

УДК 004.415.3

Приймачук В.О., Пех П.А.

Луцький національний технічний університет

## ПРОГНОЗУВАННЯ ЧАСУ ПРИБУТТЯ МАРШРУТНИХ ТРАНСПОРТНИХ ЗАСОБІВ НА ЗУПИНКУ НА ОСНОВІ ДАНИХ GPS

*Дана стаття описує модель прогнозування часу прибуття маршрутних транспортних засобів на зупинку. Ця модель використовує метод лінійної регресії та дані про положення транспортного засобу з системи GPS. Спочатку описується метод визначення положення транспортного засобу на маршруті. Для цього використовується алгоритм прив'язки даних про положення до транспортної мережі. Далі описується метод, за допомогою якого можна побудувати модель прогнозування для окремої зупинки на основі даних про відстань та час, що залишилися від поточного положення транспортного засобу до відповідної зупинки.*

*Ключові слова: інтелектуальні транспортні системи, ІТС, пасажирські перевезення, прогнозування, регресія.*

*Форм. 5. Рис. 6. Літ. 7.*

**Постановка проблеми.** Громадський транспорт відіграє значну роль в житті більшості населення України. За 2009 р. на загальних автомобільних маршрутах перевезено 4,014 млрд. пасажирів. За прогнозами обсяг перевезень пасажирів автомобільним транспортом збільшиться до 6,613 млрд. до 2025 р. [1]. Разом із задоволенням потреб населення у пасажирських перевезеннях існує очевидний дефіцит в якості, комфорті та безпеці на всіх видах транспортних засобів. Ця проблема в області міських пасажирських автобусних перевезень може бути вирішена за рахунок розробки та використання інтелектуальних транспортних систем (ІТС), зокрема, системи прогнозування прибуття транспортних засобів на конкретну зупинку. Ця ІТС є важливою як для пасажирів, так і для перевізників. Володіючи інформацією про час прибуття транспортного засобу на конкретну зупинку, перевізники можуть своєчасно реагувати на надзвичайні ситуації на дорозі та вживати заходи для вирішення проблем. В перспективі ця ІТС дозволить визначати проблемні маршрути, передбачати зміни у графіках руху тощо. Завдяки цьому якість послуг, які надають перевізники, зросте, що в свою чергу збільшить привабливість громадського транспорту в порівнянні з персональним.

**Аналіз останніх досліджень та публікацій.** За останні десятиліття було виконано багато досліджень в області розробки ІТС. Можна виділити три основні напрямки цих досліджень: моделі передбачення, що базуються на історичних даних; лінійні регресійні моделі; моделі на основі штучних нейронних мереж.

Моделі першого типу роблять висновок про поточний та майбутній час поїздки на основі історичних даних руху для того ж транспортного засобу або інших транспортних засобів, що рухаються по тому ж маршруту. Lin та Zeng запропонували набір алгоритмів прогнозування для інформаційної системи транзитних перевезень в м. Blackburg, штат Вірджинія, США[2]. Було реалізовано чотири алгоритми з різними гіпотезами щодо вхідних даних і показано перевагу над декількома іншими алгоритмами, описаними в літературі. Kidwell запропонував алгоритм прогнозування, що базується на положенні транспортного засобу в режимі реального часу[3]. Цей алгоритм ділить маршрут на кілька зон і відслідковує час, за який кожен транспортний засіб перетинає кожну зону. Прогноз базується на останніх показниках часу проходження зони. Проте, ці алгоритми не беруть до уваги нестабільність транспортних потоків і різний час стоянки на зупинці.

Моделі другого типу застосовують математичні методи для прогнозування очікуваного часу руху між зупинками та очікуваний час прибуття транспортного засобу на зупинку. Основою таких моделей є різні види рівнянь регресії. Patnik розробив набір регресійних моделей для прогнозування часу прибуття на зупинку на основі даних про кількість пасажирів з системи автоматичного підрахунку пасажирів[4]. Отримані результати показують, що дана модель має право на існування та показує достатньо гарні показники. Моделі даного типу забезпечують кращий прогноз на ділянках з великими похибками у вимірюваннях. Для покращення прогнозу в таких моделях можна використати дані про поточне положення транспортного засобу на маршруті.

Моделі третього типу передбачають використання штучних нейронних мереж (ШНМ), які можуть моделювати складні нелінійні залежності. Jeong та Rilett запропонували модель на основі ШНМ для прогнозування та показали, що вона має значні переваги в продуктивності в порівнянні з іншими методами [5]. Проте, моделі на основі ШНМ потребують інтенсивного навчання, та не гарантують стабільності результатів.

