Spatial Informatics

Prof. Soumya Kumar Ghosh
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

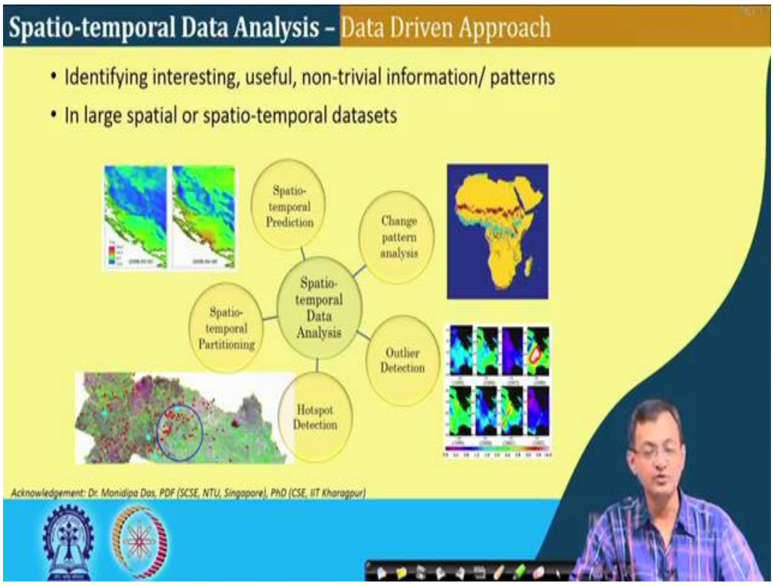
Lecture - 38
SDS / Spatial Cloud / GeoViz – III

Hello. So, let us continue our discussion on Spatial Data Science and other related topics as for today's lecture so on Spatial Informatics course, right. So, if you recollect last two lectures we discussed about different aspects of spatial data science. And in the specifically in the last lecture, we tried to look into some of the use cases or some of the major aspects of spatial data science and some use cases on that, right.

So, like we talked about spatial interpolation where or semantic interpolation, where kriging where we seen that there is a semantic kriging which was there in that which has helped in doing a better interpolation of the things, right. So, it is with a raster image where missing values are there and we do a semantic kriging by looking at the influence of a particular object or particular parameter.

On other parameter not only on the distance, but also the based on the type of parameter or type of things are there. Like what we say that influence with respect to temperature, influence of a water body will be different than influence of say vegetation or a influence of a build up area and type of things, right. So, they are semantic distances can be used to have this influences to calculate this that predicted value right or that value to be predicted.

(Refer Slide Time: 02:07)



Today we will look some other of the aspects that is spatiotemporal data analysis using a data driven approach. Like what we have a data driven a good amount of data and if you whether we can do spatiotemporal data and analysis using this data driven approach, right. So, this is again a work of one of our scholar who is presently doing our PDF, right so, scholar in our spatial data science lab.

So, if you look at the spatiotemporal data analysis there are several components; spatiotemporal prediction, change pattern analysis, outlier detection, hot spot detection, spatiotemporal partition, right. So, identify interesting useful non-trivial information and pattern is the overall goal in large spatial or spatiotemporal data sets, right. So, if you look at that the several component a spatiotemporal prediction, right.

So, if I have is set of spatiotemporal data whether I can predict the, this next occurrent you know the data, that may be climatological prediction, that may be some prediction of a particular event and type of things right. It can be like whether prediction to traffic prediction and type of thing this thing so, space time things.

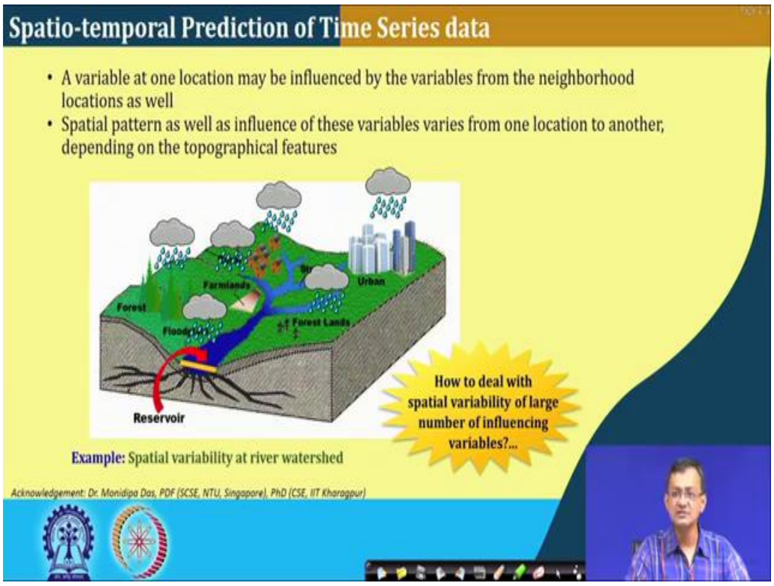
Change pattern analysis, change pattern analysis is important aspects for any large scale analysis of a particular region and type of things. Like if I see that large metros like Mumbai, Delhi, Calcutta, Kolkata, Chennai or like that, if we look at over the years how it is changing like how the harmonization changing. How this forest, how are that vegetation cover is changing and even I can have even if a way to measure this underground water etcetera, that also that how this layers are changing right even pollution layers are changing; so over the things how this change pattern analysis over space time can be looked into.

There are there can be several outliers; so, detecting out layer into a large dataset is a also a challenge. Hot spot detection typically when we talk about hot spots what we look at that the a particular region for a particular parameter is markley different from the things. Like if I say a temperature of a region of a particular X Y or in a particular region is markley higher or lower than the neighboring region right which is otherwise unlikely. So, we say that hotspot.

