

## ПЕРЕДБАЧУВАНІСТЬ СТАНУ З'ЄДНАННЯ ПРИ МАРШРУТИЗАЦІЇ В ОПОРТУНІСТИЧНИХ МЕРЕЖАХ

У статті порівнюється складність прогнозування місцезнаходження із задачею, яка розв'язується при маршрутизації, за допомогою нововведеного поняття стану з'єднання. Показано, що показники ентропії стану з'єднання є значно вищими за значення для місцезнаходження, зроблено висновки про більшу складність передбачення.

**Ключові слова:** опортуністичні мережі, передбачення місцезнаходження, ентропія, стан з'єднання.

### Вступ

Останні роботи в галузі дослідження людського переміщення вказують на високий потенціал передбачення місцезнаходження особи. Наприклад, González та ін. [2] показали, що типові відстані, які покриваються людиною протягом дня, мають степеневий закон розподілу. Це вказує на те, що існує визначена множина місць, які людина відвідує порівняно часто, і решта локацій, які відвідуються значно рідше. Аналіз передбачуваності переміщення людини, проведений Song та ін. [5], показує, що відповідний алгоритм може визначати наступне місцезнаходження з імовірністю 0.93. Це пояснюється тим, що ентропія інформації у траєкторіях руху людей є невисокою.

У контексті мобільних опортуністичних мереж, де вузлами є мобільні телефони, які переносяться людьми, результати зазначених досліджень дають підстави вважати, що алгоритм маршрутизації у таких мережах можливо побудувати на базі машинного навчання. Приклад такого алгоритму описано в [1, с. 152]. До таких алгоритмів можна віднести і PROPHEET [4], в основі якого лежить обчислення імовірності зустріти певний вузол повторно на базі історії зустрічей. Однак більшість розроблених протоколів маршрутизації для опортуністичних мереж не оперує поняттям місцезнаходження людини.

### Стан з'єднання

Мобільні опортуністичні мережі передбачають оточення, у якому кожен з вузлів, змінюючи своє місцезнаходження, набуває нових з'єднань з іншими вузлами та позбувається вже існуючих. Отже, стан вузла в такому середовищі можливо задати через множину інших вузлів, з якими у певний момент часу встановлено з'єднання.

Якщо розглянути опортуністичну мережу як граф  $G = (V, E)$ , у якому вершини відповідають вузлам мережі, а ребра відображують наявність з'єднання між вузлами, то для  $i$ -го вузла мережі  $v_i \in V$  стан з'єднання описується як множина  $X^i = \{v_j \in V \mid \exists (v_i, v_j) \in E\}$ . А послідовність з  $N_i$  змін стану з'єднання для вузла  $i$  можна розглядати як Марковський процес  $\chi^i = \{X_k^i \mid k = \overline{1..n}\}$ .

Для аналізу передбачуваності стану з'єднання використовується подібний до продемонстрованого підхід для визначення передбачуваності людського переміщення [5]. Для кожного вузла (особи) розглядаються значення ентропії трьох типів:

1. випадкова ентропія  $S_i^{rand} = \log_2 N_i$ , де  $N_i$  – кількість унікальних станів з'єднання для вузла  $i$ , – характеристика передбачуваності станів з'єднання з умовою, що поява кожного з них має однакову імовірність;

2. частотна ентропія  $S_i^{freq} = -\sum_{j=1}^{N_i} p_{ij} \log_2 p_{ij}$ , де  $p_{ij}$  –

частота появи стану  $j$  для вузла  $i$ , – враховує імовірність кожного з унікальних станів з'єднання, характеризуючи різноманіття шаблонів цих станів;

3. часова ентропія  $S_i^{time} = -\sum_{j=1}^{N_i} p_{ij}^T \log_2 p_{ij}^T$ , де  $p_{ij}^T$  –

імовірність перебування вузла  $i$  в стані  $j$  протягом усього часу спостереження, – характеристика, яка враховує час перебування користувача у певному стані.

