Mobile Phone Data Statistics as Proxy Indicator for Regional Economic Activity Assessment

Irina Arhipova[1][a], Gundars Berzins[1], Edgars Brekis[1], Juris Binde[2] and Martins Opmanis[3]
[1]*Faculty of Business, Management and Economics, University of Latvia, Aspazijas boulevard 5, Riga, Latvia*
[2]*Latvian Mobile Telephone, Ropazu street 6, Riga, Latvia*
[3]*The Institute of Mathematics and Computer Science, University of Latvia, Raina boulevard 29, Riga, Latvia*

Keywords: Efficiency, Principal Component Analysis, Theil Index.

Abstract: The mobile data analysis is an authoritative source of information for problems solving in the fields of human activity recognition, population dynamics, tourism, transport planning, traffics measuring, public administration and other activities and could be the source for valuable information as a proxy indicator. One of the obstacles to user data from mobile operators is compliance to the General Data Protection Regulation, so the development of data analytics approach that protects personal data without a necessity to identify mobility of particular persons was developed, that still provides economically relevant data. In the present research, the method for the economic activity assessment based on mobile phone data statistics of any analysed region has been developed. The data of person activity was aggregated at the area of each base station by the 15 minutes interval, where the activity is defined as the number of outgoing and incoming calls, sent and received short message service (SMS), as well the number of the unique users. The Latvian counties have used as a case study where regions were grouped into the similar categories and compared for two periods: 2015 – 2016 and 2017 by the economic activity efficiency with particular attention to the seasonality effect for mobile phone activities in counties. It was concluded, that the economic activity of counties can be estimated and in the particular case positive dynamics of regional development has been detected.

1 INTRODUCTION

Today's working life is intertwined with the use of mobile phones. They have long ceased to be just an information exchange tool. Unlike stationary phones that could be used to get list recipients and analyse frequency and duration of calls made, in the case of mobile phones, additional information can also be obtained about movements of the owner of the phone over time. Investigating particular mobile phone activities (the facts themselves, not their content) can provide insight into the mobility of the population and its economic activity. The obtained conclusions can be useful for the decision-making about regional development and could serve as metrics characterizing national economy. The mobile data analysis is an authoritative source of information for problems solving in the fields of human activity recognition, population dynamics, tourism, transport planning, traffics measuring, and public administration. The authors in previous research have analysed the mobile positioning data for residents' movements (Ahas et al., 2010; Zonghao et al., 2013), automatic recognition of population activities (Chetty et al., 2015; Lee and Cho, 2014), estimation of human trajectories (Hoteit et al., 2014; Larijani et al., 2015; Liu et al., 2013; Zilske and Nagel, 2015) and flows (Balzotti et al., 2018), patterns of population dynamics (Deville et al., 2014; Trasarti et al. 2015).

The identification of tourists' destinations (Alexander et al., 2015; Raun et al., 2016), seasonal patterns (Ahas et al., 2007; Phithakkitnukoon et al., 2015), traveler's preferences (Y. Wang et al., 2018) and behavior (Z. Wang et al., 2018), evaluation of tourism sector (Ahas et al., 2008; Kuusik et al.,

[a] https://orcid.org/0000-0003-1036-2024

Arhipova, I., Berzins, G., Brekis, E., Binde, J. and Opmanis, M.
Mobile Phone Data Statistics as Proxy Indicator for Regional Economic Activity Assessment.
DOI: 10.5220/0007772000270036
In *International Conference on Finance, Economics, Management and IT Business (FEMIB 2019)*, pages 27-36
ISBN: 978-989-758-370-4
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

2014), travel flow (Ni et al., 2018), tourist movement patterns (Zhao et al., 2018), trip modelling (Bwambale et al., 2017), analysis of number of travelers (Sørensen et al., 2018) and passengers demand (Hatziioannidu and Polydoropoulou, 2017) are the popular investigations problems.

Mobile phone data have used for transport planning (Elias et al., 2016; Liu et al., 2014), traffic measurement (Dong et al., 2015; Hongyan and Fasheng, 2013; Steenbruggen et al., 2016) and modelling (Oliveira et al., 2017), trajectories evaluation (Bonnel et al., 2015; Chen et al., 2014) and predicting travel time (Woodard et al., 2017). The general problems in urban planning (Ahas and Mark, 2005; Ricciato et al., 2017; Jonge et al., 2012) and analysis (Lee et al., 2018), land using (Ríos and Muñoz, 2017) and smart city development (Steenbruggen et al., 2015) are solved by mobile phone data analyzing.

As the basic unit of the mobile network infrastructure is a cell with its own base station, at the beginning of each call a specific base station determines the location of the mobile phone. If, during a call, the mobile phone moves beyond the limits of a particular cell, then switching to another base station takes place. If a person is sufficiently mobile, then a large number of base stations can be used to initialize calls. Although more than one base station may be used during one call, enough information can be obtained by analyzing only where the calls have started. If there are several base stations used for initializing calls during the day, then disposition of stations quite accurately reveals the habits of its owner. For example, if conversations in the mornings and evenings were initiated from one base station, but in mid-days from another, it may be assumed that these conversations geographically outline the person's working place and residence.