So, hotspots is a again a challenge or a important aspects for any urban planning. And there are issues of that experts in those areas they say there is a migration of the hotspot within their particular region and that needs to be looked into analyzed and remedial actions has to be taken.

A spatiotemporal partitioning based on the spatiotemporal things whether I can have some partitioning of the region and look at that how things works and type of things. So, there are several aspects of a spatiotemporal data analysis and we need to look into try to look into those things here.

(Refer Slide Time: 05:25)



So, this was the interesting work again done here is that finding that runoff of a particular watershed. So, when we talk of a typically a watershed I am excuse me I am not a hydrologist per se, but typically the basic philosophy of watershed is that a river along it is tributaries distributed is forming a non-compassing region called watershed of the things.

It is important to know that from the watershed how many what is the run off or the what are the how much water is coming out. Likes it is like I say, like if I say if this there are different streams and up streams and things like that. And if I have this is my the watershed right and then this is the what want to find out this runoff like here what we do finding other usually reservoirs will be there and it also dictates that what is the reservoir thing.

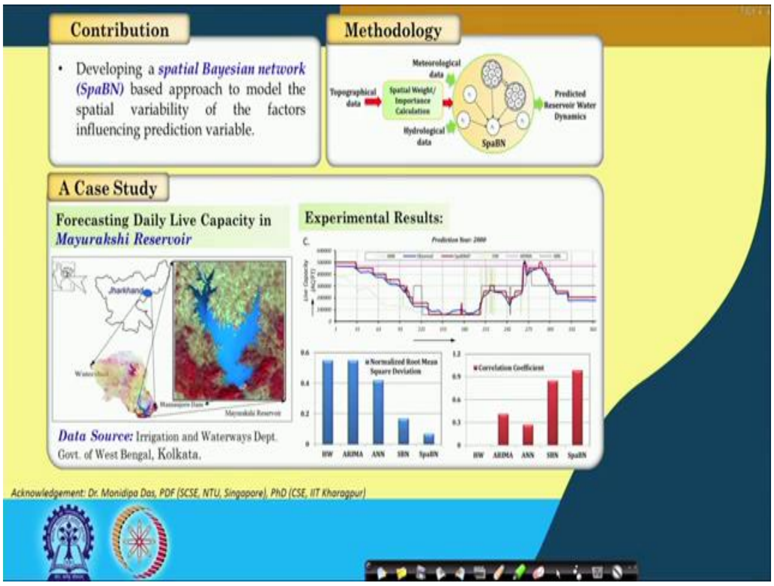
So, one way of handling those things is looking at the reservoir levels at over several years or several seasons and do a prediction right, that is also do that wet temporal prediction. Other, another along with that more thing is that, if this is my watershed then I can basically see that what are the rainfalls, what is the elevation, how much water is absorbed. Because the soil and etcetera some of the soils will absorb a lot of water some may not absorb so much water. So, that how much water is coming into the system and then which is contributing the things, right.

So, there are there can be different aspect to look at. So, a variable at one location may be influenced by the variable from the neighboring neighborhood location as well, right. So, the spatial pattern as well as the influences of this variable varies from one location to another, right. So, depending on the topographical feature right like I say that how much rainfall will be prone into the river is based on that how much water what is the elevation, how much water is being absorbed by the soil structure etcetera, right.

So, these are the different parameters which may give a domain by a domain expert. Whether as a computer science person or CSIT that other computer scientists whether we can able to model that into a generic model which dictates that how much based on that how much water will come out of the things. This is exactly the work what we try to do that and so, how to deal with that spatial variability of large number of influencing variables.

So, there can be large number of influencing variables right like temperature, humidity, rain fall, even sometimes these wind condition, soil absorption capability, elevation things, amount of vegetation cover and all those things influence they have different thing, right.

(Refer Slide Time: 09:00)

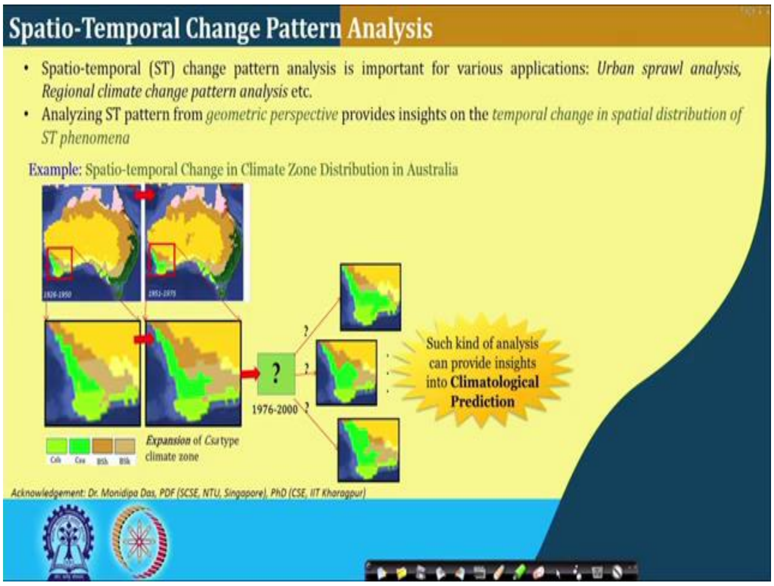


So, here developed a spatial Bayesian network based on approached network Bayesian; this spatial Bayesian network based a SpaBN based approach to model the spatial variability of the factors in placing the things, right. So, there are things if you see there are topographical data, spatial weight, importance calculation hydrological data, metrological data and then have a which this particular Bayesian network spatial, Bayesian network model named to predict the reservoir things.

