УДК 311.22:[316.32:004.73]

О. С. Корепанов,
к. е. н., доцент, доцент кафедри статистики, обліку та аудиту,
Харківський національний університет імені В. Н. Каразіна, м. Харків

ПЕРСПЕКТИВИ ВИКОРИСТАННЯ "ВЕЛИКИХ" ДАНИХ ОПЕРАТОРІВ МОБІЛЬНОГО ЗВ'ЯЗКУ В ІНФОРМАЦІЙНІЙ МОДЕЛІ "РОЗУМНОГО" СТАЛОГО МІСТА

O. Korepanov,
PhD, Associate Professor, Associate Professor of Statistics, Accounting and Auditing Department,
V.N. Karazin Kharkiv National University

PROSPECTS OF USING "BIG" DATA OF MOBILE OPERATORS IN THE INFORMATION MODEL OF "SMART" SUSTAINABLE CITY

Статтю присвячено вивченню перспектив використання "великих" даних операторів мобільного зв'язку в інформаційній моделі "розумного" сталого міста.

У статті представлені деякі існуючі докази важливої ролі "великих" даних операторів мобільного зв'язку у досягненні цілей розвитку в різних сферах політики, включаючи сталий соціально-економічний розвиток та управління надзвичайними ситуаціями (нещасними випадками). Наведені приклади використання "великих" даних операторів мобільного зв'язку для аналізу соціально-економічного розвитку та управління надзвичайними ситуаціями.

За результатами проведеного дослідження встановлено подібності та відмінності в інформації, яку виробляють оператори мобільних мереж, на відміну від операторів фіксованого зв'язку та Інтернет-провайдерів, і представлені деякі додаткові дані, зокрема про місцезнаходження та мобільність, що створюють мобільні мережі та послуги. Визначено напрями статистичного аналізу профілю клієнтів, на основі "великих" даних операторів мобільного зв'язку.

The article is devoted to studying the prospects of using "big" data of mobile communication operators in the information model of "smart" sustainable city.

The article presents some of the existing evidence of the important role of "big" data of mobile operators in achieving development goals in various policy areas, including sustainable socio-economic development and emergency management (accident). Examples are given of the use of "big" data from mobile operators to analyze socio-economic development and emergency management.

According to the results of the survey, similarities and differences in the information produced by mobile network operators, unlike fixed-line operators and ISPs, are established, and some additional data, including location and mobility, that create mobile networks and services are presented. The directions of statistical analysis of the profile of clients are determined on the basis of "big" data of mobile operators.

Ключові слова: "великі" дані, оператори мобільного зв'язку, інформаційна модель, "розумне" стале місто, статистичний аналіз.

Key words: "big" data, mobile communication operators, information model, "smart" sustainable city, statistical analysis.

ВСТУП

Оператори мобільного зв'язку володіють мобільними телекомунікаційними даними, які можуть бути класифіковані за різними типами залежно від характеру інформації, яка виробляється. До них відносяться дані трафіку, дані про сервісний доступ, дані про місцезнаходження та переміщення, характеристики пристрою, дані про клієнтів та дані тарифів.

Зусилля, спрямовані на покращення доступності та доповнення офіційної статистики, призвели до пошуку нових джерел даних, включаючи "великі" дані, які мають величезний потенціал задля створення нової і проникливої інформації.

Таблиця 1. Приклади використання "великих" даних операторів мобільного зв'язку для аналізу соціально-економічного розвитку та управління надзвичайними ситуаціями