The data obtained can also be studied on the basis of information collected by the base stations on the calls initiated and/or short message service (SMS) sent. Changes in mobile activities during the day (week, month) can provide relevant information about the habits of the region's population. Within the framework of the study, it was intended to process data of the mobile telecommunications operator Latvian Mobile Telephone (LMT) to identify social factors influencing society and the national economy. The analysis of the mobile phone data statistics makes it possible to create different types of metrics of important processes in the society. In the original model the study of all mobile calls, anonymizing only the number was planned. Therefore, it still would be possible to trace activities of each particular anonymous person - to determine which calls were made using the same number (most likely by the same person), as well as observe the habits of moving this person.

The General Data Protection Regulation (GDPR), which is determined by the Law on Personal Data Processing and started on 25 May 2018, providing uniform rules for the protection of personal data throughout the European Union (EU). The Regulation will apply to any company, entity or organization that processes or stores data from identifiable individuals living in the EU (European Parliament, 2016). Taking into account the requirements of the GDPR within the framework of this study switching of particular individual calls are not fixed and tied to the corresponding activity. It is considered that it is reasonably sufficient to identify the base stations that ensure connection to the network. Where with another approach was used to do not require identification of mobility of a particular person - the data of person activity was aggregated at the area of each base station by the 15 minutes interval; where the activity may be outgoing (calls and sent SMS) as well as incoming (calls and received SMS). Also, the number of unique users for each interval is counted. The last one gives insight about average activity of persons in the region of a particular base station. Therefore the person data protection is respected. Undoubtedly, if compared to the original, a significant part of the data is lost. However, in the current version, data can still be used to:

- explore the mobility of people and their everyday habits (such as leaving home in the morning and returning to the evenings) as well as investigate serious demographic processes (relocation to cities, natural disasters, etc.);
- explore social interactions. Mobile phone usage models in different layers of society (gender, age groups, education, etc.) vary significantly. Knowing behaviour patterns, it is possible to find out the proportion of different groups in the region, as well as to observe changes in the habits (or the proportion of different groups);
- anticipate economic activity. Mobile service cost dynamics can quite accurately predict the upcoming crises.

Call Detail Record (CDR) is a digital data recorder used for telephone communications or other equipment for telecommunications, which includes telephone conversations or other telecommunications transactions (for example, text messages) that are broadcast on the device. CDR contains call time, duration, call status, caller and calling subscriber numbers. The CDR does not contain information about the content of the conversation. In previous research was concluded that mobile phone data are suitable and updatable for the Latvian regional business index development (Arhipova et al., 2017). The regions with similar economic activity patterns using mobile communication data were identified and it was concluded that mobile phone activities have statistically significant relationship to regional economic activity such as Gross Domestic Product (GDP), number of economically active enterprises per thousand inhabitants, municipalities budget expenditures and other (Arhipova et al., 2019). As a result, the hypothesis that counties and regions with lower call activity have lower economic activity compared to other regions with higher call activity could not be rejected.

The objective of this paper is to develop method for the economic activity assessment of any region being researched, based on mobile phone data statistics. Using previous research results, the first CDR data analysis was made for the time period from July 2015 to January 2018, with particular attention to the seasonality effect for mobile phone activities in counties. The second, the counties were grouped into the similar categories based on its economic activity and compared for two periods: 2015 – 2016 and 2017. The third, the counties' economic activity efficiency in and its change dynamics was estimated. The fourth, the region's economic activity was estimated by Theil index and conclusions were made about Latvia region economic development tendencies.

2 DATA AND METHODS

In addition to the connection attributes already mentioned, CDR generated automatically by the mobile network operator initializing mobile phone call or sending SMS contains information about the base stations that provides the connection and calling side. Since the coordinates of the base stations are known, the location of the persons at the beginning of the call can be determined with appropriate precision. Each database entry includes the following parameters: the total number of calls and SMS, the total number of unique users, date and 15 minutes time interval in the daytime, antenna identifier (ID) of the mobile network base station and its coordinates.

2.1 Data

The database used for the current case study consists of the roaming data of the mobile phone call activities of Latvia Mobile Telephone for 30 months from 25 July 2015 to 20 January 2018, altogether 108 008 160 CDR (Call Data Record) from 1235 base stations for 911 days with 15 minutes time intervals per hour.

The distribution of network base stations between Latvia counties and regions was obtained using its geographical coordinates. An administrative division includes 110 counties and 9 cities (Riga, Jekabpils, Jelgava, Jurmala, Ventspils, Liepaja, Daugavpils, Rezekne and Valmiera) in Latvia, but for statistical and planning purposes six statistical regions have been formed: Kurzeme, Latgale, Pieriga, Riga, Vidzeme and Zemgale. Data analysis shows the difference between the intensity of call activity on business days and on holidays (Saturday, Sunday or public holidays), which characterizes the economic activity of the area (Arhipova et al., 2017). At the same time, the maximum of the call activity is observed on noon for all days (Arhipova et al., 2019), as well the seasonality effect in summer and winter holidays.

In Figure 1 the total number of calls and SMS for three cities Jelgava, Jurmala and Ventspils is shown from August 2015 to December 2017, where in summer time Jurmala has the highest call activity, Jelgava has the lowest call activity, but in Ventspils mobile call activity doesn't have a strong seasonal effect. All cities have a seasonal effect in December, due to the winter holidays (Christmas and New Year).

