soc. Adv. Innov. Res. Econ.* **2020**, *11*, 125–141.

205. Pejić Bach, M.; Bertoncel, T.; Meško, M.; Suša Vugec, D.; Ivančić, L. Big Data Usage in European Countries: Cluster Analysis Approach. *Data* **2020**, *5*, 25. [CrossRef]

206. Pejic-Bach, M.; Bertoncel, T.; Meško, M.; Krstić, Ž. Text mining of industry 4.0 job advertisements. *Int. J. Inf. Manag.* **2020**, *50*, 416–431. [CrossRef]

207. Fareri, S.; Fantoni, G.; Chiarello, F.; Coli, E.; Binda, A. Estimating Industry 4.0 impact on job profiles and skills using text mining. *Comput. Ind.* **2020**, *118*, 103222. [CrossRef]