Показники	Напрям дослідження	Метод аналізу	Приклади застосування
Соціально-економічний розвиток			
Показники мобільності: – частота подорожей, кількість на рік; – діапазон мобільності; – регіональна, міжрегіональна мобільність	– Зв'язок показників мобільності населення, що отримані з даних мобільної мережі та доходів населення (соціально-економічного рівня)	Кореляційний аналіз	Результати дослідження [2] показали, що населення з більш високими доходами більш сильно пов'язане із більшим діапазоном мобільності, ніж населення з нижчими доходами та соціально-економічним рівнем
– Моделі комунікації; – записи про покупку кредитів на ефірний час	– Картування бідності	Методи візуалізації	Відображення бідності в Кот-д'Івуар з використанням даних мобільної мережі [3]
– Записи про покупку кредитів на ефірний час, поєднані з вивченням соціальних мереж користувачів	– Зв'язок показників витрат на ефірний час з рівнем доходу	Кореляційний аналіз	Результати дослідження [3] показали, що люди, як правило, спілкуються з тими, хто має аналогічну купівельну спроможність (тобто аналогічний рівень доходу). Двоє користувачів розглядалися як пов'язані, якщо вони спілкувалися один з одним принаймні один раз на місяць
– Кількість транзакцій, здійснених з мобільних телефонів	– Зв'язок показників активності здійснення транзакцій, отримані з даних мобільної мережі та соціально-економічного рівня населення (доходів)	Кореляційний аналіз	Результати дослідження [6] показали, що люди, які регулярно купують ефірний час мають певний соціально-економічний рівень
– Кількість транзакцій, здійснених з мобільних телефонів, поєднані з вивченням рівня кредитоспроможності	– Оцінювання кредитоспроможності населення	Дисперсійний аналіз	Кількість транзакцій, здійснених з мобільних телефонів, демонструє передбачуваність доходу, здатність планувати наперед, що може бути позитивним свідченням спроможності повертати позику. У Бразилії компанія Cignifi побудувала 50 поведінкових змінних із 2,3 млн передплачених мобільних підписок та підтвердила результати моделі за історичними даними щодо кредитування приблизно 40 000 позичальників, які користуються кредитуванням оператора мобільного зв'язку Oi Oi Paggo [4]
Надзвичайні ситуації			
– Дані пасивного та активного позиціонування місця знаходження	– Картування місця знаходження	Методи візуалізації для відстеження мобільності населення	Дані про мобільність, зібрані відразу після катастрофи, у багатьох випадках допомагають рятувальникам особам знаходити постраждале населення та дозволяють агентствам надавати допомогу у потрібних місцях [7]

ПОСТАНОВКА ПРОБЛЕМИ

Використання "великих" даних в офіційній статистиці, а також для статистичного аналізу розвитку "розумних" сталих міст пов'язано з численними труднощами.

Дані операторів мобільного зв'язку можуть надавати уявлення про сфери економічного розвитку та соціально-економічний стан населення, часто в реальному часі. Тому вивчення перспектив використання "великих" даних операторів мобільного зв'язку в інформаційній моделі "розумного" сталого міста є надзвичайно актуальним і своєчасним.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Комітет Європейської статистичної системи (European Statistical System Committee — ESSC) у 2013 р. схвалив Меморандум про "Великі Дані та Офіційну Статистику", який зазначив, що Big Data представляють нові можливості та виклики для офіційної статистики та заохочує Європейську Статистичну Систему та її партнерів для ефективного вивчення потенціалу джерел "великих" даних.

Обговоренню питань застосування "великих" даних були присвячені міжнародні конференції з використання великих даних в офіційній статистиці, що відбулися:

I — у м. Пекіні (КНР) 28—30 жовтня 2014 року [5];

II — у м. Абу-Дабі (ОАЕ) 20—22 жовтня 2015 року [10];

III — у м. Дублін (Ірландія) 30 серпня — 1 вересня 2016 року [11];

IV — у м. Богота (Колумбія) 8—10 листопада 2017 року [12].

ПОСТАНОВКА ЗАВДАННЯ

Метою дослідження є вивчення перспектив використання "великих" даних операторів мобільного зв'язку в інформаційній моделі "розумного" сталого міста.