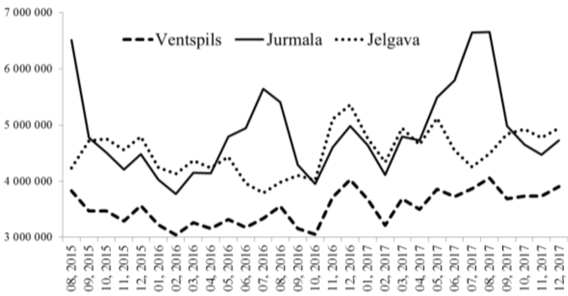

Figure 1: The total number of mobile phone activities in Jelgava, Jurmala and Ventspils.

To analyse the distribution of counties and cities by economic activity, Principal Component Analysis (PCA) was used to find out which counties are similar to the characteristics of the mobile phone call activities and group it into the similar categories or groups. In turn, the Theil index is proposed to distribute the regions according its economic activity. It is necessary to stress, that the Theil index measures the distribution of inequality not only within the same group, but also between different groups (Bellù and Liberati, 2006), for example, regional inequalities.

2.2 Grouping of Counties by Similar Economic Activity

Before grouping the counties into the similar categories based on its economic activity, the 119 variables as the linear combination of the total number of mobile phone activities and the total number of unique users for all counties were developed depending on day during the 2015 – 2017 time periods. The PCA was applied using Varimax rotation separately for two time periods:

- from 25 July 2015 to 20 January 2017 and
- from 21 January 2017 to 20 January 2018.

In the first time period 67.6% of the total variance is described by the first two principal components (PC), where the 1st PC has high values in business days as the counties with higher economic activity, but the 2nd PC has high values holidays as the counties with lower economic activity (Arhipova et al., 2019). In the next time period from 21 January 2017 to 20 January 2018 the results of applied PCA shows that 71.0% of the total variance is described by the first two principal components. To find out the interpretation of principal components, their average values were calculated based on weekdays (Fig.2) and months (Fig. 3).

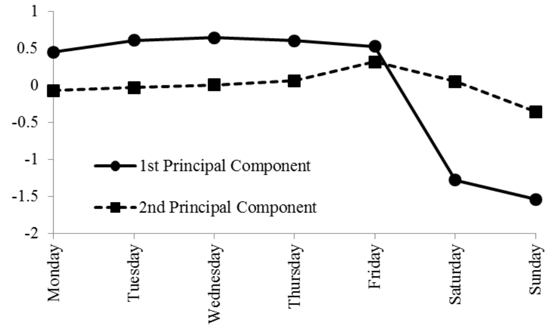

Figure 2: Principal components average values depending on weekdays in 2017.

It can be concluded that the 1st PC has the highest values on business days and lower values on holidays and summer months. In contrast, the 2nd PC has lower values on business days and higher values on holidays and summer months. The component loadings, which are the correlations of the observed 119 variables with the first two principal components, are used to interpret the meaning of components. It is hypothesized that counties and cities with higher economic activity correlate highly with the 1st PC, but with lower economic activity correlate highly with the 2nd PC. The distribution of the counties and cities by economic activity can be shown using the loading plot an orthogonal solution. Latvian counties and cities are grouped into 8 groups according to their economic activity. These groups allow you to understand the profile of each county and depend on the economic activity on business days and on holidays (Saturday, Sunday or public holidays), as well as seasonality effect.

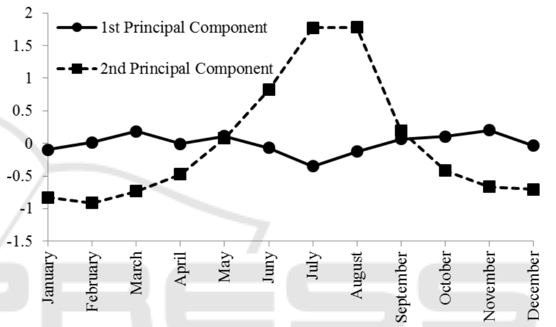

Figure 3: Principal components average values depending on months in 2017.

The summary of the proposed counties' groups, using the two components loadings or correlation coefficients, is shown in Table 1.

Table 1: Summary of counties' groups.

#	Group	Economic activity on business days (1st PC value)	Economic activity on holidays (2nd PC value)
1	Hard Workers	high (0.8 – 1.0)	average low (0.0 – 0.2)
2	Congruent	high (0.8 – 1.0)	average (0.2 – 0.6)
3	Moderate	average (0.4 – 0.8)	average (0.2 – 0.6)
4	Disinterested	low (0.0 – 0.4)	average (0.2 – 0.6)
5	Holidaymakers	average low (0.4 – 0.6)	average high (0.6 – 0.8)
6	Party Makers	low (0.0 – 0.4)	high (0.6 – 1.0)
7	Hedonists	lowest (-0.2 – 0.0)	highest (0.8 – 1.0)
8	Phenomenon	average (0.4 – 0.8)	average low (0.0 – 0.2)

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

In order to evaluate the effectiveness of the economic activity strategy chosen by each county, 40% to 100% efficiency curves were constructed. It is necessary to stress, that the amount of variance in each variable explained by the principal components or the component communalities are computed by taking the sum of the squared loadings for that variable, where component loadings are the correlations of the observed variables with the principal component.