**Невирішені частини проблеми.** Огляд літературних джерел показує, що вирішенню проблеми розроблення моделі прогнозування часу прибуття маршрутних транспортних засобів на зупинку на основі даних GPS приділялося недостатньо уваги.

**Метою даного дослідження** є розроблення моделі прогнозування часу прибуття маршрутних транспортних засобів на зупинку на основі даних GPS.

**Основні результати дослідження.** Для дослідження обрано маршрут №10 громадського транспорту м. Луцька, на якому здійснюються регулярні пасажирські перевезення в режимі маршрутного таксі. Було проаналізовано рух транспортних засобів на маршруті за період від 01.06.2012 до 31.12.2012. Було побудовано лінійну модель маршруту руху, досліджено можливості прив'язки даних з пристроїв відстеження положення до транспортної мережі та побудовано лінійну регресійну модель для прогнозування часу прибуття маршрутних транспортних засобів на зупинку.

На маршруті №10 можна виділити 2 напрямки руху:

- Шота Руставелі - Інститут розвитку людини
- Інститут розвитку людини - Шота Руставелі.

Даний маршрут обслуговують 11 транспортних засобів. Всі транспортні засоби маршруту обладнані пристроями відстеження місцеположення. Вони визначають місцеположення транспортного засобу на основі сигналів від штучних супутників землі. Для цього використовуються GPS, Глонасс чи подібні технології. Похибка показів для транспортних засобів в загальному становить менше 10м і залежить від точності годинника на пристрої та від методів корекції, що використовуються.

Можна виділити два типи даних, що надходять від пристроїв відстеження: location-at-time та time-at-location. Дані першого типу є періодичною інформацією про місцеположення маршрутного транспортного засобу з інтервалом в 30 с. Повідомлення даного типу містить ідентифікатор транспортного засобу, місцеположення в даний момент часу (координати довготи та широти). Дані другого типу містять інформацію про час, коли транспортний засіб перетнув деяку область, задану координатами.

Геометрія маршруту, інформація про зупинки та географічне положення останніх отримані з баз даних проекту OpenStreetMap [6].

**Модель руху транспортного засобу на маршруті.** Для визначення траєкторії руху транспортного засобу та дистанції до зупинок необхідно виконати проекцію координат на відповідну транспортну мережу для кожного повідомлення GPS. В цифровій моделі транспортної мережі маршрут представляє собою послідовність лінійних сегментів, які є наближенням їхнього істинного географічного положення. Модель руху необхідна для наступних кроків визначення положення транспортного засобу на маршруті та обчислення віддалі від транспортного засобу до всіх наступних зупинок.

**Модель прогнозування часу прибуття.** Алгоритм прогнозування складається з двох етапів. На першому етапі виконується попередня обробка даних про положення: проектування координат на векторну модель маршруту, ідентифікація рейсу, обчислення відстані до кожної зупинки. На другому етапі виконується безпосередній прогноз часу прибуття до кожної з зупинок на основі поточного положення транспортного засобу на маршруті.

Перший крок в процедурі визначенні часу прибуття становить визначення точного положення транспортного засобу на маршруті. Для цього адаптовано метод map-matching, описаний в [7]. Алгоритм map-matching ставить за мету визначення найкращого сегменту транспортної мережі на основі вибірки координат, яка представляє траєкторію руху транспортного засобу.

Алгоритм працює наступним чином. Нехай для точки  $p_{i-1}$  попереднього положення вже визначено сегмент, який їй відповідає –  $\tilde{n}_3$ . Необхідно визначити сегмент мережі, що відповідає положенню  $p_i$ . Сегменти-кандидати для визначення поточного положення визначаються, як множина сегментів, безпосередньо з'єднаних з попередньо визначеним сегментом  $\tilde{n}_3$ . На рис.1 це сегменти  $c_1$  та  $c_2$ .