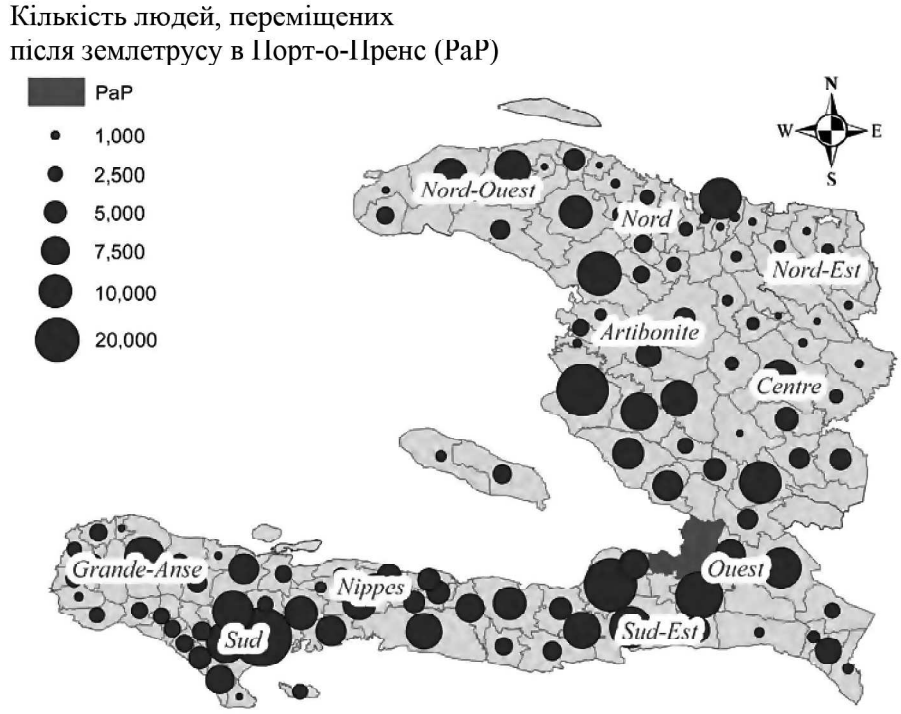

Рис. 1. Відстеження мобільності населення у надзвичайних ситуаціях за допомогою мобільних телефонів

Джерело: [1].

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Створена Глобальною робочою групою СК ООН цільова група з питань даних мобільного телефонного зв'язку завершила розроблення першого повного проекту відповідного довідника, який заснований на досвіді здійснення кількох проектів Євростату та проекту в Омані за участю компанії "Позітіум". Крім того, він також спирається на роботу, виконану в рамках центрів і мереж передового досвіду Європейської статистичної системи (ESSNet) по проектам у Бельгії, Іспанії, Нідерландах, Словенії і Франції.

Довідник включає в себе бізнес-моделі партнерства між національними статистичними управліннями і операторами мобільного зв'язку, додатки для підготовки статистики туризму (на міжнародному, національному та регіональному рівнях), "денної" демографічної статистики (статистики туризму на певну дату), статистики подій, транспортної статистики та статистики щільності населення. Довідник завершується тематичними прикладами за двома країнами — Францією та Індонезією. В цьому довіднику містяться практичні інструкції, інформація про запроваджену країнами практику і безліч довідкових відомостей, що повинно бути особливо корисно для країн, що розвиваються.

Надалі доцільно представити деякі існуючі докази важливої ролі "великих" даних операторів мобільного зв'язку у досягненні цілей розвитку в різних сферах політики, включаючи сталий соціально-економічний розвиток та управління надзвичайними ситуаціями (нещасними випадками) (табл. 1).

Дані операторів мобільного зв'язку можуть надавати уявлення про сфери економічного розвитку та соціально-економічний стан населення, часто в реальному часі.

Нещодавно "великі" дані почали оброблятися та аналізуватися з метою знаходження взаємозв'язків в економічних і соціальних системах, що раніше здійснювалося за допомогою опитувань, експериментів та інших видів збору даних, на основі чого розроблялись рішення і будувалися прогнози.

Приклад того, як дані мобільної мережі можуть відслідковувати переміщення населення — це землетрус у Гаїті-2010. На рисунку 1 показано кількість людей, які, за оцінками, були в Порт-о-Пренс (Port-au-Prince (PaP)) в день землетрусу в Гаїті 2010 року, та опинились за межами столиці через 19 днів.

Оператори фіксованих та мобільних телекомунікаційних мереж, включаючи провайдерів Інтернет-послуг (ISP), є важливим джерелом даних, і для цілей аналізу розглядаються всі форми "великих" даних телекомунікацій (обся-

Таблиця 2. Огляд "великих" даних телекомунікаційної мережі*

	Оператор фіксованого зв'язку	Оператор мобільного зв'язку	Інтернет-провайдер
1	2	3	4
Інформація про рух даних	Обсяг даних		
	Кількість дзвінків		
		Кількість SMS / MMS	
		Інтенсивність навантаження (Ерланг)	
	технологія накопичення статистичних даних, перевірки і фільтрації контенту мережевих пакетів (DPI)		
Детальні записи про доступ до сервісу	Часова позначка використання		
	Тривалість використання		
	Контактна мережа		
	Застосовані типи оплати послуг		
		Тип телефону	
	Технологія, що використовується (2G, 3G, DSL / ADSL та ін.)		
Дані про місцезнаходження	Платіжна адреса		
		Пасивні дані позиціонування (наприклад, ідентифікатор стільника)	
		Дані активного позиціонування (наприклад, трикутник GPS)	
Характеристики пристроїв		Міжнародний ідентифікатор мобільного устаткування IMEI	
	MAC-адреса		
Інформація про клієнта	Демографічні дані клієнтів (наприклад, вік, стать, номер національної ідентифікаційної картки)		
	Історія платежів за післяплатою		
	Платіжна адреса		
	Історія платежів за передоплата		
	Історія сервісного замовлення		
Тарифні дані	Тарифний план		
	* – Дані, що генеруються при транзакціях		
	* –Дані, що накопичені в сховищах даних		

Джерело: [4; 9].

гу, швидкості чи різноманіття). Більшість телекомунікаційних даних можна розглядати як результат дії (наприклад, здійснення дзвінка, відправлення SMS, доступ до Інтернету або перезаряджання передплаченої картки).

"Великі" дані для ініціатив з розвитку переважно спираються на "великі" дані мобільної мережі, а не на операторів фіксованого зв'язку або Інтернет-провайдерів. У таблиці 2 показано деякі подібності та відмінності в інформації, яку виробляють оператори мобільних мереж, на відміну від операторів фіксованого зв'язку та Інтернет-провайдерів, і представлені деякі додаткові дані, зокрема про місцезнаходження та мобільність, що створюють мобільні мережі та послуги.

Оскільки найширше охоплення та найбільше поширення та популярність мають по-

НАПРЯМИ СТАТИСТИЧНОГО АНАЛІЗУ ПРОФІЛЮ КЛІЄНТІВ

Рис. 2. Напрями статистичного аналізу профілю клієнтів на основі "великих" даних операторів мобільного зв'язку

Джерело: побудовано автором за [4; 8].

слуги мобільного стільникового зв'язку, дані операторів мобільного зв'язку мають найбільший потенціал для отримання репрезентативних результатів та виявлення тенденцій соціально-економічного розвитку територіальних громад, і все частіше — територій з низьким рівнем доходу.

Телекомунікаційні оператори фіксують цілий ряд поведінкових даних щодо своїх клієнтів.

Профілі клієнтів містять інформацію про моделі мобільності клієнтів, соціальні мережі та споживчі переваги. У сукупності ці "цифрові сухарі" дозволяють операторам профілювати та сегментувати своїх клієнтів на основі різноманітних показників (метрик) (рис. 2).