Therefore the efficiency criterion *EC* was calculated according to the formula (1):

$$EC = \sqrt{r_{1n}^2 + r_{2n}^2} \quad (1)$$

where r_{1n}^2 is correlation coefficient of the observed n^{th} variable (linear combination of the total number of mobile phone activities and the total number of unique users in n^{th} county) with the first principal component loadings, but r_{2n}^2 is correlation coefficient of the observed n^{th} variable with the second principal component loadings.

2.3 Theil Index for Regions Economic Activity Evaluation

Theil index is a measure of overall inequality and primarily used to measure economic inequality. According the proposed hypothesis that mobile phone activities have statistically significant relationship to region economic activity, Theil index is calculated for the each county mobile phone activities data, aggregated by day for the time period from 25 July 2015 to 20 January 2017, using the formula (2):

$$E(1) = \frac{1}{n}\sum_{i}\left(\frac{y_i}{\bar{y}}\right) ln\left(\frac{y_i}{\bar{y}}\right) \quad (2)$$

where y_i is total number of mobile phone activities on a particular day, n is the number of days, and $\bar{y}$ is the average number of mobile phone activities for observed time period. The higher is the value of the Theil index, the greater is the data inequality.

Theil index is used to compare changes in mobile phone activities on business days and holidays. The higher the value of the Theil index, the higher the inequality between the number of mobile phone activities on business days and holidays.

It should be noted that Theil indexes are equal for the two possible cases: high mobile phone activities on business days and low on holidays or low mobile phone activities on business days and high on holidays.

Theil index was calculated for six statistical regions in Latvia: Kurzeme, Latgale, Pieriga, Riga, Vidzeme and Zemgale, using the number of mobile phone activities that are grouped by days for each statistical region.

3 RESULTS

Based on the PCA obtained results, Latvian 110 counties and 9 cities are grouped into 8 groups (Table 1) according to their mobile phone activity for two periods: 2015 – 2016 and 2017.

The efficiency of the economic activity strategy was calculated, using formula (1), and compared by 40% to 100% efficiency curves. The mobile phone activity of six regions was estimated by Theil index and the obtained results have compared with counties' mobile phone activity for the next Latvia region economic development tendencies evaluation.

3.1 Latvian Counties Distributions by Groups

Latvian 110 counties and 9 cities distribution by groups, according to their mobile phone activity for time period from 25 July, 2015 to 20 January, 2017 is shown on Figure 4.

The first groups "Hard Workers" is characterized by high activity on business days, but on average low activity on holidays. It is the driving force behind the Latvian economy, but does not fully exploit the holiday potential. The counties are highly dependent on fluctuations in economic activity, and it is necessary to develop the service sector. For example, the capital of Latvia the city of Riga is the central metropolis of the Baltic States, the international level port and infrastructure hub (The Freeport of Riga).

Riga is characterized by transit and logistics companies, developing social infrastructure. At the same time, Riga is a monocentric city, which insufficient uses tourism potential, only 5% of small and medium enterprises (SMEs). Riga has a congested traffic infrastructure and a population reduction. The city of Jelgava is only 40 km far from to Riga and has excellent infrastructure, besides there is situated the Latvia University of Life Sciences and Technologies. However, there is small business activity on holidays and insufficient leisure opportunity, only 4% of SMEs work in tourism. The second group "Congruent" of Latvia's counties is