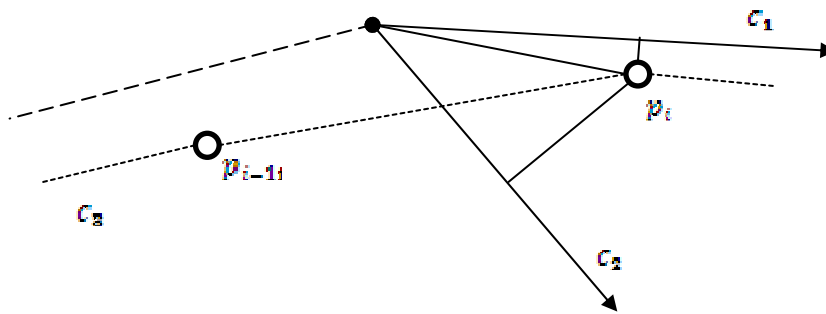


Рис. 1 – Приклад застосування алгоритму map-matching

Для оцінювання сегментів-кандидатів використовуються два показники –  $s_d$  та  $s_\alpha$ . Показник  $s_d$  відображає дистанцію від тестового положення до сегмента транспортної мережі, і обчислюється як середньозважена лінійна відстань до сегмента  $d_i$  для поточного положення  $p_i$  та кожного кандидата  $c_j$ , використовуючи фактори масштабування  $\mu_d$ ,  $n_d$  та  $a$ .

$$s_d(p_i, \tilde{n}_j) = \mu_d - a * d(p_i, \tilde{n}_j)^{n_d}. \quad (1)$$

Показник  $s_\alpha$  характеризує орієнтацію траєкторії руху для відповідного сегмента-кандидата  $c_j$  та поточного положення  $p_i$ . Він обчислюється на основі різниці в кутах  $\alpha_{i,j}$  між сегментом-кандидатом  $c_j$  та лінійним сегментом  $l_i = \overline{p_{i-1}, p_i}$ , використовуючи фактори масштабування  $\mu_a$  та  $n_a$ .

$$s_\alpha(p_i, c_j) = \mu_a * \cos(\alpha_{i,j})^{n_a} \quad (2)$$

Фактори масштабування  $\mu_{[d|\alpha]}$  та  $n_{[d|\alpha]}$  обмежують максимальну оцінку та силу параметра відповідно.

Комбінована оцінка  $s$  обчислюється як сума показників  $s_d$  та  $s_\alpha$ :

$$s = s_d + s_\alpha \quad (3)$$

Чим вища ця оцінка, тим кращий результат.

Для маршруту №10 було встановлено, що показник  $s_d$  відіграє більшу роль. Експериментально встановлено наступні параметри для алгоритму map-matching:  $\mu_d = 10$ ,  $n_d = 0.5$ ,  $a = 1.3$ ,  $\mu_a = 0.5$ ,  $n_a = 0.5$ .

Наступний крок – ідентифікація рейсів. В процесі ідентифікації визначаються рейси: сукупність послідовних у часі та напрямі точок відмітки, які формують траєкторію руху маршрутного транспортного засобу на маршруті. Рейс можна представити як одновимірний лінійний простір, по якому рухаються транспортні засоби (рис. 2).

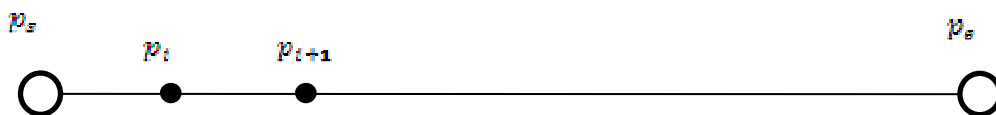


Рис. 2 – Лінійне представлення руху транспортного засобу по маршруту

На цьому рисунку  $P_s$  – точка старту маршруту,  $P_e$  – точка закінчення маршруту,  $P_i$  та  $P_{i+1}$  – довільні положення транспортних засобів на лінії маршруту.

Якщо продиференціювати функцію відносного положення транспортного засобу на рейсі по часу, отримаємо графік, зображений на рис. 3.

Можна помітити, що різниця між двома послідовними повідомленнями додатна в межах одного рейсу. В процесі ідентифікації рейсів виникає дві проблеми:

- визначення початку нового та закінчення попереднього рейсу;
- випадкові відхилення при стоянці автобуса на зупинці, пов'язані з похибкою GPS.