Очевидно, що  $S_i^{freq} \leq S_i^{time} \leq S_i^{rand}$ , оскільки випадкова ентропія розраховується з припущенням, що імовірність появи будь-якого стану має рівномірний розподіл, тоді як способи обчислення кожної наступної ентропії передбачають додаткові закономірності у розподілі станів з'єднання.

## Обробка даних

Зібраний у Дартмутському коледжі набір даних [3], на якому проводився аналіз, містить інформацію про те, до якої точки доступу та в який час було підключено ноутбук певного студента. Ці дані дають можливість оцінити стан з'єднання студентів протягом усього їхнього часу перебування на території коледжу, оскільки спостереження проводилося безперервно протягом періоду тривалістю більш ніж три роки, починаючи з квітня 2001 року, закінчуючи червнем 2004 року. Стан з'єднання є невідомим для нас, коли студент знаходиться за межами кампусу, однак, оскільки цей час є порівняно невеликим, його впливом на значення ентропії можна знехтувати.

Користувачі (студенти) вважаються з'єднаними між собою, якщо вони підключені до однієї точки доступу в один момент часу. Таким чином, історію станів з'єднання кожного зі студентів можна отримати, розглядаючи кожен запис вхідного набору даних як подію зміни стану з'єднання та обробивши всі події у хронологічному порядку. Кожна з таких подій несе інформацію про відключення або підключення користувача до відповідної точки доступу. Обробка цієї події створює новий запис в історії станів кожного з користувачів, підключених до даної точки.

Із десятка тисяч студентів, про яких є дані у вхідному наборі, було розглянуто 4666, у кожного з яких було не менше 1500 записів.

## Результати

Для характеристики передбачуваності станів з'єднання було розраховано значення ентропій  $S_i^{rand}$ ,  $S_i^{freq}$ ,  $S_i^{time}$  для кожного з розглянутих користувачів. Розподіл значень цих ентропій  $P(S^{rand})$ ,  $P(S^{freq})$ ,  $P(S^{time})$ , округлених до цілого, наведено на рис. 1.

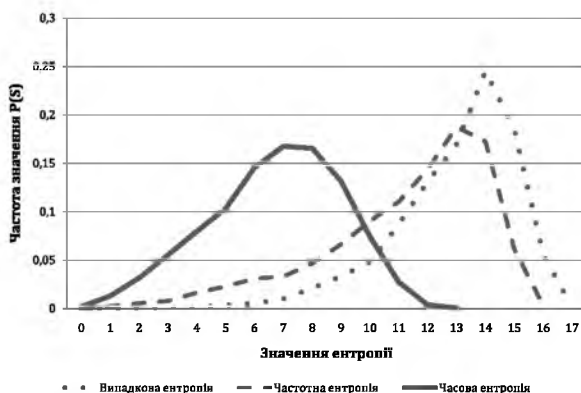


Рис. 1. Розподіл значень ентропій стану з'єднання, обчислених для кожного з користувачів

На графіках розподілу значень ентропії (рис. 1) чітко простежується співвідношення  $S_i^{time} < S_i^{freq} < S_i^{rand}$ . При цьому спостерігається значний зсув розподілу часової ентропії у порівнянні з частотною та випадковою. Пік у розподілі випадкової ентропії  $P(S^{rand})$  знаходиться біля значення  $S^{rand} = 14$ . Це означає, що, якби наступний стан з'єднання обирався випадково, то він би мав одне з  $2^{14}$  значень. Однак урахування часу перебування в станах дає можливість звужити область пошуку до  $2^8$  варіантів (для піку в  $S^{time} = 8$ ).

Разом із зсувом розподіл часової ентропії  $P(S^{time})$  також стає більш подібним до розподілу Коші. Цей факт говорить про те, що у часової ентропії немає очікуваного значення і вказує на ускладнення задачі передбачення стану з'єднання.

Цікаво порівняти отримані результати передбачуваності стану з'єднання з передбачуваністю наступної позиції при переміщенні користувачів. Для цього точка доступу у вхідному наборі даних може розглядатися як унікальне місцезнаходження. Застосувавши той же підхід, що й для стану з'єднання, було побудовано графіки розподілу значень ентропій для локацій кожного з користувачів (рис. 2).