And it has been shown that this has been a studied with in a nearby a, not nearby a Mayurakshi database or Mayurakshi reservoir where which is that particular Mayurakshi river watershed which data which received from of the data. We got it from irrigation department some also hill from remote sensing center and like that.

And we tried to look at these things and it has been seen that it this our methodology this proposed methodology able to predicted in a much better way. The idea is here see this spatial informatics is hill can be deployed or to find out this runoff of a particular watershed using different this techniques. Like in this case whatever view a spatial Bayesian network technique to use that, right.

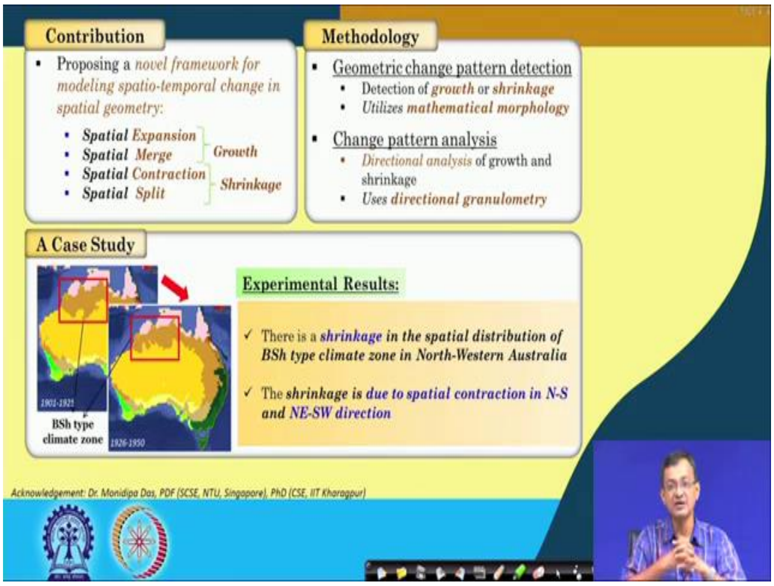
(Refer Slide Time: 10:46)



There is another interesting thing you can see that spatiotemporal change pattern analysis, right. So, it is spatiotemporal change pattern analysis is important for various application like we want to like as we are discussing urban sprawl analysis that how much urbanization is going on, regional climate change pattern analysis. Whether the there is a climate change analyzing spatiotemporal pattern from geometric perspective provides insights on the temporal change in the spatial distribution of the ST phenomena, right.

Spatiotemporal change in climatic zone distribution in Australia this is typically a region; we have taken up this region because the data set was available. This is something a dataset of 1926 to 1950 and 1950 to 75 and try to look at that how to predict that with some other 76 to 2000; what should be the prediction thing such kind of you can provide an insight into the climatological pattern, right. So, again this is a spatiotemporal dataset. So, these are again some dataset something which are remotely sensed data and which are basically segmented or classified into on finding out the different climatic zones, right.

(Refer Slide Time: 12:03)



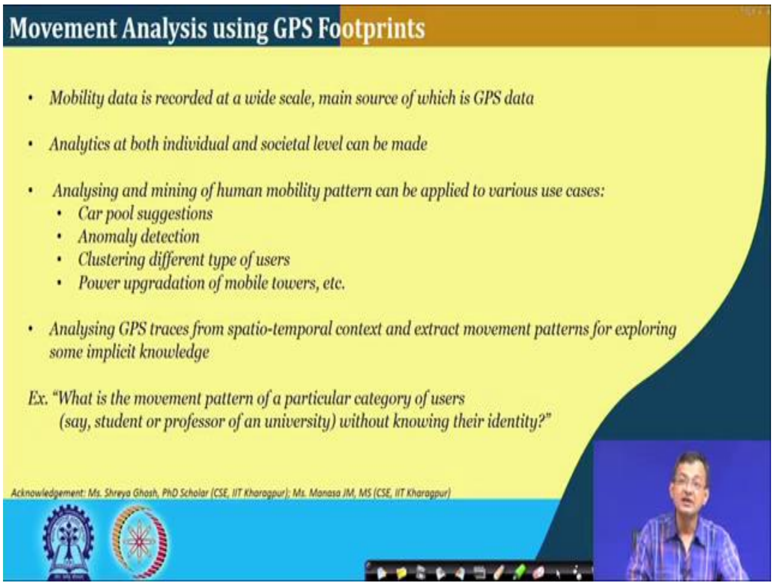
So, here what we have used that different framework for modeling spatiotemporal change in a spatial geometry. So, there are four geometry; one is spatial expansion and spatial merge which are growth right, other things are spatial contraction or spatial split which are shrinking right either split or contraction is shrinking and things.

So, what we did here geometric change pattern detection; detection of the growth or shrinking using mathematical morphology right. So, like did some directional analysis of growth and shrinkage use directional granulometry and for that purpose. Like it can be seen that is shrinkage in the spatial distribution of a particular type of climate BSh, climate zone in North Western Australia. The shrinkage is due to spatial contraction in North South and North East South West direction.

So, this is some of the insight which brings which may help in predicting the climatological pattern right. Again what we I try to emphasize here see again with the data we can do all this analysis, right. The analysis is done based on these data sets right. So, take a good amount of data which is space time variable make over the years like in this case we are having data from 1901 to a 1990. So, this volume of data some we can use for training and some use for testing purpose, but nevertheless we have a huge amount of data which are analyzed.