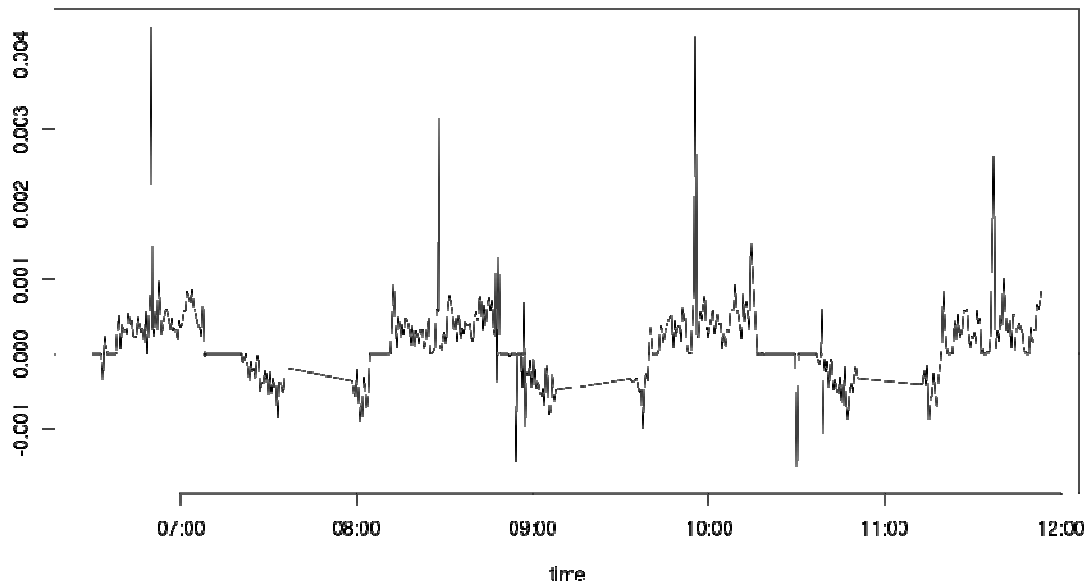


Рис. 3 – Похідна функції положення транспортного засобу по часу

Можна сказати, що рейс починається, коли похідна функції змінює свій знак на додатній. Для уникнення хибних спрацювань, можна дещо трансформувати похідну, сумуючи попередні  $N$  показників (рис. 4).

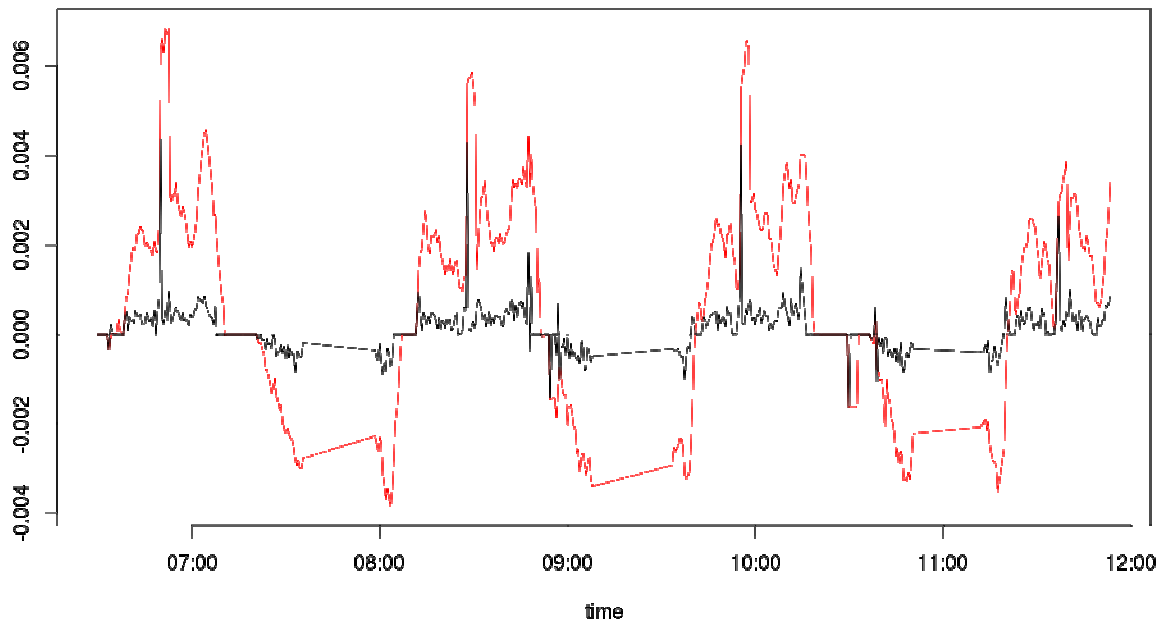


Рис. 4 – Результат сумування кількох попередніх показників

Ця трансформація підкреслює межі рейсу та згладжує випадкові відхилення при стоянці автобуса на зупинці. В результаті точність алгоритму ідентифікації рейсів на тестових даних складає 97,4%.