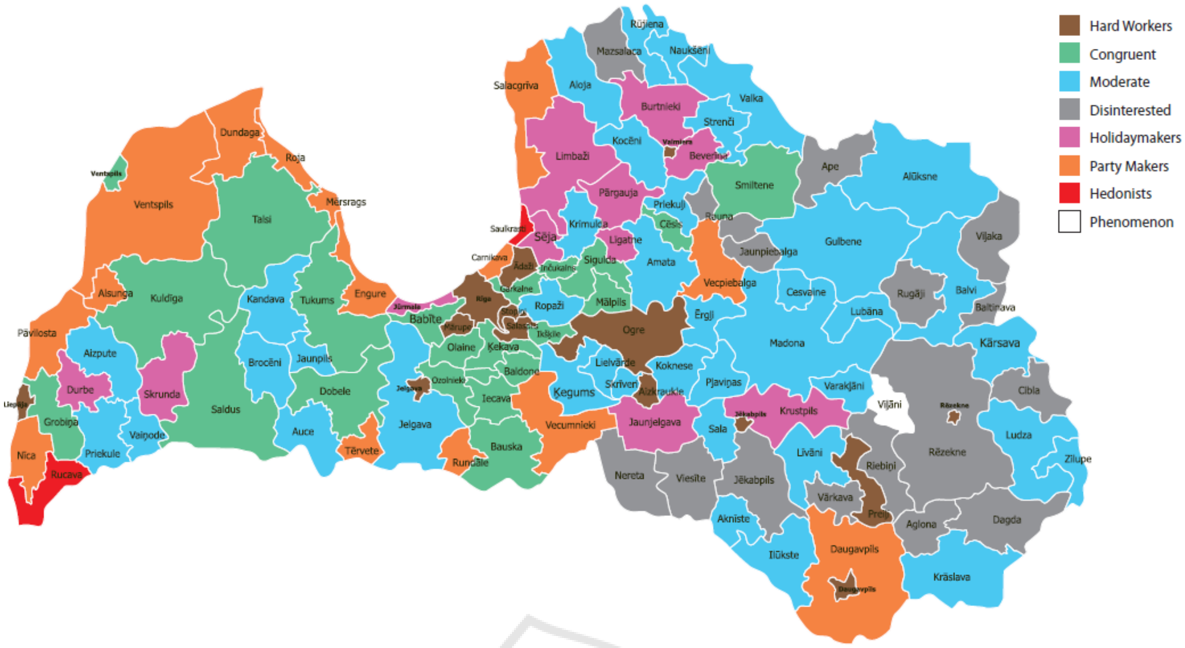

Figure 4: Latvian 110 counties and 9 cities distribution by groups in 2015 – 2016.

characterized by high and moderate activity on business days and average activity on holidays. The group is characterized by balanced development, but insufficient resources for the next breakthrough. Depending on the priorities, it is necessary to develop the production or service sector, but the wrongly selected priorities can be broken down by the available resources. For example, the city of Ventspils has a developed production and logistics sector, a well-developed port, Ventspils University College and in tourism sector there is 5% of SMEs.

The group "Holidaymakers" is characterized by an average low activity on business days, but on average high activity on holidays. The counties sufficiently use the holiday potential, but insufficiently on business days. It is necessary to develop the production sector and change the county's development strategy. For example, the city of Jurmala is a historic resort town near Riga with favourable strategic positioning, where 7% of SMEs are in the tourism sector. However, it does not fully use its potential on holidays.

In both cities of Riga and Jelgava from group "Hard Workers" exist high activity on business days, but on average low activity on holidays, as well a strong negative seasonal effect in summer time.

The city of Ventspils from the second group "Congruent" is characterized by high and moderate activity on business days and average activity on holidays and summer time. In turn the city of Jurmala from the group "Holidaymakers" has an average low activity on business days, but on average high activity on holidays, as well the strong positive seasonal effect in summer time.

The average number of mobile phone call activities of LMT per month in Riga, Jelgava, Ventspils, and Jurmala in 2017 is given in Figure 5.

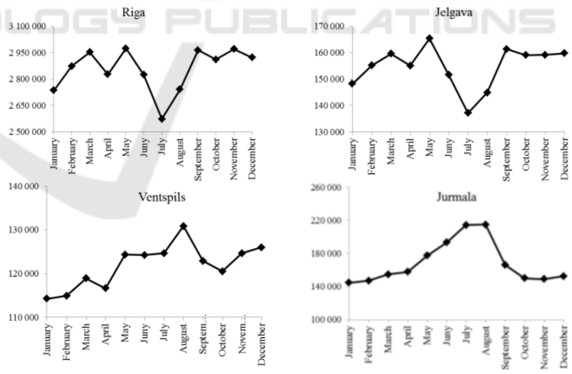

Figure 5: The average number of mobile phone call activities in cities per month.

The group "Moderate" is characterized by the average economic activity on all days. It is characterized by uniform activity, where the resource potential is not sufficiently used. It is necessary to increase labour productivity and economic potential, otherwise economic activity and regional development will decrease.

The group "Disinterested" is characterized low activity on business days and average activity on holidays. There is a holiday potential, but low economic activity on business days. It is necessary to develop the service sector and to change the development strategy of the region, threatening the region's degradation.

The group “Party Makers” is characterized by low activity on business days, but high activity on holidays. It is necessary to develop the production sector and change the region's development strategy. There is a high dependence on the purchasing power of the population.

The group “Phenomenon” is characterized by average activity on business days and moderate low activity on holidays, where in 2015 – 2016 it was Vilanu county, but in 2017 - Rugaju county.

The “Hedonists” group is characterized by the lowest activity on business days, but the highest activity on holidays. There is no economic potential for the manufacturing sector. It is necessary to develop the production sector and change the region's development strategy. There is a maximum dependence on the purchasing power of the population. For example, Rucavas county has a strong positive seasonal effect in summer months and it is unique because the number of mobile phone activities on holidays is higher than in business days (Fig. 6).

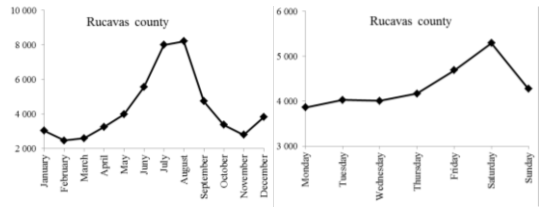

Figure 6: The average number of mobile phone call activities in Rucavas county per month and week days.

3.2 Latvian Counties Economic Activity Efficiency

The distribution of the counties by economic activity can be shown using the principal components loads with data aggregated by days. Latvian counties are divided into 8 groups according to their economic activity.