So, these are the data driven techniques which can be seen, which we can see that can be used for prediction type of purposes, which actually so, that how spatial informatics or spatial data science help in different supporting in different decision making or decision makers to work on it.

(Refer Slide Time: 14:22)



Another aspects we want to see that the movement analysis using GPS footprints, right. So, mobility of the data is recorded in wide scale main sources is a GPS data. These days all our smart phones maximum many most of the devices GPS is connected on vehicles, on buses and public transports. So, there is a huge volume of data is being collected which is space time variable right, I know the location and also that at what time it has been collected.

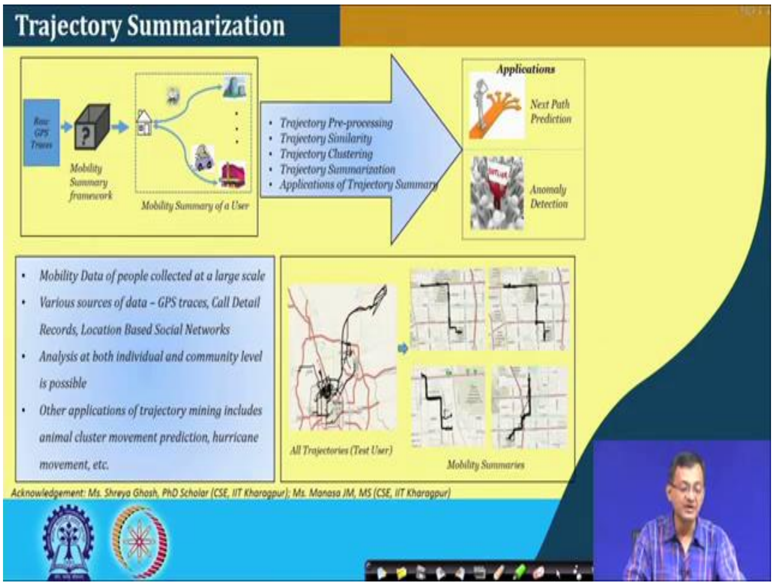
So, analytics at both individual and societal or a aggregate level is possible. Analyzing and mining of human mobility pattern may apply to the various use cases say car pool suggestion. How that location of a particular bus anomaly detection like it can say that there is a anomaly what are the reporting clustering different types of users, right. So, I can say this is the category of users which are this sort of cluster.

And even if we see that in some portion of the day that there are lot of accumulation of users and the mobile devices, whether there is a power upgradation or if it is not degradation or reduction in of the mobile towers etcetera, right; so, sort of a power management so to say instead of gain.

So, analyzing GPS traces from spatiotemporal context and extract movement patterns for exploring some implicit knowledge include. So, whatever this analyzing this spatiotemporal GPS traces to find out whether there is inherent pattern into the things, either in aggregate movement or individual movement based on the things.

Most of the studies what we treat is basically for aggregate movement it is not to look at individual things because that also have privacy issues, but aggregately whether we can say something, right. Like what is the movement pattern of a particular category of user without knowing their identity say student, or professor of an university on a particular time of that day. Like, what is the typical pattern that may help in arranging their traffic things; if I have a larger scale in a city it will be much helpful for the traffic management.

(Refer Slide Time: 16:48)

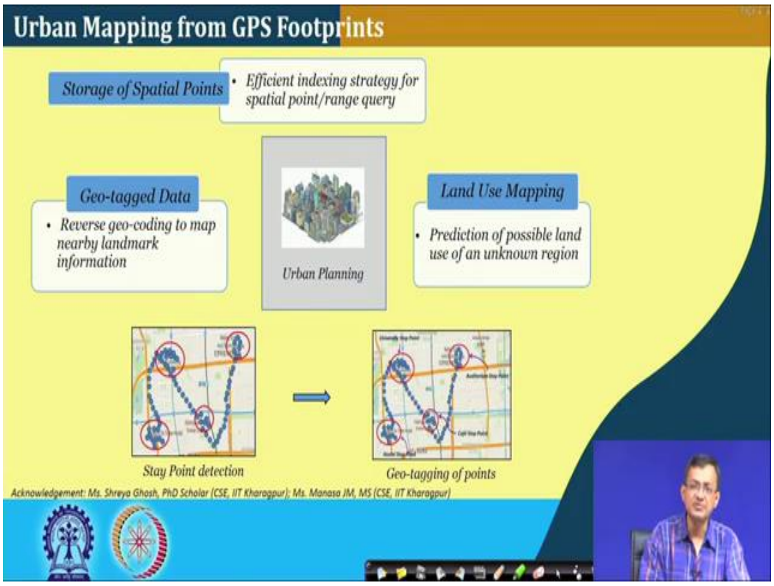


So, this trajectory summarization is important like mobility data for people collected a large scale, various sources of data GPS trace, even if the CDR is this call data records of a mobile users chopping of that user id or the mobile number. So, it is a call data record, location based social networks, if the social networks sharing with the location driven things, analysis at both individual and community level is possible, right. Other applications of trajectory mining includes animal cluster movement prediction, hurricane movement, etcetera.

So, that way we can basically try to predict that what are the different type of movement patterns of a individual user or group of users right. So, that we can suggest a proper travel path or also find out that overall traffic condition on a particular region on a particular day or date of name.

So, major application is next path prediction whether there is a anomaly or outlier detecting and type of things that it is a not a appropriate behavior. We can say that is neither a for say a academic campus neither a student not a faculty neither a staff, but it is likely to be outlier because this sort of things are not the matching with any movement patterns.