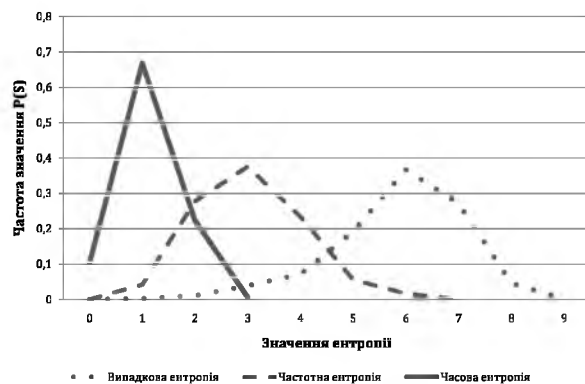


Рис. 2. Розподіл значень ентропій місцезнаходження, обчислених для кожного з користувачів

Отримані розподіли для місцезнаходження користувачів близькі до вже раніше відомих [5] і говорять про високий потенціал у передбаченні позиції людини (оскільки пік часової ентропії є яскраво вираженим і знаходиться біля дуже малого значення  $S^{time} = 1$ , що вказує лише на 2 можливі наступні локації для користувача). Це дуже сильно відрізняється від результатів для стану з'єднання.

Однак такий контраст виглядає достатньо природним, якщо взяти до уваги, що стан з'єднання є залежною величиною від позицій багатьох користувачів і провести аналогію

з утворенням розподілу Коші через відношення двох випадкових величин з нормальним розподілом (який дає значно кращий потенціал для передбачення).

шим за визначення наступного місцезнаходження вузла. На це вказує виконаний аналіз ентропії стану з'єднання, поняття, введеного для дослідження алгоритмів опортуністичної маршрутизації.

### Висновки

Хоча високий потенціал передбачуваності місцезнаходження людини і дає перспективи у впровадженні алгоритмів машинного навчання в опортуністичній маршрутизації, завдання передбачення для таких систем є значно складні-

### Подяка

За допомогу у підготовці статті та її коректуванні висловлюю вдячність Семену Самуїловичу Гороховському, к. ф.-м. н., професору факультету інформатики Національного університету «Києво-Могилянська академія».

### Список літератури

1. Теоретичні та прикладні аспекти побудови програмних систем: праці конференції «ТААПСД'2012» (Київ, 3–7 грудня 2012 р.) / ред. Д. Б. Буя, М. С. Нікітченка. – 348 с.
2. González M. C. Understanding individual human mobility patterns / M. C. González, C. A. Hidalgo, A. Barabási // *Nature*. – 2008. – Vol. 453, № 7196. – P. 779–782.
3. CRAWDAD data set dartmouth/campus (v. 2009–09–09) [Електронний ресурс] / D. Kotz, T. Henderso, I. Ayzov, J. Yeo. – 2009. – Режим доступу: <http://crawdad.cs.dartmouth.edu/dartmouth/campus>. – Назва з екрана.
4. Lindgren A. Probabilistic Routing in Intermittently Connected Networks / A. Lindgren, A. Doria, O. Schele'n // *SIGMOBILE Mobile Computing Comm. Rev.* – 2003. – Vol. 7, № 3. – P. 19–20.
5. Song C. Limits of predictability in human mobility / C. Song, Z. Qu, N. Blumm, A. L. Barabási // *Science*. – 2010. – Vol. 327, № 5968. – P. 1018–1021.

R. Mazur

## CONNECTIVITY STATE PREDICTABILITY IN OPPORTUNISTIC NETWORKS ROUTING

*The paper compares complexity of location prediction with a task being solved in opportunistic routing using an newly introduced term of connectivity state. It is shown that connectivity state entropy values are much higher than location entropy, inducing higher prediction complexity.*

**Keywords:** opportunistic networks, location prediction, entropy, connectivity state.

Матеріал надійшов 10.10.2013