Для побудови регресійної моделі необхідно визначити час та відстань до зупинки для кожного повідомлення про зміну положення. Відстань до зупинки  $d_{i,j}$  можна визначити як різницю між дистанцією  $dz_i$ , що показує відстань від початкової точки маршруту до зупинки, та фактичним положенням транспортного засобу на маршруті  $p_j$ .

$$d_{i,j} = dz_i - p_j \quad (4)$$

Час до зупинки можна отримати з історичних даних, які надають інформацію про точний час прибуття транспортного засобу на відповідну зупинку. Це дані типу time-at-location, і вони потребують об'єднання з даними про фактичне положення. Для цього в нагоді стають визначені на першому етапі рейси. Знаючи час початку та закінчення рейсу (необов'язково точний), можна побудувати вибірку всіх даних про відвідання зупинки для даного рейсу.

```
library('rjson')
library('forecast')
default_par <- par(no.readonly=TRUE)
par(pch=19, lty=1)

data <- fromJSON(file="data/october_78.json")
testData <- data.frame(data$time_to_stop, 1 - data$progress)
names(testData) <- c("time", "distance")
plot(jitter(time) ~ jitter(distance), data=testData,
      xlab="distance to stop", ylab="time to stop", type="p")
```

На рис. 5 побудовано залежність часу до зупинки від дистанції до зупинки для зупинки "Депо" маршруту №10 в напрямку Конякіна–Шота Руставелі за жовтень 2012. На осі абсцисс дистанція до зупинки у відсотках, на осі ординат – час, що залишився до зупинки в секундах.

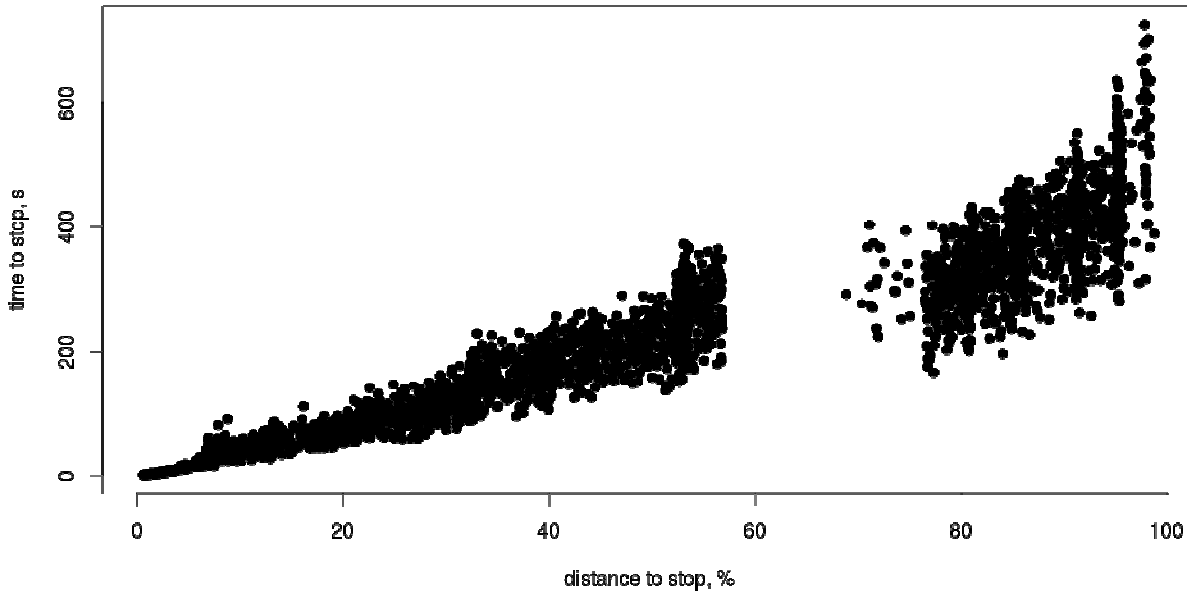


Рис. 5 – Графік залежності часу до зупинки від дистанції до зупинки

Авторська розробка

Графік показує, що існує деяка лінійна залежність між цими двома величинами. Побудуємо лінійну регресійну модель для даних.

```
fit <- lm(time ~ distance, data=testData)
summary(fit)
abline(fit, col="red", lwd=4)
```

Параметри побудованої моделі:

```
Call:
lm(formula = time ~ distance, data = testData)
Residuals:
    Min       1Q   Median       3Q      Max
-175.672  -21.889   -0.844   21.879  291.516
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -4.581     1.566  -2.926  0.00346 **
distance      448.867     2.624 171.040 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 46.39 on 3497 degrees of freedom
Multiple R-squared:  0.8932,    Adjusted R-squared:  0.8932
F-statistic: 2.925e+04 on 1 and 3497 DF,  p-value: < 2.2e-16
```

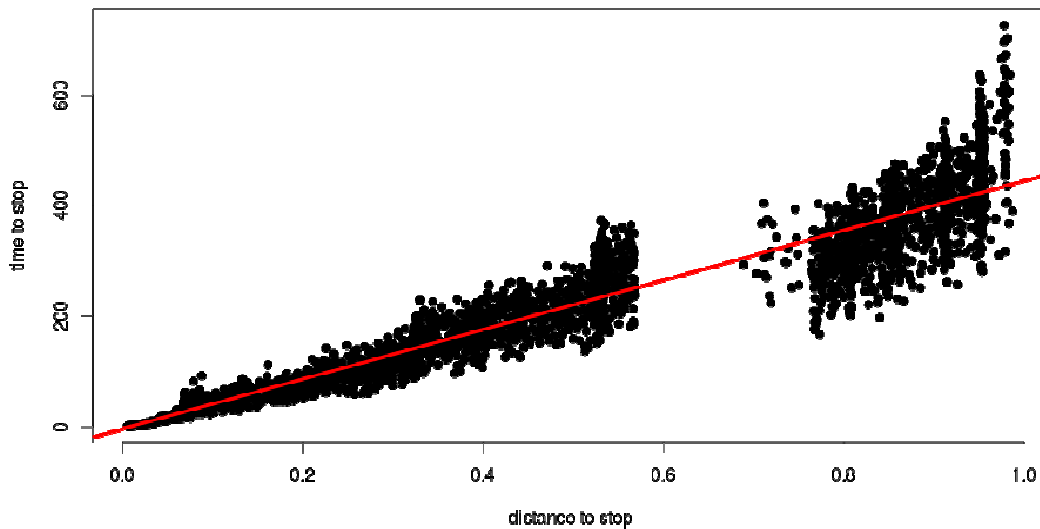


Рис. 6 – Результат побудови регресійної моделі

Авторська розробка

Побудована лінійна залежність має наступний вигляд:

$$y = 448.867 * x - 4.581. \quad (5)$$

**Висновки.** Точний прогноз часу прибуття маршрутних транспортних засобів на зупинку не тільки допомагає пасажиром планувати свій вихід з дому або роботи та зменшити час очікування на зупинці, а ще й дозволяє перевізникам реагувати на стан перевезень в режимі реального часу. Модель прогнозування, описана в даній статті, дозволяє робити точний прогноз на основі даних про поточне положення транспортних засобів на маршруті. Етап попередньої обробки даних є важливим і займає значне місце в процесі побудови моделі прогнозування та безпосереднього прогнозування часу прибуття. Є доцільним продовжити подальші дослідження в цьому напрямку, зокрема щодо більш точного визначення положення транспортного засобу на маршруті в режимі реального часу. Це дозволить адаптувати та коректувати прогноз по мірі наближення транспортного засобу до зупинки.

1. Підтримка інтеграції України до Транс-Європейської транспортної мережі ТСМ-Т.РК4. Автомобільний транспорт. Заключний звіт 4.1. – 2010.
2. Lin, W.-H. and J. Zeng. Experimental Study on Real-Time Bus Arrival Time Prediction with GPS Data. In Transportation Research Record: Journal of the Transportation Research Board, No. 1666, TRB, National Research Council, Washington, D.C., 1999, pp. 101–109.
3. Kidwell, B. Predicting Transit Vehicle Arrival Times. GeoGraphics Laboratory, Bridgewater State College, Bridgewater, Mass., 2001.
4. Patnaik, J., S. Chien, and A. Bladikas. Estimation of Bus Arrival Times Using APC Data. Journal of Public Transportation, Vol. 7, No. 1, 2004, pp. 1–20.
5. Jeong, R., and L. R. Rilett. Bus Arrival Time Prediction Using Artificial Neural Network Model. Proc., IEEE Intelligent Transportation Systems Conference, Washington, D.C., 2004, pp. 988–993.
6. Вільний проект спільного розвитку загальнодоступних мап і схем міст, вулиць, доріг. [Електронний ресурс] Режим доступу: <http://openstreetmap.org>.
7. S. Brakatsoulas, D. Pfoser, R. Salas, and C. Wenk. On map-matching vehicle tracking data. In Proc. Intl. Conf. Very Large Data Bases, pages 853–864, 2005.