(Refer Slide Time: 18:17)

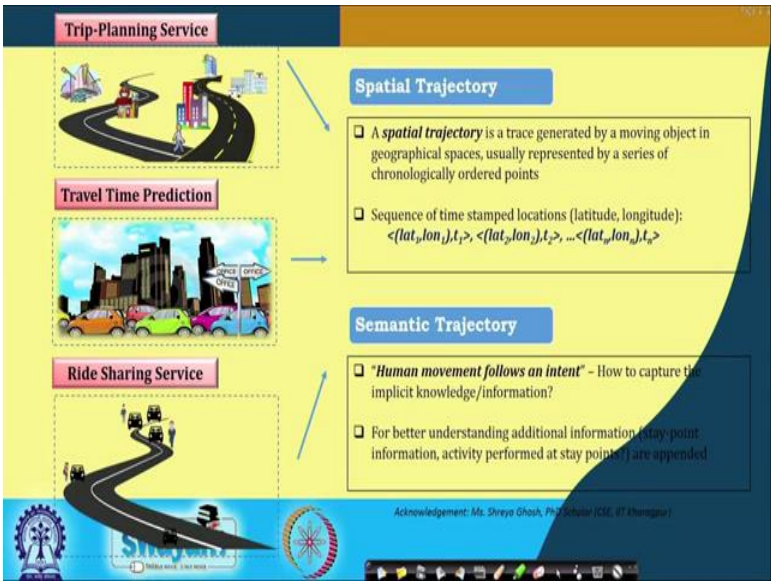


So, urban mapping from GPS footprint is another thing like storage of spatial point. If you see an indexing strategy for spatial point and query is important, because huge volume of data need to be accessed to appropriate indexing is that is one aspects. Geo tagged data reverse geo coding to map nearby landmark of the information finding out the nearby landmark.

Land use mapping predicting a possible land of unknown region right whether that is possible right I with looking at that particular things patterns we can do whether I can say this is a things. Like say for example, I have this point of view or point of interest of Kharagpur campus, IIT Kharagpur campus and the GPS traces of the this campus whether with only the GPS traces of say IIT Delhi campus. I can say this is a possible lecture hall, this is a possibly a cafeteria, this is a possibly a library, or hall of residence and etcetera.

In other sense we are whether I can a possible land use map or land cover land use map based on these GPS traces having some a priori knowledge of the same thing right. So, without actually surveying, I can say this type of pattern first of all this type of pattern looks like academic campus and this could be these are the p o is right. So, there is there are several possibilities and this may immensely help in planning upcoming city, extension of a thing like city, canals and type of things.

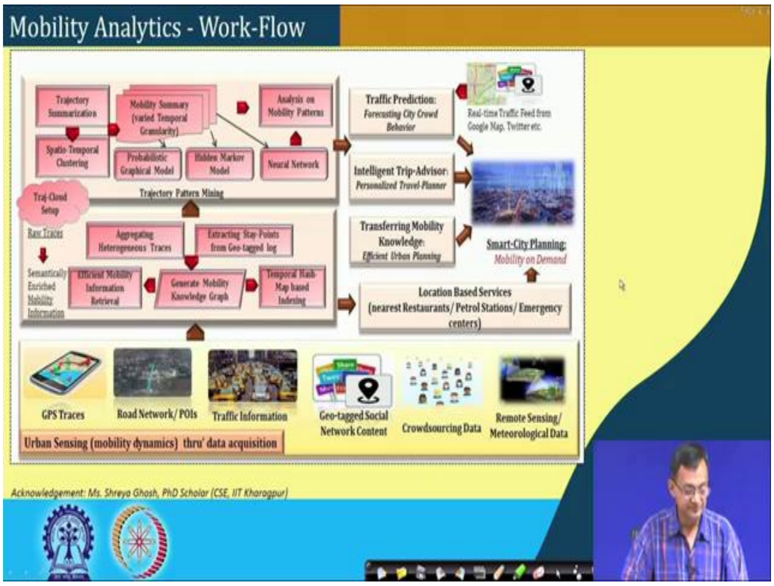
(Refer Slide Time: 20:02)



So, if we look at that different aspects of this GPS traces and so forth. So, trip planning services, travel time prediction, right sharing services, these are the some major things which are directly helpful, right. So, spatial trajectories if we define a spatial trajectory is a trace generate by moving objects in the geographical space usually represented by a series of chronological ordered points right, X Y lat long or X Y t or lat long t 1, lat long t 2 and so and so forth.

Semantic trajectory human movements follows an intent right. What we say that human moves with some intent right, how to capture that implicit knowledge or informations that how of that, right. So, for better understanding of additional information, stay point, information activity perform at stay are appended.

(Refer Slide Time: 21:05)



So, this is typically a mobility analytics work flow things what we are working out here by one of our research scholar in this spatial data science lab is working. She is working on this area. So, there can be different type of sources like urban sensing, mobility dynamics to data acquisition. So, there can be GPS traces, road network, point of interest, traffic information, geo tag, social network, crowd sourcing data, remote sensing, or even met data.