Розглянемо більш детально наведені напрями статистичного аналізу профілю клієнтів, на основі "великих" даних операторів мобільного зв'язку:

— Інтереси клієнтів: ці дані можуть бути зафіксовані або, у деяких випадках, виведені з урахуванням часу та / або обсягу пошукових запитів для різних послуг. Такі дані також можуть використовуватися для класифікації інтересів на основі відвідуваних сайтів (на відміну від перегляду контенту). Статистичний аналіз структури пошукових запитів також може окреслити більш детальні класи інтересів;

— Соціально-економічний клас: хоча деталі клієнта часто дозволяють класифікувати соціально-економічний статус своїх клієнтів, такі дані не завжди надійні. З іншого боку, "великі" дані можуть допомогти уточнити цю класифікацію шляхом статистичного аналізу рівнів споживання різних послуг, у тому числі на основі витрат (часто стосовно інших послуг), типів використовуваних пристроїв, частоти зміни телефонної трубки тощо;

— Ймовірність відмови клієнтів: це показник кількості клієнтів, які залишають мережу або особливу послугу, запропоновану оператором. Аналіз "великих" даних може допомогти операторам змоделювати ймовірність клієнтів залишити мережу (або відмовитись від певної послуги), зосереджуючись на поточній поведінці клієнта та використаних послугах.

— Рівень впливу на клієнтів: оператори прагнуть поширювати послуги та технології серед своїх абонентів з метою маркетингу додаткових, індивідуальних послуг. Це часто вимагає розуміння рівня впливу соціальних мереж на кожного абонента як в мережі (тобто в межах одного оператора), так і поза мережею (тобто в мережах конкурента). Визначивши що велика кількість користувачів перебувають в мережі конкурента, оператори можуть націлювати абонента та / або користувачів поза мережею на рекламні акції та стимули, спрямовані на залучення користувачів у мережу певного оператора.

— Профіль мобільності: оператори мобільного зв'язку надають першочергову увагу визначенням місць, де клієнти найчастіше користуються запропонованими послугами, з тим щоб не лише забезпечити високу якість обслуговування, але також для створення профілів мобільності своїх клієнтів, які можуть бути використані для надання послуг на основі місцезнаходження.

ВИСНОВКИ

Залежно від країни чи регіону можуть існувати різні правила конфіденційності даних, що регулюють спосіб, яким оператори можуть зберігати та / або використовувати клієнтські дані. У цьому випадку, міра використання профілю поведінки, що використовується операторами, може сильно відрізнятися. Удосконалений профіль клієнтів дозволяє операторам більш ефективно персоналізувати та продавати нові послуги. Наприклад, розуміючи відносини своїх клієнтів з соціальними мережами, оператори можуть моделювати поширення послуг та проводити цільові рекламні акції.

У подальших дослідженнях доцільно провести детальний аналіз джерел "великих" даних за напрямами: дані супутникових зображень; дані соціальних мереж; дані скануючих пристроїв тощо з метою подальшого розроблення концепції статистичного аналізу "великих" даних задля оцінювання розвитку "розумних" сталих міст.

Література:

1. Bengtsson L., Lu X., Thorson A., Garfield R., Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti // PLOS Medicine. 2011. Vol. 8, Issue 8. URL: http://journals.plos.org/plosmedicine/ article?id=10.1371/journal.-pmed.1001083

2. Frias-Martinez V., Virseda J. On the relationship between socio-economic factors and cell phone usage // In Proceedings of the Fifth International Conference on Information and Communication Technologies and Development-ICTD'12 New York, USA. ACM Press. 2012. 76 p. URL: 10.1145/2160673.2160684

3. Gutierrez T., Krings G., Blondel V. D. Evaluating socio-economic state of a country analyzing

airtime credit and mobile phone datasets. 2013. — P. 1—6. URL: http://arxiv.org/abs/1309.4496

4. ITU GSR discussion paper (2014) Big Data — Opportunity or Threat // International Telecommunication Union. URL: https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2014/Discussion%20papers%20and%20presentations%20-%20GSR14/Session3_GSR14-DiscussionPaper-BigData.pdf

5. International Conference on Big Data for Official Statistics / United Nations. URL: https://unstats.un.org/unsd/trade/events/2014/beijing/default.asp

6. Kumar K., Muhota K. Can Digital Footprints Lead to Greater Financial Inclusion? — Washington DC. 2012. — P. 1—4.

7. Lu X., Bengtsson L., Holme P. Predictability of population displacement after the 2010 Haiti earthquake // Proceedings of the National Academy of Sciences of the United States of America. 2012.109(29). P. 11576-11581. URL: http://www.pnas.org/content/109/29/11576.full.pdf

8. Measuring the Information Society Report 2014 // ITU. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/ publications/ mis2014/MIS2014_ without_Annex_4.pdf

9. Naef E., Muelbert P., Raza S., Frederick R., Kendall J., Gupta N. Using mobile data for development // Gates Foundation Whitepaper. 2014. — 52 p.

10. 2nd Global International Conference on Big Data for Official Statistics // United Nations / UN Trade statistics. URL: https://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp

11. 3rd International Conference on Big Data for Official Statistics // United Nations / UN Trade statistics. URL: https://unstats.un.org/unsd/bigdata/conferences/2016/default.asp

12. 4th International Conference on Big Data for Official Statistics // United Nations / UN Trade statistics. URL: https://unstats.un.org/unsd/bigdata/conferences/2017/default.asp

References:

1. Bengtsson, L. Lu, X. Thorson, A. Garfield, R. Schreeb, J. (2011), "Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti", PLOS Medicine, Vol. 8, Issue 8, available at: http://journals.plos.org/plosmedicine/ article?id=10.1371/journal.pmed.-1001083 (Accessed 20 May 2018).

2. Frias-Martinez, V. Virseda, J. (2012), On the relationship between socio-economic factors and cell phone usage. In Proceedings of the Fifth International Conference on Information and Communication Technologies and Development-ICTD'12, New York, USA, ACM Press, available at: 10.1145/2160673.2160684 (Accessed 20 May 2018).

3. Gutierrez, T. Krings, G. Blondel, V. D. (2013), "Evaluating socio-economic state of a country analyzing airtime credit and mobile phone datasets", [Online], available at: http://arxiv.org/abs/1309.4496 (Accessed 20 May 2018).

4. International Telecommunication Union (2018), "ITU GSR discussion paper (2014) Big Data — Opportunity or Threat", available at: https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2014/Discussion%20papers%20and%20presentations%20-%20GSR14/Session3_GSR14-DiscussionPaper-BigData.pdf (Accessed 20 May 2018).

5. United Nations (2018), "International Conference on Big Data for Official Statistics", available at: https://unstats.un.org/unsd/trade/events/2014/beijing/default.asp (Accessed 20 May 2018).

6. Kumar, K. Muhota, K. (2012), Can Digital Footprints Lead to Greater Financial Inclusion?, Washington DC.

7. Lu, X. Bengtsson, L. Holme, P. (2012), "Predictability of population displacement after the 2010 Haiti earthquake", Proceedings of the National Academy of Sciences of the United States of America [Online], 109(29), pp. 11576-11581, available at: http://www.pnas.org/content/109/29/11576.full.pdf (Accessed 20 May 2018).

8. ITU (2018), "Measuring the Information Society Report 2014", available at: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_ without_-Annex_4.pdf (Accessed 20 May 2018).

9. Naef, E. Muelbert, P. Raza, S. Frederick, R. Kendall, J. Gupta, N. (2014), Using mobile data for development, Gates Foundation Whitepaper.

10. United Nations (2018), "2nd Global International Conference on Big Data for Official Statistics", available at: https://unstats.un.org/unsd/trade/events/2015/abudhabi/default.asp (Accessed 20 May 2018).

11. United Nations (2018), "3rd International Conference on Big Data for Official Statistics", available at: https://unstats.un.org/unsd/bigdata/conferences/2016/default.asp (Accessed 20 May 2018).

12. United Nations (2018), "4th International Conference on Big Data for Official Statistics", available at: https://unstats.un.org/unsd/bigdata/conferences/2017/default.asp (Accessed 20 May 2018).

Стаття надійшла до редакції 05.06.2018 р.