The negative 2nd PC values in group “Hedonists” means, that the mobile phone activity during the holiday in absolute values is higher, that in business days. It is a character only for counties from group “Hedonists”: Saulkrastu county in 2015 – 2016 and Rucavas county in 2015 – 2016 and 2017 time periods.

Groups allow you to understand the profile of each county, but the efficiency curve makes it possible to assess the effectiveness of the strategy chosen by each county (Fig. 7).

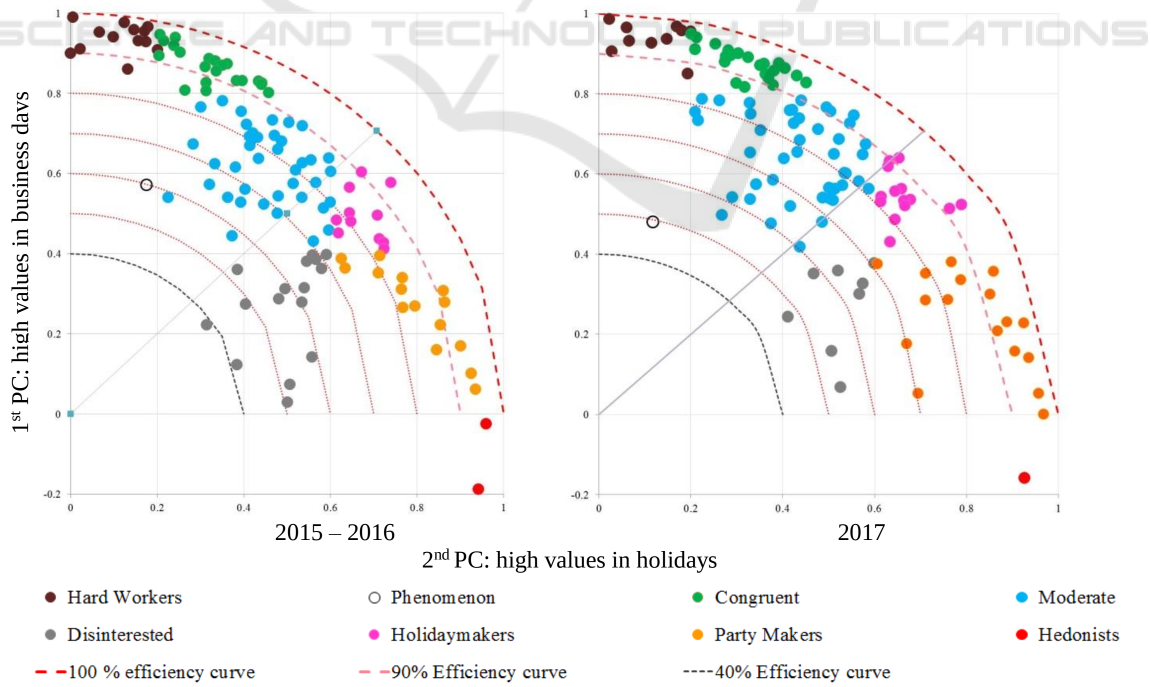

Figure 7: Distribution of Latvian counties in groups after the effectiveness of economic activity.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

In 2017, in comparison with 2015 - 2016, the economic activity of Latvian counties is improving, as indicated by the distribution of groups, for example, the number of counties included in the group “Disinterested” decreased two times.

The distribution of 119 Latvian counties in 2017 according to the efficiency criterion *EC* displays that the number of the counties lies between the two different efficiency curves are the following:

- 16 counties from 95% to 100%;
- 31 counties from 90% to 95%;
- 36 counties from 80% to 90%;
- 36 counties from 45% to 80%.

In comparison in 2015 – 2016 time period there were two counties with efficiency less than 45%.

3.3 Distribution of Regions by Economic Activity using Theil Index

In order to evaluate the regional differences in call activity between business days and holidays, Theil index *E(*1) for each region was calculated for period from 1 August 2015 to 31 December 2016.

The highest index value indicates a greater difference in call activity between business days and holidays, which indicates a higher economic activity. According to the Theil index the higher economic activity is in Riga, than Kurzeme and Pieriga, but Latgale and Vidzeme have the lowest results in economic activity (Fig. 8).

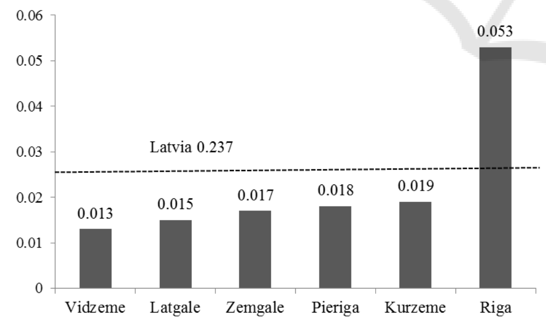

Figure 8: Theil index for Latvian regions in 2015-2016.

The similar results were obtained, using the PCA, which shows that cities and regions with lower call activity have lower economic activity compared to other regions with higher call activity (Arhipova et al., 2019).

To estimate the dynamics of the difference between regional call activities in business days and holidays, the Theil index for each region is calculated per month. There is a close correlation between regions’ Theil index change dynamics by months, except city of Riga, and indicates the related tendency of regions economic development.