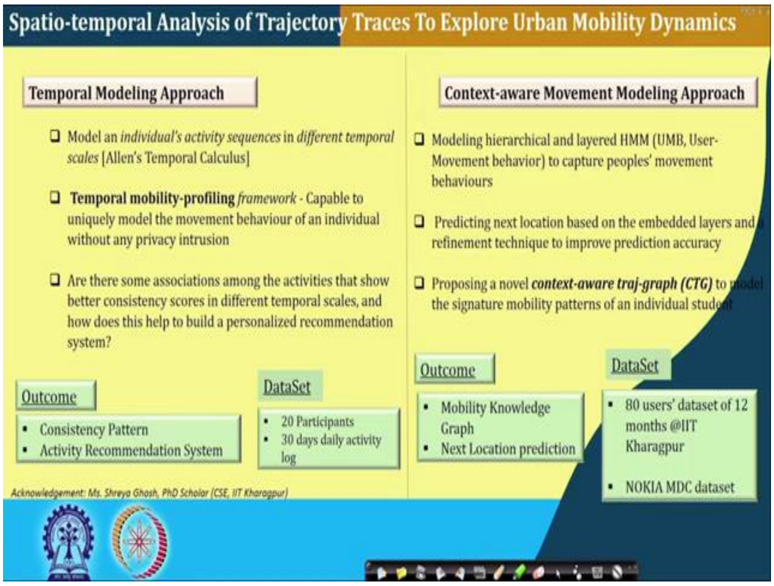
So, there are different sources of data which are used to do semantically enriched with the things with proper tagging etcetera, we enrich the thing. Based on that whether we are trying to get a trajectory pattern mining on the things, so find it a active mining which has different applications, right.

So, what we say different quote unquote smart city planning right, like traffic prediction, intelligent trip advisor right, personalized traffic planner, transferring mobility knowledge of and I can say, if this region is behaving like this. It is likely that after some time this will be there right or even if this is the mobility pattern of this sort of structure, I can say in a different city of a different region of the same city this is likely to be affected.

Even if I say I feel like this is semantically or sometimes from the knowledge over the things people know that like if there is congestion in the region A. There may be a similar type of congestion after some time in some region B which may not be directly adjacent region; it may be something other region. There can be different reason for that there is may be a lot of movement pattern on the things. So, this will be migrating to the things and etcetera, right.

So, there are transferring of mobility knowledge or transfer knowledge things is there, if you see an urban planning. So, look as; several location based service like nearest restaurant, petrol station, emergency centers etcetera are all being influenced out here.

(Refer Slide Time: 23:15)



So, what we see the spatiotemporal analysis for trajectory traces to explore urban mobility dynamics is one of the things. So, one is temporal modeling approach, model of a individual activity sequence in different temporal scale.

So, what is done through Allen's temporal calculus; so, temporal mobility profiling framework capable to uniquely model or movement behavior of an individual without any privacy intrusion. Like, there can be things like a particular say if you look at a particular individual. So, there are there may be a particular pattern, right.

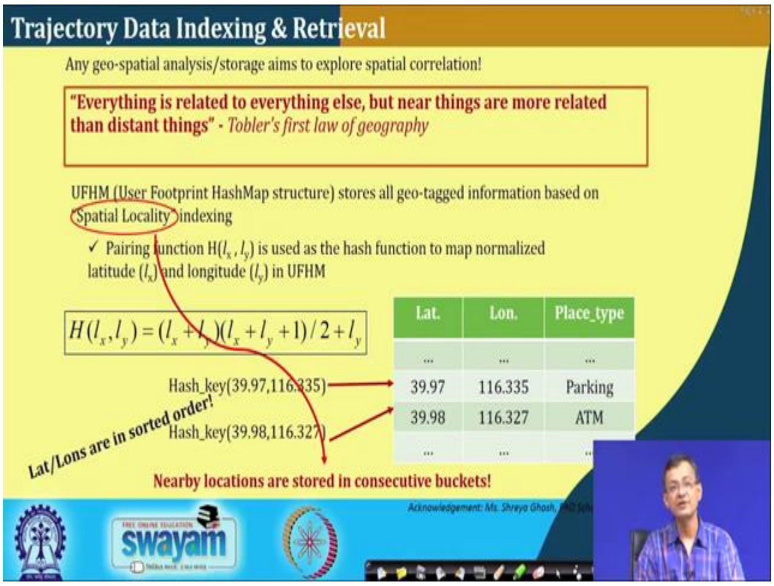
So, like he works for some time going to office then board a bus on that while going to the bars read a newspaper and there are activity, sub activity, sub sub activities right. So, by that which are in a temporal scale right over the time even looking not looking at the spatial thing, because spatial thing uniquely identified is location only the temporal profiling can be good that is it does not look at the privacy intrusion of the things. You do not know where it is, but at the same times it says that about how what are the activities.

So, are they are some association among the activities that. So, better consistency code in better temporal scale and how does this help to build a personalized recommendation system. So, if I have this time; this sort of things then I can have a personalized recommendation systems, right. So, like I can say like if I have a particular for some senior citizens or in old age homes. I say that this particular activity has to be like medicine taking before meal or after meal and type of things which can learn from it is activities. And basically recommend that whether he or she is in place A or place B, this is the temporal sequence.

There is another context ever movement modeling approach modeling hierarchical and layered HMM for user movement behavior to capture peoples movement, predicting next location based on a embedded layers and refinement techniques to prove prediction. And proposing a novel context aware trajectory graph to model the signature of the mobility patterns of individuals.

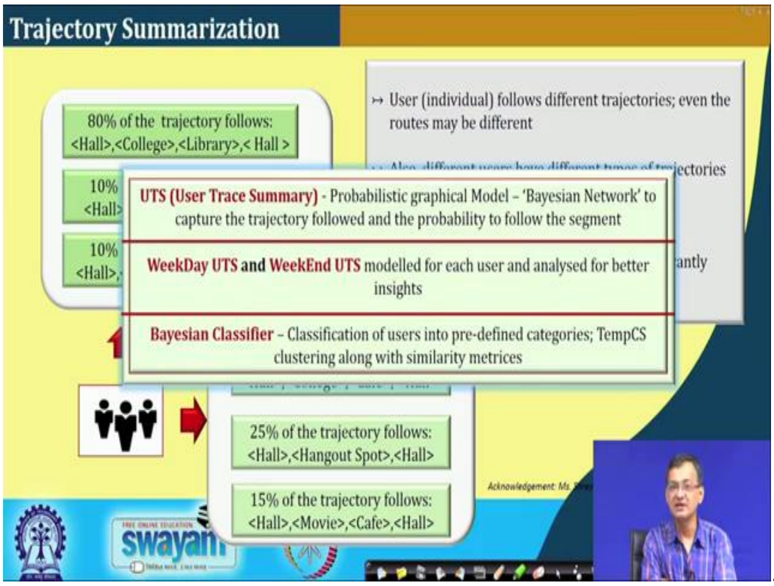
So, it is a context aware things right, like I say that the student movement during normal class hours a normal days are somewhere whereas, the student seems that student movement during some festive season like animal faced, or something will be something different right. So, it is or similarly people movement during some event and some other weekdays and weekends may be different, right. So, this is context aware.

(Refer Slide Time: 26:09)



And as we are discussing we are not going detail into these things like trajectory data indexing and the table is also an important factor that how could we index, because this there are huge volume of data need to be replayed retrieved at a much efficient way. So, we require some sort of a better as seeing and a better indexing techniques.

(Refer Slide Time: 26:33)

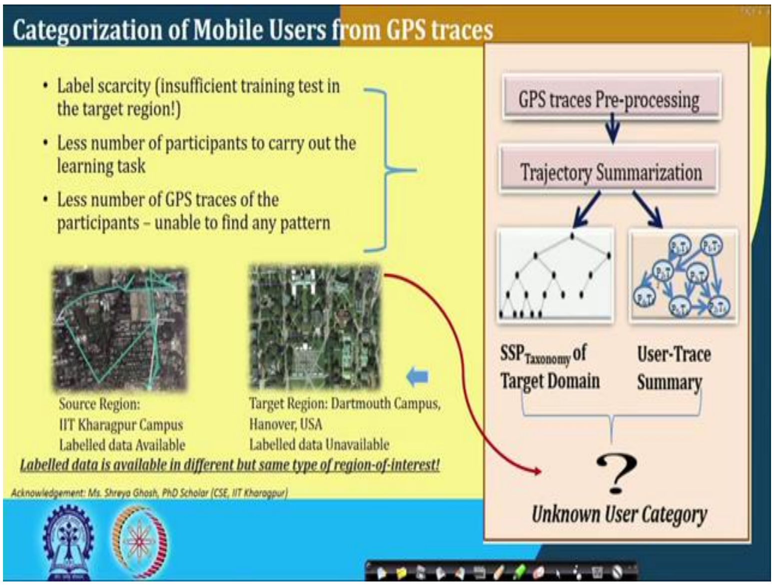


.

So, and as we discussed that this trajectory summarizes like 80 percent of the trajectory follows hall college, say this is a particular study what we carried out, 10 percent of the trajectory follow these etcetera. So, we can say user individual follow different trajectory even roots may be different. Also different user of different trajectories in our have different significant similarities weekdays and weekend patterns may be different, right.

So, user trace summary probabilistic graphical model. So, Bayesian network to capture trajectory followed and the probability to follow the segments, right. Weekdays UTS, weekend UTS model for each user analyzed for better insights, right. Bayesian classifier classification of user into predefined categories; TempCS clustering among with similarity matrix and etcetera are there.

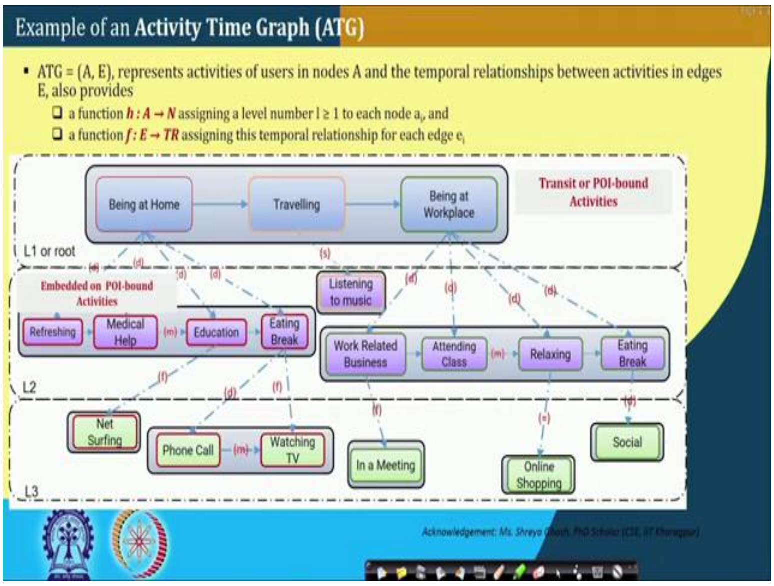
(Refer Slide Time: 27:32)



So, here also one interesting war on categorization of the user from the GPS traces like level scarcity, insufficient training test in the target region. So, using the GPS traces, pre-processing, trajectory summarization and then finding out that which category of the user cities, right.

So, it is sometimes there is insufficient training data can be there or less number of GPS traces of a participants unable to find any pattern whether this helps me in doing that, right. Like here we need one source region our IIT, Kharagpur campus, another is a Dartmouth campus in Hanover USA, labeled data was unavailable. But as both are academic campus we tried to map that this knowledge into the other knowledge level data is available in different. But the same type of region of interest both are academic regions and in this case.

(Refer Slide Time: 28:35)

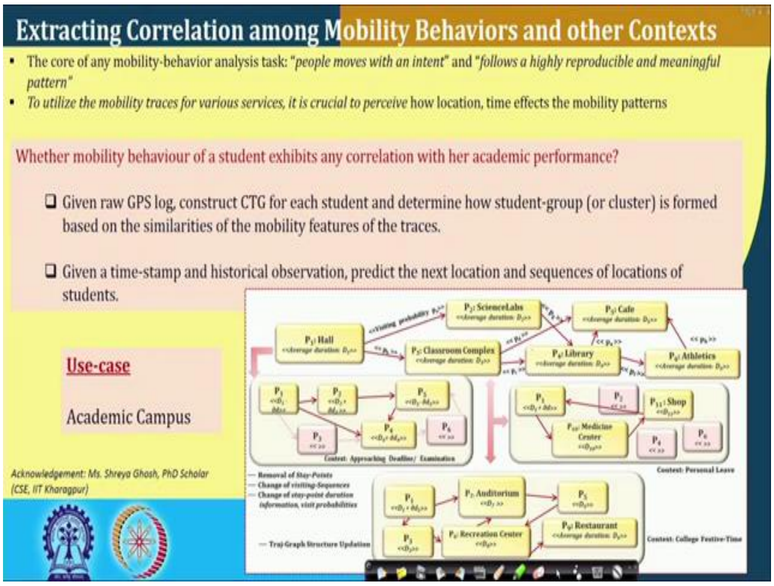


Similarly, as we discussed there is a activity time graph being at home, traveling and being at work place and type of things. So, that we have different category of things so being at home, refreshing, medical help education. So, there are different levels of activities right and while doing a say educating or education or reading something, then net surfing or eating while eating doing phone call watching TV and type of things.

So, what we see that this is a different levels of activities are there, activity with a sub activity, with a sub activities whether I can have this pattern into the things right. So transit or POI bound activities. So, these are either in transit or point of interest bound activities.

So, this helps me into finding out that activity time graph of a particular individual or a particular group based on the things. It may help in different planning or look or personalized recommendation and things like that, and similarly extracting correlation among mobility behaviors of other context, right.

(Refer Slide Time: 29:57)

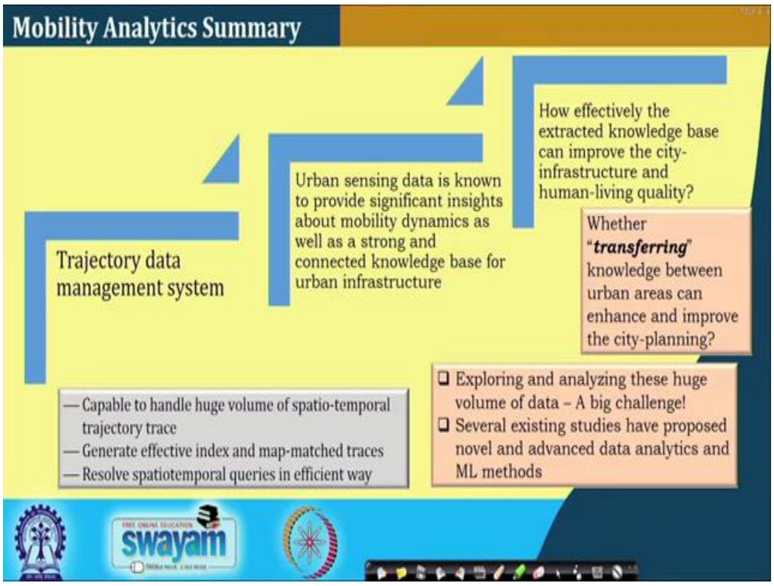


So, the core of any mobility behavior people moves with an indent flows a highly reproducible and meaningful patterns, right. So, what it does it moves with an intent and follows a highly reproducible or meaningful patterns. Like every day I come office in a particular route, a particular things with particular means, right. So, those whether I can map them, right.

So, like here in this case if you see there is some P 1 hall what is the visiting probabilities of science lab or classroom complex and so and so forth, right. So, this is context aware how I can have these based on the context. So, college festive time; some context will be there normal time some other context.

So, given a raw GPS log construct these CTG or the context have a trajectory graph of each student and determine how student group or cluster form on the similarities etcetera; so, given a time stamp historical observation of the location and location. So, we see that different type of prospective and things are like this.

(Refer Slide Time: 31:17)



So, not only that it helps us in looking at that whether this movement pattern can is also associated with any with their academics, right whether how, whether same group of or what group of people meet etcetera together etcetera. So, there is a lot of analysis of can be done along with the things.

So, there are from the trajectory data management to urban sensing data known for providing significant insights to effective extraction of knowledge base can be city infrastructure or human living, whether transferring knowledge between the urban areas can enhance improve city planning. So, there are several type of activities can be done with these spatiotemporal trajectory traces.

So, what we try to again to emphasize what we dare to discuss again today. What we discuss today is more that how what are the different applicability of this different spatio temporal analysis, spatial data science for different domains, right. Starting from climatological prediction or interpolation to this last one or last phase what we have discussed which is a big area of research across the world that what how to do trajectory analysis finding from the GPS traces, right.

So, we will continue our discussion on this line for 1 or 2 lectures to finding out that how this spatial data science help in different decision makings and decision support systems.

Thank you.