4 CONCLUSIONS

The number of mobile phone call activities is an indicator of economic activity for counties and regions and could be used to develop a reliable tool for continuous and dynamic monitoring of the region’s economic performance.

There is a seasonal effect in mobile phone activities, as well a significant difference between business days and holidays, that indicates and distinguishes counties and regions with different economic activity patterns.

Data obtained from Latvian counties resulted in the identification of distinct 8 groups representing the unique pattern of economic activity. The efficiency curve makes it possible to evaluate the effectiveness of the strategy chosen by each county.

Characteristics of the distribution of Latvian regional groups depend on economic activity on business days and holidays, as well on seasonal effect.

In 2017, in comparison with 2015 - 2016, the economic activity of Latvian counties is improving, as indicated by the distribution in groups, for example, the number of counties included in the group "Disinterested" decreased by two times.

The Theil index characterizes the difference in mobile phone activities between business days and holidays. The results of PCA and Theil Index give the same distribution of counties and regions by its economic activity.

The authors propose the regional development index as a real-time or periodic monitoring tool. Developed method provides a practical tool for regional governments in keeping track on strategy implementation and strategic gap analysis.

This method also provides dynamic visualization of strategic direction particular municipality have achieved between periods of measurement and can be used as central additional performance indicator regional governments measure regularly.

The developed method tested on Latvian mobile telephone data sets as proxies, and the regional development index was created.

Regional economic activity efficiency evaluation is based on mobile phone data statistics, taking into account the requirements of the GDPR, where switching of particular individual calls are not fixed, and the personal data protection is respected.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

ACKNOWLEDGEMENTS

This work was supported by the University of Latvia and LMT Ltd. [grant number AAP2016/B089].

REFERENCES

Ahas, R., Aasa, A., Silm, S., Tiru, M., 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C*, 18(1), pp.45–54.

Ahas, R., Aasa, A., Roose, A., Mark, U., Silm, S., 2008. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29, pp.469–486.

Ahas, R., Aasa, A., Mark, U., Pae, T., Kull, A., 2007. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 28, pp.898–910.

Ahas, R., Mark, U., 2005. Location based services — new challenges for planning and public administration? *Futures*, 37, pp.547–561.

Alexander, L., Jiang, S., Murga, M., Gonzalez, M. C., 2015. Origin – destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C*, 58, pp.240–250.

Arhipova, I., Berzins, G., Brekis, E., Opmanis, M., Binde, J., Steinbuka, I., Kravcova, J., 2019. Pattern Identification by Factor Analysis for Regions with Similar Economic Activity Based on Mobile Communication Data. *Advances in Intelligent Systems and Computing*, 886, pp.561–569.

Arhipova, I., Berzins, G. Brekis, E., Kravcova, J., Binde, J., 2017. The methodology of region economic development evaluation using mobile positioning data. In D. Cingula, ed. *20th International Scientific Conference on Economic and Social Development*. Prague: Varazdin Development and Entrepreneurship Agency, Varazdin, Croatia. City of Prague. University North, Koprivnica, Croatia. Faculty of Management University of Warsaw, Warsaw, Poland, pp. 111–120.

Balzotti, C., Bragagnini, A., Briani, M., Cristiani, E., 2018. Understanding Human Mobility Flows from Aggregated Mobile Phone Data. *IFAC-PapersOnLine*, 51(9), pp.25–30.

Bellù, L.G., Liberati, P., 2006. *Describing Income Inequality. Theil Index and Entropy Class Indexes*, Available at: www.fao.org/tc/easypol.

Bonnel, P., Hombourger, E., Olteanu-Raimond, A.-M., Smoreda, Z., 2015. Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations. *Transportation Research Procedia*, 11, pp.381–398.

Bwambale, A., Choudhury, C.F., Hess, S., 2017. Modelling trip generation using mobile phone data: A latent demographics approach. *Journal of Transport Geography*, (August), p.article in press.

Chen, C., Bian, L., Ma, J., 2014. From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C*, 46, pp.326–337.

Chetty, G., White, M., Akther, F., 2015. Smart Phone Based Data Mining For Human Activity Recognition. *Procedia Computer Science*, 46, pp.1181–1187.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J., 2014. Dynamic population mapping using mobile phone data. *PNAS (Proceedings of the National Academy of Sciences of the United States of America)*, 111(45), pp.15888–15893.

Dong, H., Wu, M., Ding, X., Chu, L., Jia, L., Qin, Y., Zhou, X., 2015. Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C*, 58, pp.278–291.

Elias, D., Nadler, F., Stehno, J., Krosche, J., Lindorfer, M., 2016. SOMOBIL – improving public transport planning through mobile phone data analysis. *Transportation Research Procedia*, 14, pp.4478–4485.

European Parliament, 2016. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da. *Official Journal of the European Union*, L 119, pp.1–88.

Hatziioannidu, F., Polydoropoulou, A., 2017. Passenger Demand And Patterns Of Tourists' Mobility In The Aegean Archipelago With Combined Use Of Big Datasets from Mobile Phones and Statistical Data from Ports and Airports. *Transportation Research Procedia*, 25, pp.2309–2329.

Hongyan, G., Fasheng, L., 2013. Estimating freeway traffic measures from mobile phone location data. *European Journal of Operational Research*, 229(1), pp.252–260.

Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., Pujolle, G., 2014. Estimating human trajectories and hotspots through mobile phone data. *Computer Networks*, 64, pp.296–307.

Jonge, E. de, Pelt, M. van, Roos, M., 2012. *Time patterns, geospatial clustering and mobility statistics based on mobile phone network data*. The Hague: Statistics Netherlands.

Kuusik, A., Nilbe, K., Mehine, T., Ahas, R., 2014. Country as a free sample: the ability of tourism events to generate repeat visits. Case study with mobile positioning data in Estonia. *Procedia - Social and Behavioral Sciences*, 148, pp.262–270.

Larijani, A.N., Olteanu-Raimond, A.-M., Perret, J., Bredif, M., Ziemlicki, C., 2015. Investigating the mobile phone data to estimate the origin destination flow and analysis; case study: Paris region. *Transportation Research Procedia*, 6, pp.64–78.

Lee, K.-S., You, S.Y., Eom, J.K., Song, J., Min, J.H., 2018. Urban spatiotemporal analysis using mobile phone data: Case study of medium- and large-sized

Korean cities. *Habitat International*, 73 (October 2017), pp.6–15.

Lee, Y.-S., Cho, S.-B., 2014. Activity recognition with android phone using mixture-of-experts co-trained with labeled and unlabeled data. *Neurocomputing*, 126, pp.106–115.

Liu, F., Janssens, D., Wets, G., Cools, M., 2013. Annotating mobile phone location data with activity purposes using machine learning algorithms. *Expert Systems With Applications*, 40(8), pp.3299–3311.

Liu, F., Janssens, D., Cui, J.X., Wang, Y.P., Wets, G., Cools, M., 2014. Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications*, 41(14), pp.6174–6189.

Ni, L., Wang, X.C., Chen, X.M., 2018. A spatial econometric model for travel fl ow analysis and real-world applications with massive mobile phone data. *Transportation Research Part C*, 86(April 2017), pp.510–526.

Oliveira, E.M.R., Viana, A.C., Naveen, K.P., Sarraute, C., 2017. Mobile data traffic modeling: Revealing temporal facets. *Computer Networks*, 112, pp.176–193.

Phithakkitnukoon, S., Horanont, T., Witayangkurn, A., Siri, R., Sekimoto, Y., Shibasaki, R., 2015. Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan. *Pervasive and Mobile Computing*, 18, pp.18–39.

Raun, J., Ahas, R., Tiru, M., 2016. Measuring tourism destinations using mobile tracking data. *Tourism Management*, 57, pp.202–212.

Ricciato, F., Widhalm, P., Pantisano, F., Craglia, M., 2017. Beyond the "single-operator, CDR-only" paradigm: An interoperable framework for mobile phone network data analyses and population density estimation. *Pervasive and Mobile Computing*, 35, pp.65–82.

Ríos, S.A., Muñoz, R., 2017. Land Use detection with cell phone data using topic models: Case Santiago, Chile. *Computers, Environment and Urban Systems*, 61, pp.39–48.

Sørensen, A.Ø., Bjelland, J., Bull-Berg, H., Landmark, A.D., Akhtar, M.M., Olsson, N.O.E., 2018. Use of mobile phone data for analysis of number of train travellers. *Journal of Rail Transport Planning & Management*, 8(2), pp.123–144.

Steenbruggen, J., Tranos, E., Nijkamp, P., 2015. Data from mobile phone operators : A tool for smarter cities ? *Telecommunications Policy*, 39(3-4), pp.335–346.

Steenbruggen, J., Tranos, E., Rietveld, P., 2016. Traffic incidents in motorways : An empirical proposal for incident detection using data from mobile phone operators. *Journal of Transport Geography*, 54, pp.81–90.

Trasarti, R., Olteanu-Raimond, A.-M., Nanni, M., Couronne, T., Furletti, B., Giannotti, F., Smoreda, Z., Ziemlicki, C., 2015. Discovering urban and country dynamics from mobile phone data with spatial correlation patterns. *Telecommunications Policy*, 39(3-4), pp.347–362.

Wang, Y., Correia, G.H. de A., Arem, van B., Timmermans, H.J.P., 2018. Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. *Transportation Research Part C*, 90(October 2017), pp.247–259.

Wang, Z., He, S.Y., Leung, Y., 2018. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, 11, pp.141–155.

Woodard, D., Nogina, G., Koch, P., Racz, D., Goldszmidt, M., Horvitz, E., 2017. Predicting travel time reliability using mobile phone GPS data. *Transportation Research Part C*, 75, pp.30–44.

Zhao, X., Lu, X., Liu, Y., Lin, J., An, J., 2018. Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an , China. *Tourism Management*, 69(February), pp.368–383.

Zilske, M., Nagel, K., 2015. A simulation-based approach for constructing all-day travel chains from mobile phone data. *Procedia Computer Science*, 52, pp.468–475.

Zonghao, R., Dongyuan, Y., Zhengyu, D., 2013. Resident Mobility Analysis Based on Mobile-Phone Billing Data. *Procedia - Social and Behavioral Sciences*, 96(Cictp), pp.2032–2041.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS