



## Fusion of travel time data in Niayesh tunnel using Bayesian inference

M. Mesbah<sup>1\*</sup>, Z. Pakdaman<sup>2</sup>, M. Mollajafari<sup>1</sup>, A. Golroo<sup>1</sup>

<sup>1</sup>Department of Civil & Environmental Engineering, Amirkabir University of Technology, Tehran, Iran.

<sup>2</sup>Department of Civil Engineering, the University of Science & Technology, Tehran, Iran.

**ABSTRACT:** As data collection costs decrease, transportation systems have shifted from systems requiring data to systems requiring data analysis. Since the accuracy of these data varies with the sources of data collection, acquiring higher-accuracy data from a combination of multiple sources is the main challenge of working with such data. Data fusion is a very efficient mechanism that can interconnect data from different sources to increase the accuracy of data in line with the purpose of the study. The main goal of this article is to get the most accurate travel time possible from multiple sources. Among the data fusion methods are the Kalman filter, Bayesian inference, artificial neural networks and Dumpster-Scheffer theory, from which the Bayesian inference is used, and its results are investigated. It is proposed that by combining different data sources with different temporal and spatial coverage, the most accurate travel time with maximum spatial and temporal coverage would be achieved. The Niayesh tunnel in Tehran was selected as a case study, where extensive equipment for intelligent transportation systems is installed. In this study, considering the possibility of simultaneous access to multiple data sources at the same location, the following source, Google travel time data, Bluetooth travel time data and Inductive loop detectors, were fused. The improved travel time can increase the accuracy of travel time costs in transportation planning, information on variable message signs and routing software.

### Review History:

Received: Apr. 05, 2020

Revised: Jun. 19, 2020

Accepted: Jul. 14, 2020

Available Online: Aug. 22, 2020

### Keywords:

Data Fusion

Google Travel Time

Bayesian inference

Travel Time

Intelligent Transportation Systems

## 1. INTRODUCTION

In order to truly understand the state of the network, traffic control and evaluation of network performance, the collection and analysis of traffic data is a basic requirement. Good quality traffic information is essential to improve traffic models that contribute to short-term planning and regional policies. The proposal for this project is to combine information obtained from fixed & mobile sensors using data fusion techniques that can provide a complete picture of what we need. Data fusion (DF) is a set of techniques that uses statistical methods to combine information from different sources in order to reach a better inference. It is useful to use data fusion techniques to predict the situation as well as provide accurate network performance indicators because predicting these indicators plays an effective role in traffic. Van et al. [1] used a linear fitting model to predict travel time by combining Loop Detector data and floating machine data. They showed that linear fit on current, measurements of occupancy, travel time, and day of the week are useful for short-term forecasting of travel time. For long-term forecasts, however, the use of historical data will be more effective. Hillman et al. [2] also estimated travel time using the Kalman filter and travel time fusion from local sensors and electronic toll data. The results showed that travel time estimation using data fusion methods

is better than estimated travel time from each data source alone. Among the existing methods for data fusion, it can be concluded that Bayesian inference is a logical and valid method for data fusion. Studies have shown that more than two data sources have been used in only three studies. On the other hand, none of the studies used Google data, and in this research, Google travel time data has been used as a new and suggested source.

## 2. METHODOLOGY

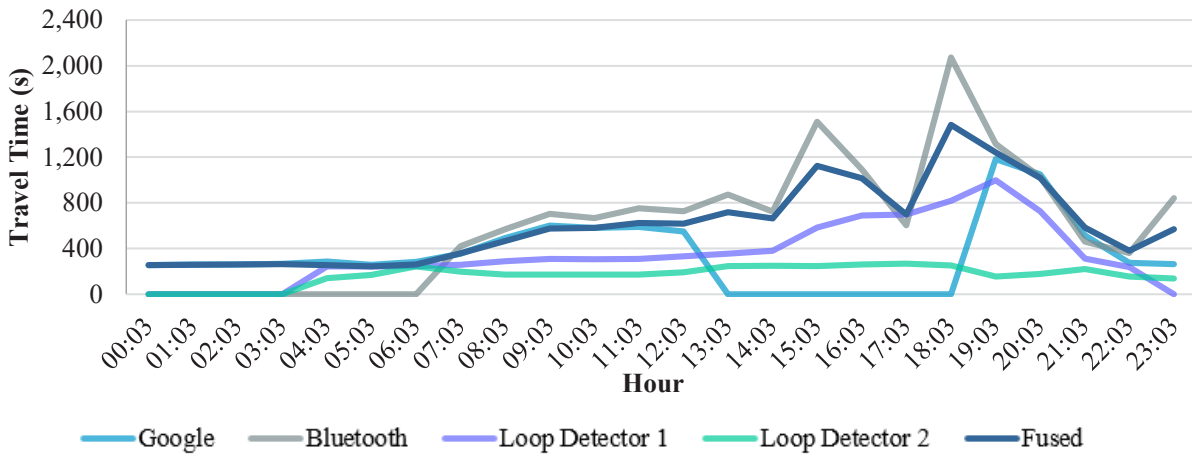
Bayesian inference is a probability-based method. This method is based on the principle that for each quantity, there is a probability distribution that by observing new data and reasoning about its probability distribution, optimal decisions can be made. This theorem is useful in that it can be used to calculate the probability of an event by making it conditional on the occurrence or non-occurrence of another event. In many cases, it isn't easy to calculate the probability of an incident directly. Using this theorem and conditioning one event on another, the required probability can be calculated. The main reason for choosing the Bayesian method in this article is that the Bayesian method has good accuracy despite its simplicity. One of the main advantages of this method is the use of the anterior distribution, which

\*Corresponding author's email: mmesbah@aut.ac.ir



**Table 1. Comparison of the square root of the errors in different sources**

Sensors	Bluetooth	Google	Loop Detector 1	Loop Detector 2
RMSE	212	335	602	614



**Fig. 1. Comparison chart of resource travel time**

due to the lack of sensor data in some time intervals, this model uses a pre-distribution to consider a value for that interval [3].

### 3. RESULTS AND DISCUSSION

Considering the coverage area of each intelligent system, the Niayesh tunnel was selected as a place where several sources of traffic information are available simultaneously for a case study. Among the available technologies, the following data for the north and south tunnels of Niayesh are the information recorded by the monitoring system via Google, Bluetooth and traffic loop detector. RMSE index is used to compare the results of the study with reality and is one of the most widely used indicators in assessing the accuracy of the results.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \tag{1}$$

In order to merge the data, the travel time accuracy of the sensors is calculated and the error of different sources relative to each other is shown in Table 1.

Finally, the travel time obtained by separate and fused sensors is Fig. 1:

### 4. CONCLUSION

The results show that first, fusion is easily available if data is available from two sources (Google and Bluetooth), three sources (Google, Bluetooth and Loop Identifier 1),

or four sources (Google, Bluetooth, Loop Detector 1 and Loop Detector 2). Second, the answers are close to the less error-prone Bluetooth source. Because Bluetooth is not available at all times and, for example, no Bluetooth data is available at midnight, this sensor has limited time coverage but full spatial coverage. On the other hand, because the Loop Detector is only available at the beginning and end of the tunnel, the information is only available at limited points along the route, and as a result, it has limited spatial coverage, but full temporal coverage. The combination of these two sensors with fusion models makes the final results have complete spatial and temporal coverage. The results of the hypothetical data show that the travel time output of the fusion models has higher accuracy than the travel time obtained from each of those sensors separately. This comparison has been investigated as the root mean square error (RMSE), and its results are shown in Table 1. Because fused travel time has higher accuracy than individual sensors, it can help to improve all travel time applications in the operation of the transportation network.

### REFERENCES

[1] Kwon, J., B. Coifman, and P. Bickel, *Day-to-day travel-time trends and travel-time prediction from loop-detector data*. Transportation Research Record, 2000. **1717**(1): p. 120-129.

[2] Heilmann, B., et al., *Predicting motorway traffic performance by data fusion of local sensor data and electronic toll collection data*. Computer-Aided Civil and Infrastructure Engineering, 2011. **26**(6): p. 451-463.

[3] Mahmoud Mesbah, A.G., Farzaneh Abdollahi, Fatemeh

Banani, Zahra Pakdaman, Afarin Kheirati, Mohsen Kheirandish & Amirhossein Alikhani, *Research plan for transportation systems data fusion Smart city of*

*Tehran to estimate travel time*. Amirkabir University of Technology (Tehran Polytechnic), 2016: p. 149.

**HOW TO CITE THIS ARTICLE**

M. Mesbah, Z. Pakdaman, M. Mollajafari, A. Golroo , *Fusion of travel time data in Niayesh tunnel using Bayesian inference*, Amirkabir J. Civil Eng., 53(10) (2022) 923-926.

DOI: [10.22060/ceej.2020.18214.6803](https://doi.org/10.22060/ceej.2020.18214.6803)







## استفاده از استنتاج بیز برای همجوشی داده‌های زمان سفر در تونل نیایش

سید محمود مصباح نمینی<sup>۱\*</sup>، زهرا پاکدامن<sup>۲</sup>، ماهان ملاجعفری<sup>۱</sup>، امیر گلرو<sup>۱</sup>

۱- دانشکده مهندسی عمران و محیط‌زیست، دانشگاه صنعتی امیرکبیر، تهران، ایران  
۲- دانشکده مهندسی عمران، دانشگاه علم و صنعت، تهران، ایران

### تاریخچه داوری:

دریافت: ۱۳۹۹/۰۱/۱۷

بازنگری: ۱۳۹۹/۰۳/۳۰

پذیرش: ۱۳۹۹/۰۴/۲۴

ارائه آنلاین: ۱۳۹۹/۰۶/۰۱

### کلمات کلیدی:

همجوشی داده‌ها

زمان سفر گوگل

استنتاج بیز

زمان سفر

سیستم‌های حمل و نقل هوشمند

**خلاصه:** با کاهش هزینه‌های جمع‌آوری داده، سیستم‌های حمل و نقل از سیستم‌های نیازمند به داده، به سیستم‌های نیازمند به تحلیل داده‌ها تبدیل شده‌اند. از آنجایی که دقت این داده‌ها متناسب با منابع جمع‌آوری آن متفاوت است، شناسایی داده با دقت بالاتر با ترکیبی از چند منبع، چالش اصلی کار با داده‌هایی از این دست می‌باشد. روش همجوشی داده‌ها می‌تواند با بهم پیوند دادن داده‌های حاصل از منابع مختلف دقت داده‌ها را در راستای هدف مورد مطالعه افزایش دهد. هدف اصلی از این مقاله، به دست آوردن دقیق‌ترین زمان سفر ممکن است. از جمله روش‌های همجوشی داده می‌توان به فیلتر کالمن، استنتاج بیز، شبکه‌های مصنوعی و تئوری دمپستر-شیفر اشاره کرد که در این تحقیق از استنتاج بیز استفاده شده و نتایج آن بررسی می‌شود. بر اساس این روش پیشنهاد شد که با ترکیب منابع مختلف داده که پوشش‌های زمانی و مکانی متفاوتی دارند، دقیق‌ترین زمان سفر با پوشش زمانی و مکانی حداکثر از طریق همجوشی داده به دست آید. مطالعه موردی تونل نیایش در شمال تهران انتخاب شد که در آن تجهیزات گسترده سیستم‌های حمل و نقل هوشمند نصب شده است. پس از تعیین دقت هر یک از منابع با استفاده از داده‌های خودروی شناور، زمان سفر همجوشی شده در یک روز که داده‌های همه منابع در آن موجود بود محاسبه گردید. در این تحقیق، با در نظر گرفتن امکان دسترسی همزمان به چند منبع داده در یک یا چند کمان مشترک، منابع داده‌های زمان سفر گوگل، داده‌های زمان سفر از حسگرهای بلوتوث و شناساگرهای حلقه در همجوشی مورد استفاده قرار گرفت. زمان سفر به دست آمده دقت هزینه‌های زمان سفر در برنامه‌ریزی حمل و نقل، اطلاعات تابلوهای پیام متغیر و نرم‌افزارهای مسیریاب را افزایش می‌دهد.

### ۱- مقدمه

و زمان سفر به منظور کاربردهای گسترده در بهره‌برداری، مورد نیاز است. اندازه‌گیری داده‌های به هنگام در سیستم پیشرفته اطلاع‌رسانی مسافران، که اطلاعات مورد نیاز کاربران راه را برای کمک به تصمیماتی مانند انتخاب مسیر و زمان حرکت فراهم می‌کند، بسیار ضروری است. این اطلاعات می‌تواند به شکل‌های متفاوتی مانند تابلوها، وب‌سایت‌ها، موبایل‌ها برای مسافران ارائه شود. علاوه بر این، گزارشات دقیق از الگوی روزانه سرعت ترافیک و زمان سفر برای تصمیم‌گیری در مورد سرمایه‌گذاری در زیر ساخت‌های حمل و نقل بسیار مورد نیاز است. در حال حاضر شاخص‌های عملکرد شبکه از جمله سرعت، زمان سفر، چگالی و تردد توسط حسگرهای مختلفی

به منظور شناخت واقعی از وضعیت شبکه، کنترل ترافیک و ارزیابی عملکرد شبکه، جمع‌آوری و تحلیل داده‌های ترافیکی نیازی اساسی است. داده‌های ترافیکی مانند ابزاری در اختیار اپراتورهای شبکه برای کمک به بهبود جریان ترافیک هستند. تبدیل این داده به اطلاعات قابل اطمینان، منجر به بهبود همه کنترل‌ها و تحلیل‌های شبکه خواهد شد. اطلاعات ترافیکی با کیفیت خوب، به منظور بهبود مدل‌های ترافیکی که به برنامه‌ریزی‌های کوتاه مدت و سیاست‌های منطقه‌ای کمک می‌کند، ضروری است. اندازه‌گیری سرعت ترافیک

\* نویسنده عهده‌دار مکاتبات: mmesbah@aut.ac.ir



همجوشی داده‌ها است که می‌تواند تصویر کاملی از آن چه نیاز داریم را فراهم کند. همجوشی داده‌ها (DF)<sup>۴</sup> مجموعه‌ای از تکنیک‌ها است که با استفاده از روش‌های آماری، اطلاعات را از منابع مختلف به منظور رسیدن به یک استنتاج بهتر، ترکیب می‌کند. ترکیب داده‌های چند منبع، ممکن است درک بهتری از وضعیت و موقعیت، با کاهش عدم اطمینان مربوط به هر منبع جداگانه تولید کند. استفاده از تکنیک‌های همجوشی داده به منظور پیش بینی وضعیت و همچنین ارائه دقیق شاخص‌های عملکرد شبکه به این دلیل که پیش‌بینی این شاخص‌ها نقش مؤثری در ترافیک دارند، مفید می‌باشد. سیستمی که تنها به یک حسگر وابسته باشد پایدار نیست، به این معنا که اگر آن حسگر نتواند عملکرد درستی داشته باشد، کل عملیات سیستم با شکست مواجه خواهد شد. روش‌های همجوشی داده‌ها می‌تواند این ضعف را برطرف کند و قابلیت اطمینان<sup>۵</sup> و استواری<sup>۶</sup> را افزایش دهد. ترکیب انواع مختلفی از حسگرها می‌تواند به سیستم، اطلاعات دقیق‌تری بدهد. ترکیب چند قرائت از همان حسگر باعث می‌شود یک سیستم حساسیت کمتری نسبت به نویزها و اشکالات موقت داشته باشد، بنابراین این روش می‌تواند به بهبود دقت<sup>۷</sup> و صحت کمک کند، همچنین می‌تواند با از بین بردن ابهام در داده‌ها، قطعیت<sup>۸</sup> را اضافه کند. در این روش از داده‌های یک حسگر که می‌تواند به اطلاعاتی که در دیگر حسگرها در دسترس نیست، پرت است، یا مشابه است کمک کند که این می‌تواند باعث افزایش استواری<sup>۹</sup> شود. در مواقعی که یک حسگر نمی‌تواند در شرایط خاص زمانی و مکانی، اطلاعات ترافیکی را اختیار اپراتورها قرار دهد، می‌توان با استفاده از حسگر دیگر به این اطلاعات دست یافت یا به عبارتی باعث افزایش پوشش مکانی و زمانی می‌شود. همچنین با مجموعه‌ای از حسگرهای حد وسط (معمولی) می‌توان با کاهش هزینه‌ها<sup>۱۰</sup> به همان سطح از عملکردی دست یافت که یک حسگر با قابلیت اعتماد بالا می‌رسد [۲]. در ادامه توضیحاتی درباره تاریخچه این روش‌ها در بخش دوم، توضیحاتی درباره روش پیشنهاد و انتخاب شده در بخش سوم، مطالعه موردی در بخش چهارم، بررسی محاسبات و نتایج در بخش پنجم و نتیجه‌گیری را در

به دست می‌آیند که هر کدام از این حسگرها به منظوری خاص به کار گرفته شده‌اند، اما ممکن است در بخشی از بزرگراه چندین حسگر وجود داشته باشد که علاوه بر برآورده کردن نیاز اصلی که به منظور آن در بزرگراه نصب شده‌اند، شاخص‌های دیگری از عملکرد شبکه را نیز، در اختیار قرار دهند و در نتیجه این فرصت فراهم شود که چندین داده از حسگرهای مختلف برای همجوشی، در دسترس قرار گیرد. همچنین داده‌های ترافیکی به وسیله حسگرهای مختلفی جمع‌آوری می‌شوند که هر کدام از این حسگرها دارای مزایا و معایبی هستند. به طور مثال شناساگرهای حلقه که اندازه‌گیری‌های مربوط به شمارش خودروها، اشغال خط و سرعت خودروها را ارائه می‌دهند دارای دقت بالایی هستند و همچنین تمامی خوردوهای که از آن مقطع راه سفر می‌کنند را پوشش می‌دهند. در نتیجه دارای پوشش زمانی<sup>۱</sup> بالایی هستند. اما به دلیل اینکه در یک نقطه ثابت هستند و جریان سرعت ترافیک را تنها در محل حسگر، نه در کل راه بیان می‌کنند (شناساگرهای حلقه سرعت لحظه‌ای را اندازه‌گیری می‌کنند)، دارای پوشش مکانی<sup>۲</sup> محدودی می‌باشند. از طرفی دیگر، تکنولوژی‌های ارتباطی بی‌سیم مانند جی‌پی‌اس<sup>۳</sup>، وای‌فای و کاوشگر تلفن همراه به منظور جمع‌آوری اطلاعات ترافیکی به طور فزاینده‌ای در سیستم‌های حمل و نقل هوشمند به کار گرفته می‌شوند. این حسگرها حالت ترافیک را در طول قابل توجهی از راه بیان می‌کنند و در نتیجه دارای پوشش مکانی بالایی هستند اما به دلیل اینکه آن‌ها تنها درصد کوچکی از وسایل نقلیه‌ای هستند که شبکه ترافیکی را تشکیل می‌دهند، جامع نمی‌باشند و نیز پوشش زمانی کمی دارند، در نتیجه هر کدام از این حسگرها از طریق دیگری تکمیل می‌شود و همجوشی این دو حسگر می‌تواند محدودیت‌های موجود در هر حسگر جداگانه را کاهش دهد [۱]. علاوه بر مزیت فوق در همجوشی داده حسگرها، هر کدام از این حسگرها ممکن است در شرایط خاص زمانی و مکانی اطلاعات ترافیکی را با دقت پایینی در اختیار اپراتورها قرار دهند. در نتیجه تعیین شاخص عملکرد شبکه با استفاده از داده‌های به دست آمده از چند حسگر موضوعی است که سال‌های اخیر بسیار مورد توجه قرار گرفته است. پیشنهاد این پروژه ترکیب اطلاعات به دست آمده از حسگرهای ثابت و موبایل با استفاده از تکنیک‌های

4 Data Fusion  
5 Reliability  
6 Robustness  
7 Accuracy  
8 Certainty  
9 Robustness  
10 Cost Effectiveness

1 Temporal coverage  
2 Spatial coverage  
3 Global Positioning System

بخش ششم خواهیم داشت.

## ۲- مروری بر ادبیات

روش های متفاوتی برای همجوشی داده وجود دارد. این روش ها از میانگین گیری حسابی تا به کارگیری مدل های پیچیده تر همجوشی داده تغییر می کنند. ون و همکاران [۳] از مدل برازش<sup>۱</sup> خطی برای پیش بینی زمان سفر با ترکیب داده های شناساگر حلقه و ماشین شناور بهره گرفتند. آن ها نشان دادند که برازش خطی بر روی جریان کنونی، اندازه گیری های میزان اشغال، زمان حرکت و روز هفته برای پیش بینی کوتاه مدت زمان سفر مفید است. حال آن که برای پیش بینی های بلند مدت استفاده از داده های تاریخی مؤثرتر خواهد بود. ژانگ و رایس [۴] از یک مدل خطی با ضرایب متغیر به منظور پیش بینی زمان سفر بزرگراه ها از طریق همجوشی داده های شناساگر حلقه و ماشین شناور استفاده کردند. این ضرایب به صورت تابعی از زمان حرکت تغییر می کردند. در ضمن تخمین این ضرایب به صورت آفلاین بوده و پس از ذخیره سازی آن ها، مدل قادر به پیش بینی در لحظه بوده است. الفوزی و همکاران [۵] مدلی را که بر پایه روش دمپستر-شیفر ساخته شده بود به کار گرفتند. آن ها از زمان سفرهای حاصل از شناساگر حلقه و داده های منتج از جمع آوری عوارض برای تخمین زمان سفر استفاده کردند. این مدل به احتمال آن که هر یک از منابع داده، اطلاعات صحیحی را گزارش دهند نیاز داشت. هیلمن و همکاران [۶] همچنین با استفاده از فیلتر کلمن و همجوشی زمان سفرهای حاصل از حسگرهای محلی و داده های الکترونیکی عوارض به تخمین زمان سفر پرداختند. نتایج نشان دادند که تخمین زمان سفر با به کارگیری روش های همجوشی داده بهتر از زمان سفرهای تخمینی حاصل از هر منبع داده به تنهایی است. ایوان [۷] با به کارگیری شبکه عصبی مصنوعی به تشخیص حوادث ترافیکی در راه های شریانی چراغ دار از طریق همجوشی زمان سفرهای حاصل از شناساگر حلقه شبیه سازی شده و ماشین شناور پرداخت. مدل تحلیلی ساده دیگری که از داده های شمارش جریان در بالادست و پایین دست یک قطعه از مسیر به منظور تخمین زمان سفر آن قطعه استفاده می کند، روش شمارش تجمعی (ورودی - خروجی) [۸] است. اشکال عمده این روش وابستگی آن به دقت شمارش جریان

برای تخمین زمان سفر است [۹]. برخی از روش های گزارش شده دیگر نیز پایه های تئوری جریان ترافیک را شامل می شوند [۱۰ و ۸].

چو و رکر [۱۱] از ترکیب داده های شناساگر حلقه شبیه سازی شده و ماشین شناور برای تخمین زمان سفر با استفاده از مدلی بر اساس فیلتر کالمن بهره گرفتند. نتایج نشان داد استفاده از چند منبع برای تخمین زمان سفر بهبود قابل توجهی نسبت به استفاده از یک منبع داشته است. پارک و لی [۱۲] تخمینی از زمان سفر در مناطق شهری به وسیله دو روش شبکه عصبی و استنتاج بیز به صورت جداگانه انجام دادند و در این تخمین از داده های ماشین شناور و حلقه های القایی استفاده کردند. ژو و همکاران [۱۳] یک مدل همجوشی داده بر اساس شبکه عصبی، با استفاده از داده های ماشین شناور و شناساگر حلقه پیشنهاد دادند. این مدل توسعه داده شد و با داده های شهر گوانگژو اعتبارسنجی گردید، نتایج نشان داد که مدل دارای دقت مناسبی است. در پژوهش پنگ و همکاران [۱۴] بر روی داده های شهر هونگژوای چین<sup>۲</sup> با استفاده از فیلتر کالمن و تکنیک همجوشی، قادر به حل مشکل نظارت بر بخشی از مسیر که بدون داده های ترافیکی<sup>۳</sup> GPS بود، شدند. نتایج همجوشی برای متغیر مورد نظر با داده های تاریخچه ای مقایسه شد که خطای نسبی<sup>۴</sup> حاصل ۱۹ درصد شد. این در حالی است که میانگین خطای نسبی داده های تاریخچه ای برای متغیر مورد نظر ۴۴ درصد است. نتایج بیون و همکاران [۱۵] نشان داد همجوشی داده های کم کیفیت شناساگرهای حلقه در حضور حسگر بهتر تنها منجر به افزایش خطا می شود. سوریگورا و همکارش مدلی را برای پیش بینی کوتاه مدت زمان سفر در بزرگراه بیان کردند. مدل ارائه شده تخمین مناسبی از زمان سفر مورد انتظار رانندگان بود. آن ها در مرحله اول با استفاده از همجوشی دو زمان سفر که از روش های تئوری به وسیله شناساگر حلقه به دست آمده اند، به یک زمان سفر با پوشش زمانی و مکانی کامل رسیدند و در مرحله دوم همجوشی، زمان سفر مرحله اول را با زمان سفر به دست آمده از درگاه اخذ عوارض<sup>۵</sup>، برای رسیدن به یک زمان سفر پیش بینی شده ترکیب کردند که نتایج به دست آمده دارای دقت مناسبی بود [۱۶].

لی و یانگ یک مدل همجوشی داده ارائه کردند که از مدل میانگین زمان سفر کمان بر اساس جمعیت وسایل نقلیه با ترکیب داده های

2 Hangzhou in China

3 Global Positioning System

4 The Relative Mean Error (RME)

5 Toll ticket

1 Regression



جدول ۱. مروری بر ادبیات همجوشی داده‌ها در تخمین زمان سفر

Table 1. A review of the data fusion literature in estimating travel time

روش استفاده شده		داده‌های استفاده شده							سال	پژوهشگر			
داده‌های GPS	ماشین شناور	داده‌های شناساگر حلقه	گیت‌های جمع آوری عوارض	بلوتوث	پلاک ماشین	الگوریتم‌های وزن دهی	فیلتر کالمن	منطق فازی			شبکه عصبی	استنتاج بیز	دمپستر-شیفر
		۱						۱				۱۹۹۹	Palacharla and Nelson
						۱						۲۰۰۰	El Faouzi & Simon
					۱							۲۰۰۴	Park & Lee
					۱							۲۰۰۵	El Faouzi
									۱			۲۰۰۵	van Lint et al
								۱	۱			۲۰۰۶	Sazi-Murat
									۱			۲۰۰۶	zho & zhu
											۱	۲۰۰۹	El Faouzi
												۲۰۰۹	Peng et al
												۲۰۰۹	Guo et al
												۲۰۱۰	Byon
									۱	۱		۲۰۱۰	Chung & Choi
									۱	۱		۲۰۱۱	& Bachmann Christian
												۲۰۱۲	Anusha et al
									۱			۲۰۱۳	Huiping
												۲۰۱۴	zheng et al
												۲۰۱۴	Chang et al
												۲۰۱۴	Xia et al
												۲۰۱۵	Zhang et al
												۲۰۱۶	Liu et al
												۲۰۲۰	Guo & Yang
												۲۰۱۸	Mil et al

را برای همجوشی داده‌ها پیشنهاد دادند که در این روش ابتدا داده‌های شناساگر حلقه و خودرو شناور به وسیله مدل بیز، برای تخمین زمان سفر ترکیب می‌شوند. گو و یانگ در مقاله دیگر یک روش تخمین معتبر از زمان سفر در مسیرهای کلان شهری به کمک تلفیق کردن داده‌های چند حسگر و استفاده از روش همجوشی داده‌ها پرداختند. آن‌ها مطالعه موردی خود را برای جاده جینگشی در جینان چین انجام دادند و نشان دادند همجوشی داده‌ها به طور موثری دقت و اطمینان از

شناساگرهای ثابت و ماشین‌های شناور به دست آمده بود. بچمن و همکاران [۱] با استفاده از داده‌های شناساگر حلقه و خودرو شناور به تخمین دقیق سرعت ترافیک آزادراه پرداختند و با بررسی چندین تکنیک همجوشی از جمله فیلتر کالمن، منطق فازی، شبکه عصبی و وزن‌دهی به این نتیجه رسیدند که استفاده از هر تکنیک بستگی به مواردی همچون درصد خودرو شناور و شرایط ترافیکی (اشباع، نیمه اشباع، جریان با سرعت آزاد) دارد. لیو و همکاران [۱۷] روشی



استنتاج بیز یک فرآیند یادگیری به وسیله داده‌ها می‌باشد. در این روش با استفاده از داده‌های مشاهده شده  $t_1$ ، پارامتر  $\mu$  آموزش داده شده است. قبل از مشاهده داده‌ها اطلاعات به صورت احتمال پیشین<sup>۲</sup> است و بعد از مشاهده داده‌ها به صورت احتمال پسین<sup>۳</sup> می‌باشد و در نتیجه باور و دانش ما درباره اطلاعات پیشین در این فرآیند یادگیری، ارتقاء می‌یابد. در این فرمول توزیع پیشین را با  $P(\mu)$  و توزیع پسین با  $P(\mu|t_1)$  نشان داده شده است. برای به دست آوردن این توابع توزیع باید در ابتدا بررسی شود که متغیرها از نوع گسسته هستند و یا پیوسته می‌باشند. از آنجا که متغیرهای مورد بررسی در این پروژه از به دلیل آن که متغیرها از جنس زمان سفر هستند از نوع پیوسته می‌باشند، از تابع چگالی احتمال استفاده شده است. استنتاج بیز، احتمال پسین را از ماحصل احتمال پیشین و تابع درست‌نمایی<sup>۴</sup> برگرفته از یک مدل آماری روی داده‌های مشاهده شده، محاسبه می‌کند. فرمول کلی استنتاج بیز به صورت رابطه (۱) است:

$$P(\mu|t_1) = \frac{P(t_1|\mu) P(\mu)}{P(t_1)} \quad (1)$$

$\mu$  = زمان سفر همجوشی شده از دو منبع

$t_1$  = زمان سفر به دست آمده از منبع ۱

$P(\mu)$  توزیع پیشین است و پیش از آنکه مشاهده صورت گیرد با استفاده از آگاهی قبلی (تجربه) وضع را مشخص می‌کند و باور مدل‌ساز از واقعیت را بیان می‌کند، که این توزیع معادل اطلاعات موجود در نمونه‌ی دیگر (غالباً مرسوم به نمونه "خیالی" یا "شبه نمونه") است. باور پیشین در مورد یک پارامتر، ممکن است نقش اساسی در برآورد کردن آن‌ها بازی کند، در واقع روش بیز وسیله‌ای است برای آنکه یک چنین اطلاعات پیشینی رسماً به حساب آورده شود. همچنین اگر تعداد مشاهدات کافی وجود داشته باشد، انتخاب توزیع پیشین چندان مهم نیست و می‌توان به یک توزیع پسین با دقت مناسب دست یافت.  $P(t_1|\mu)$  تابع درست‌نمایی است که مبتنی بر اطلاعات نمونه می‌باشد و به صورت یک تابع شرطی بیان می‌شود.  $P(\mu|t_1)$  تابع پسین است که بعد از مشاهده اطلاعات به دست می‌آید. این تابع توزیع احتمال با استفاده از ترکیب اطلاعات

برآورد زمان سفر را در سیستم‌های جاده‌های شهری بهبود می‌بخشد [۱۹]. در تحقیقی دیگر میل و همکارش [۲۰] چارچوبی برای توسعه مدل تخمین زمان سفر با استفاده از چندین منبع به کمک همجوشی داده‌ها پرداخته است. آن‌ها با ترکیب کردن روش بیز و مدل ترکیبی گاوسی برای همجوشی داده‌های زمان سفر جهت بهبود دقت و تکمیل داده‌ها از نظر مکانی و زمانی پرداختند. نتایج نشان داد آن‌ها توانستند بهبود معناداری در برآورد زمان سفر از نظر میانگین خطای درصد مطلق (MAPE) در محدوده ۳/۱۶ درصد تا ۴۶/۳ درصد را به وجود بیاورند. جدول ۱ مروری بر ادبیات گذشته در ارتباط با روش‌های اصلی همجوشی داده‌ها در مبحث تخمین زمان سفر را نشان می‌دهد. وجود داده‌های تاریخیچه‌ای و منابع مناسب مانند شناساگر حلقه، گوگل، ماشین شناور و بلوتوث و استفاده از روش فیلتر کالمن و استنتاج بیز<sup>۱</sup> در همجوشی پارامترهای ترافیکی به منظور محاسبه زمان سفر دقیق‌تر در دهه اخیر افزایش یافته است. همچنین روش‌هایی مانند شبکه عصبی در صورت وجود داده‌های تاریخیچه‌ای کافی، کاربرد مناسبی خواهند داشت. می‌توان نتیجه گرفت که استنتاج بیز روشی منطقی و معتبر برای همجوشی داده‌ها می‌باشد. همانطور که در جدول ۱ مشاهده می‌شود تنها در سه پژوهش است که از بیش از دو منبع داده استفاده شده است. از طرفی در هیچ‌کدام از بررسی‌های صورت گرفته در جدول ۱ از داده‌های گوگل استفاده نشده است که در این تحقیق از داده‌های زمان سفر گوگل به عنوان یک منبع جدید و پیشنهادی نیز استفاده شده است.

### ۳- روش پیشنهادی

استنتاج بیز روشی بر پایه احتمالات است. این روش بر این اصل استوار است که برای هر کمیتی یک توزیع احتمال وجود دارد که با مشاهده یک داده جدید و استدلال در مورد توزیع احتمال آن می‌توان تصمیمات بهینه‌ای اتخاذ کرد. این قضیه از آن جهت مفید است که می‌توان از طریق آن، احتمال یک پیشامد را با مشروط کردن نسبت به وقوع و یا عدم وقوع یک پیشامد دیگر محاسبه کرد. در بسیاری از حالت‌ها، محاسبه احتمال یک پیشامد به صورت مستقیم کاری دشوار است. با استفاده از این قضیه و مشروط کردن پیشامد مورد نظر نسبت به پیشامد دیگر، می‌توان احتمال مورد نظر را محاسبه کرد.

2 Prior Probability  
3 Posterior Probability  
4 Likelihood Function

1 Bayesian Inference

توزیع پیشین و یک تابع درست‌نمایی حاصل می‌شود که در نتیجه این توزیع دارای دقت بالاتری می‌باشد و یک محدوده باریک‌تری برای مقادیر  $\mu$  را نسبت به هر کدام از توزیع‌های پیشین و درست‌نمایی در برمی‌گیرد. در همجوشی داده‌ها، از چندین منبع برای رسیدن به یک استنتاج با دقت بالاتر استفاده می‌شود. در این حالت هر کدام از منابع برای رسیدن به یک استنتاج دقیق‌تر، اطلاعاتی را در اختیار ما قرار می‌دهند که به عنوان یک تابع درست‌نمایی در روش بیز به شمار می‌آید؛ در نتیجه به ازای هر منبع باید از یک تابع درست‌نمایی در فرمول بیز استفاده شود. برای به دست آوردن این احتمال شرطی باید از یک جدول یادگیری استفاده کرد. این جدول به داده‌های واقعی در چندین روز که دوره آفلاین یادگیری نامیده می‌شود، نیاز دارد و هرچه تعداد این روزها بیشتر باشد، این جدول دقیق‌تر می‌شود. به دلیل دسترسی نداشتن به داده‌های واقعی گسترده نمی‌توان از این جدول استفاده کرد، بنابراین نیاز است که یک توزیع برای خطای داده‌هایی که توسط هر کدام از منابع به دست می‌آید را در نظر گرفت. در شرایطی که اطلاعاتی از توزیع یک پارامتر وجود ندارد؛ بهترین توزیعی که می‌شود برای آن فرض کرد توزیع نرمال می‌باشد. همچنین توزیع پیشین را به صورت نرمال در نظر گرفته و در نهایت از حاصل ضرب سه توزیع نرمال، یک توزیع برای تابع پسین به دست می‌آید که توزیع حاصله نیز نرمال است [۲۱]. فرمول بیز در همجوشی به صورت رابطه (۲) است:

$$P(\mu|t_1, t_2) = \frac{P(\mu) P(t_1|\mu) P(t_2|\mu)}{P(t_1) P(t_2)} \quad (2)$$

$t_2$  = زمان سفر به دست آمده از منبع ۲

فرض نرمال بودن تمامی توزیع‌ها به صورت روابط (۳)، (۴) و (۵)

است:

$$P(\mu) = K_1 e^{-\frac{1}{2} \frac{(\mu-\mu_0)^2}{\sigma_0^2}} \quad (3)$$

$$P(t_1|\mu) = K_2 e^{-\frac{1}{2} \frac{(t_1-\mu)^2}{\sigma_1^2}} \quad (4)$$

$$P(t_2|\mu) = K_3 e^{-\frac{1}{2} \frac{(t_2-\mu)^2}{\sigma_2^2}} \quad (5)$$

$\sigma_0$  و  $\sigma_1$  و  $\sigma_2$  = به ترتیب خطای توزیع پیشین، خطای منبع اول و خطای منبع دوم

که در روابط بالا مقادیر  $K_1$ ،  $K_2$  و  $K_3$  عبارتند از:

$$K_1 = \frac{1}{\sqrt{2\pi} \times \sigma_0} \quad K_2 = \frac{1}{\sqrt{2\pi} \times \sigma_1}$$

$$K_3 = \frac{1}{\sqrt{2\pi} \times \sigma_2}$$

از ضرب این سه توزیع در یکدیگر:

$$P(\mu|t_1, t_2) = K_1 K_2 K_3 e^{-\frac{1}{2} \frac{(\mu-\mu_0)^2}{\sigma_0^2}} e^{-\frac{1}{2} \frac{(t_1-\mu)^2}{\sigma_1^2}} e^{-\frac{1}{2} \frac{(t_2-\mu)^2}{\sigma_2^2}} = K_1 K_2 K_3 e^g$$

$$g = -\frac{1}{2} \left( \frac{(\mu-\mu_0)^2}{\sigma_0^2} + \frac{(t_1-\mu)^2}{\sigma_1^2} + \frac{(t_2-\mu)^2}{\sigma_2^2} \right) = -\frac{1}{2} \left[ \mu^2 \left( \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - 2\mu \left( \frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2} \right) + K_4 \right]$$

$$= -\frac{1}{2} \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \left[ \mu^2 - 2\mu \left( \frac{\frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right) + K_5 \right] = -\frac{1}{2} \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \left[ \mu - \left( \frac{\frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right) \right]^2$$

$K_4$  و  $K_5$  نیز عبارتند از:

$$K_4 = \frac{\mu_0^2}{\sigma_0^2} + \frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2}$$

$$K_5 = \frac{K_4}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

که در نهایت توزیع پسین، توزیعی نرمال با میانگین و واریانس

زیر می‌باشد:

$$\text{واریانس} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \quad (7)$$

$$\text{میانگین} = \left( \frac{\frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right) \quad (8)$$

مستقیم در بازه های زمانی مشخص و در دو حالت "در لحظه" یا "پیش بینی" زمان سفر را گزارش می کند. داده های گوگل از طریق کدنویسی و اجرای برنامه در بازه زمانی چند ماهه جمع آوری شدند. این داده ها به صورت لحظه ای هر ۳ دقیقه و به صورت پیش بینی شده هر ۱۵ دقیقه در حالت های ۱۵ دقیقه، ۳۰ دقیقه، ۱ ساعت و ۲ ساعت قبل از وقوع هر بازه برداشت شده اند. برای جمع آوری این داده ها و کاهش داده های گم شده دو سیستم از آزمایشگاه هوش مصنوعی دانشکده عمران دانشگاه صنعتی امیرکبیر به صورت شبانه روزی و به مدت ۳ ماه در حال برداشت اطلاعات لحظه ای از گوگل بودند. گوگل با استفاده از برنامه مخصوص نوشته شده، مانند شکل a در پیوست برداشت شده است که به صورت یک ردیف داده پس از مرتب سازی به صورت شکل b در پیوست نشان داده می شود. یکی از کاربردهای اصلی سامانه نظارت بر بلوتوث، محاسبه زمان سفر خودروها می باشد. برای محاسبه زمان سفر، داده های ارسالی توسط حسگرهای موجود در مسیر با هم تطبیق داده شده و با به دست آوردن فاصله زمانی که بلوتوث یک خودرو توسط دو حسگر متوالی عبور کرده است، زمان سفر خودروی مورد نظر محاسبه می شود. به ازای هر داده تطبیق داده شده، یک داده زمان سفر تولید می گردد. محل قرارگیری حسگرهای بلوتوث در ابتدای تونل نیایش (تقاطع صدر) و انتهای تونل نیایش (تقاطع بزرگراه شهید چمران) است. شرکت کنترل ترافیک با تطبیق کد شناسه بلوتوث یک خودرو در دو حسگر متوالی زمان سفر خودرو مورد نظر را محاسبه می کند. (بلوتوث خودرو هنگام عبور از ابتدای تونل نیایش یک بار شناسایی می شود و بار دوم هنگام عبور از انتهای تونل نیایش (تقاطع بزرگراه شهید چمران) شناسایی می شود. با محاسبه اختلاف زمان شناسایی خودرو مورد نظر هنگام عبور از این دو نقطه زمان سفر این قطعه محاسبه می شود). بلوتوث می تواند بر روی وسایلی مانند تلفن همراه، ضبط صوت، هدفون و موارد دیگر وجود داشته باشد. همانطور که گفته شد مسئله مهم شناسایی بلوتوث در ورود و خروج تونل و بررسی کد شناسایی (ID) که برای هر بلوتوث توسط حسگرها برداشت می شود، می باشد. شناساگر حلقه بر اساس محل قرارگیری در کف مسیر سرعت و تعداد خودروهای عبوری در هر بازه زمانی مشخص از دو مقطع ابتدایی و انتهایی تونل را ثبت می کنند. با استفاده از این دو داده و با توجه به طول تونل، زمان سفر خودروهای داخل تونل محاسبه می گردد. اکثر داده های

بنابراین با داشتن میانگین و واریانس توزیع پیشین، خطای منبع ۱ و خطای منبع ۲ می توان توزیع پسین را به صورت توزیعی نرمال با میانگین رابطه (۸) و واریانس معادله (۷) محاسبه کرد. در صورت وجود بیش از دو منبع به سادگی می توان نشان داد، بسط معادلات فوق برای n منبع منجر به یک توزیع پسین نرمال با میانگین و واریانس روابط (۱۰ و ۱۱) خواهد بود. برتری روابط زیر در آن است که بدون نیاز به روش های عددی توزیع پسین به صورت تحلیلی قابل محاسبه است.

$$P(\mu|t_1, t_2, \dots, t_n) = \frac{P(\mu)P(t_1|\mu)P(t_2|\mu)\dots P(t_n|\mu)}{P(t_1)P(t_2)\dots P(t_n)} \quad (9)$$

$$\text{واریانس} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_n^2}} \quad (10)$$

$$\text{میانگین} = \left( \frac{\frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2} + \dots + \frac{t_n}{\sigma_n^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \dots + \frac{1}{\sigma_n^2}} \right) \quad (11)$$

علت اصلی انتخاب روش بیز در این مقاله این است که روش بیز وجود سادگی دارای دقت مناسبی است. یکی از مزیت های اصلی این روش به کارگیری توزیع پیشین می باشد که با توجه به موجود نبودن داده های حسگرها در بعضی از بازه های زمانی، این مدل با استفاده از توزیع پیشین مقداری برای آن بازه در نظر می گیرد [۱۸].

#### ۴- مطالعه موردی

با در نظر گرفتن حوزه پوشش هر سامانه هوشمند تونل نیایش به عنوان محلی که در آن چند منبع اطلاعات ترافیکی به صورت همزمان در دسترس هستند، جهت انجام مطالعه موردی انتخاب گردید. از میان فناوری های موجود، داده های زیر برای تونل شمالی و جنوبی نیایش، اطلاعات ثبت شده توسط سامانه نظارت از طریق گوگل، بلوتوث و شناساگر حلقه بر ترافیک است. گوگل به صورت

شناساگر حلقه دریافتی شامل تعداد خودروهای عبوری و سرعت آن‌ها در مقاطع ورودی و خروجی تونل می‌باشند. به منظور محاسبه خطای زمان سفر شناساگر حلقه طرح ماشین شناوری در سه روز و در تاریخ‌های ۹۵/۷/۱۷، ۹۵/۷/۱۸ و ۹۵/۸/۲۴ انجام شده و زمان سفرهای واقعی تونل شمالی و تونل جنوبی (هر دو از ابتدای تونل تا انتهای آن) برداشت شده‌اند. لازم به ذکر است که دلیل انتخاب این نقاط (ابتدای تونل و انتهای آن) برای اندازه‌گیری زمان سفر واقعی بین آن‌ها انطباق این نقاط با محل قرارگیری شناساگر حلقه است. اما از آنجا که داده‌های شناساگر حلقه موجود متعلق به تونل جنوبی می‌باشند، خطا تنها برای تونل جنوبی محاسبه شده است. همچنین با توجه به آن که داده‌های شناساگر حلقه و در نتیجه زمان سفرهای حاصل از این داده‌ها تنها در بازه‌های یک ساعته موجود هستند، زمان سفرهای واقعی برداشت شده نیز در بازه‌های یک ساعته میانگین‌گیری شده و به عنوان زمان سفر ابتدای آن بازه در نظر گرفته شده‌اند.

داده‌های بلوتوث و شناساگر حلقه از شرکت کنترل ترافیک دریافت شده است. داده‌های گوگل به دلیل این که آنلاین برداشت می‌شدند و برنامه مخصوصی برای آن نوشته شده بود، نیاز به پاکسازی نداشت. تنها در مواردی به علت قطعی برق و یا اینترنت، داده گوگل برداشت نشده بود که در برنامه همجوشی این موارد در نظر گرفته شد. در مورد پاکسازی داده‌های بلوتوث ستونی مبنی بر وضعیت رکورد موجود بود که قبول یا رد رکورد به صورت عددی در آن صورت مشخص شده است. بدین منظور با فیلتر کردن آن ستون داده‌های غیرقابل قبول حذف می‌شوند. همچنین به منظور حذف داده‌های پرت با استفاده از نمودار جعبه و آماره Z داده‌ها با کمک نرم‌افزار SPSS تحلیل و پاکسازی شدند. در فایل‌هایی که سرعت میانگین وجود دارد، از این سرعت برای ادامه محاسبات استفاده می‌شود. در فایل‌هایی که این سرعت وجود ندارد و بازه‌بندی سرعت وجود دارد، با ضرب تعداد خودروها در میانگین بازه سرعت و در نهایت میانگین‌گیری وزنی، سرعت در هر زمان مشخص می‌شود. در بعضی فایل‌ها تعداد خودرو با سرعت بالای ۱۳۰ وجود دارد که برای میانگین‌گیری سرعت، این سرعت لحاظ نشده ولی برای استفاده از جریان تعداد کل خودروهای عبوری لحاظ شده است. نکته حائز اهمیت این است که داده‌های شناساگر حلقه در بازه‌های یک ساعته داده شده بود در حالی که داده‌های گوگل و بلوتوث که

در ادامه می‌آید در بازه‌های ۱۵ دقیقه‌ای (و حتی یک دقیقه‌ای) هم موجود بود. بنابراین در محاسبات خطای شناساگر حلقه و محاسبات همجوشی از بازه‌های یک ساعته استفاده شده است. به منظور محاسبه خطای منابع طرح ماشین شناوری انجام شده است که زمان سفرهای منتج از این طرح به عنوان زمان سفر واقعی در نظر گرفته شدند. محاسبه خطای زمان سفر منابع در دو حالت هم‌مفزون شده ۱۵ دقیقه‌ای و یک ساعته انجام شد. علت بررسی زمان سفرهای هم‌مفزون شده ۱۵ دقیقه‌ای آن است که زمان سفرهای گوگل و بلوتوث هر ۱۵ دقیقه یکبار ثبت شده‌اند. زمان سفرهای هم‌مفزون شده یک ساعته نیز به دلیل ثبت داده‌های شناساگر حلقه در بازه‌های یک ساعته بررسی شدند. با توجه به نتایج محاسبه خطای منابع زمان سفر لحظه‌ای میانگین هم‌مفزون شده ۱۵ دقیقه‌ای به دست آمده از گوگل به علت بیشترین میزان دقت نسبت به سایر زمان سفرهای گوگل به عنوان ورودی گوگل به مدل‌های همجوشی انتخاب شده است. خطای این زمان سفر در دو مسیر به صورت میانگین (۳۶/۵۶٪) ۳۷۰ ثانیه می‌باشد. خطای زمان سفرهای هم‌مفزون شده ۱۵ دقیقه‌ای شناساگر حلقه، بلوتوث و گوگل در جدول ۲ آمده است که در آن حسگر بلوتوث بالاترین دقت را دارد و پس از آن منبع گوگل و شناساگر حلقه قرار دارند. این سه دسته زمان سفر ورودی مدل‌های همجوشی نیز می‌باشند. خطای زمان سفرهای هم‌مفزون شده یک ساعته شناساگر حلقه، بلوتوث و گوگل نیز مطابق جدول ۲ است. لازم به ذکر است که خطای حسگرهای "شناساگر حلقه ۱ و ۲ و بلوتوث در روزهای ۱۷ام و ۱۸ام و ۲۴ام آبان به دلیل موجود بودن داده‌های ماشین شناور محاسبه شد اما خطای گوگل به دلیل عدم دسترسی به داده‌های آن در روز ۱۷ ام و ۱۸ام (قطعی برق و کار نکردن سیستم آزمایشگاه) تنها با استفاده از داده‌های ماشین شناور موجود در روز ۲۴ام محاسبه شد. در شکل ۱ نقشه محل حسگر آمده است:

#### ۵-نتایج

همان‌طور که در بخش سوم بیان شد این روش به سادگی قابل تعمیم برای هر تعداد منبع می‌باشد. در نتیجه از حاصل ضرب ۵ توزیع نرمال که ۴ توزیع مربوط به خطای منابع و دیگری مربوط به توزیع پیشین می‌باشد، می‌توان به یک توزیع نرمال برای تابع پسین دست یافت. فرمول بیز در همجوشی برای چهار منبع به صورت رابطه



شکل ۱. محل قرارگیری حسگرها در طول تونل نیایش  
 Fig. 1. Location of sensors along the Niayesh Tunnel

جدول ۲. خطای زمان سفر منابع نسبت به زمان سفر واقعی در تونل جنوبی

Table 2. Travel time resources' error relative to actual travel time in the southern tunnel

میانگین ۱ ساعته	میانگین ۱۵ دقیقه‌ای	منبع زمان سفر
۶۰۲ (/۴۰/۷۰)	۵۷۹ (/۳۲/۷۹)	شناساگر حلقه ۱
۶۱۴ (/۴۹/۵۷)	۵۷۵ (/۳۴/۰۹)	شناساگر حلقه ۲
۲۱۲ (/۲۱/۸۴)	۲۵۳ (/۱۹/۱۶)	بلوتوث
۳۳۵ (/۳۵/۸۸)	۳۷۰ (/۳۶/۵۶)	گوگل

یافته، که توزیعی نرمال با میانگین و واریانس به دست آمده مطابق روابط (۱۵) و (۱۶) می‌باشد:

$$\text{واریانس} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2} + \frac{1}{\sigma_4^2}} \quad (15)$$

(۱۶)

لازم به ذکر است که دقت (RMSE) در جدول بالا از رابطه (۱۷) محاسبه شده است در این رابطه  $x_i$  داده  $i$  ام منبع مورد بررسی،  $\hat{x}_i$  داده واقعی نظیر  $x_i$  و  $N$  تعداد داده های منبع مورد نظر است. به صورت کلی جهت مقایسه نتایج مطالعه با واقعیت از شاخص RMSE استفاده می‌کنند و یکی از پرکاربردترین شاخص‌ها در بررسی دقت نتایج می‌باشد.

(۱۲) می‌باشد:

$$P(\mu | t_1, t_2, t_3, t_4) = \frac{P(\mu) P(t_1 | \mu) P(t_2 | \mu) P(t_3 | \mu) P(t_4 | \mu)}{P(t_1) P(t_2) P(t_3) P(t_4)} \quad (12)$$

با فرض نرمال بودن تمامی توزیع‌ها:

$$P(\mu) = K_1 e^{-\frac{1}{2} \frac{(\mu - \mu_0)^2}{\sigma_0^2}} \quad (13)$$

$$\text{for } i = 1 \text{ to } 4 \rightarrow P(t_i | \mu) = K_{i+1} e^{-\frac{1}{2} \frac{(t_i - \mu)^2}{\sigma_i^2}} \quad (14)$$

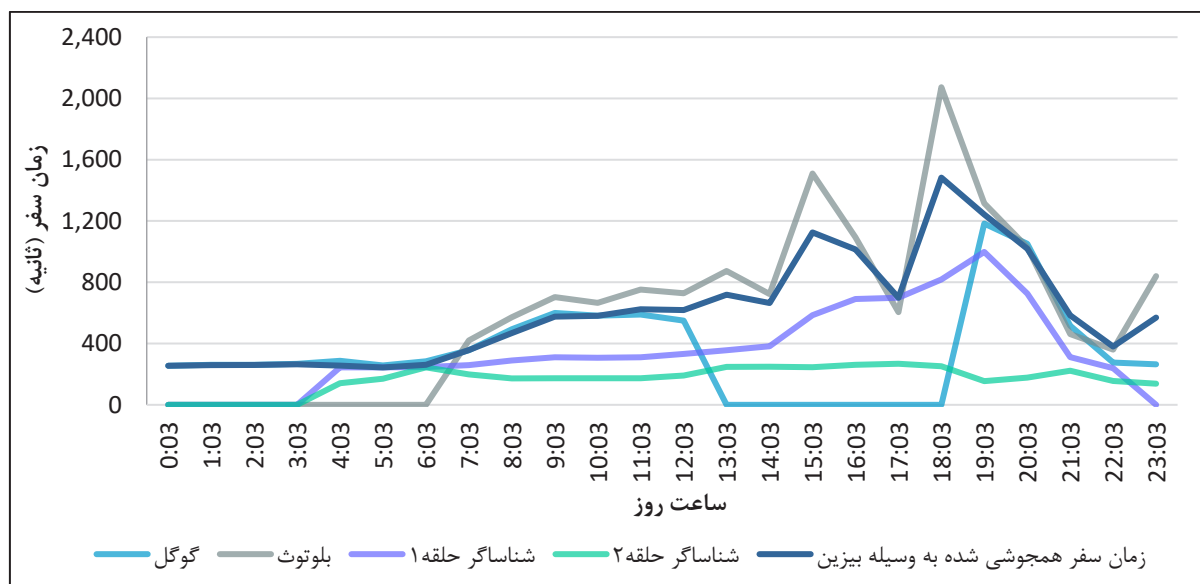
$$\text{میانگین} = \left( \frac{\frac{\mu_0}{\sigma_0^2} + \frac{t_1}{\sigma_1^2} + \frac{t_2}{\sigma_2^2} + \frac{t_3}{\sigma_3^2} + \frac{t_4}{\sigma_4^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2} + \frac{1}{\sigma_4^2}} \right)$$

از ضرب این ۵ توزیع در یکدیگر، در نهایت به توزیع پسین دست

جدول ۳. مقایسه جذر میانگین مربع خطاها در منابع مختلف

Table 3. Comparison of the root mean square of errors in different sources

حسگرها	بلوتوث	گوگل	شناساگر حلقه ۱	شناساگر حلقه ۲
دقت (RMSE)	۲۱۲	۳۳۵	۶۰۲	۶۱۴



شکل ۲. نمودار مقایسه زمان سفرهای منابع (گوگل، بلوتوث، شناساگر حلقه ۱ و شناساگر حلقه ۲) و همجوشی شده همفزون شده ۱ ساعته

Fig. 2. Comparison chart of travel time resources (Google, Bluetooth, Loop Detector 1 and Loop Detector 2) and 1-hour aggregation

همجوشی شده به دلیل دقیق تر بودن زمان سفر حسگر بلوتوث نسبت به سایر حسگرها به آن نزدیک تر است.

همانطور که از شکل ۲ می توان برداشت کرد، در شرایط زیر اشباع و فوق اشباع حسگرهای مختلف دقت های متفاوتی دارند. مثلا حسگر بلوتوث در شرایط اشباع فراز و نشیب های تندی را نشان می دهد یا داده گوگل شرایط فوق اشباع را با تاخیر نشان می دهد. دقت مدل خروجی در همه شرایط به دقت منابع ورودی بستگی دارد و در همه حال (شرایط زیر اشباع و فوق اشباع) دقت خروجی مدل دقت بالاتری از دقت تک تک منابع ورودی دارد.

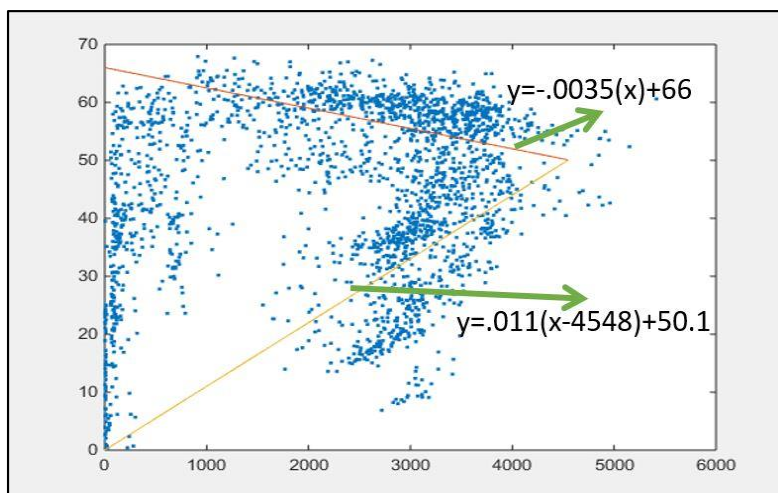
با استفاده از اطلاعات شناساگر حلقه، می توان از دو روش زمان سفر را به دست آورد، دسته نخست از تبدیل اطلاعات سرعت شناساگر حلقه به سرعت متوسط مکانی و سپس تبدیل آن به زمان سفر مقطع

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (17)$$

در مرحله نهایی به منظور همجوشی داده ها، دقت زمان سفرهای حاصل از هر ۴ حسگر بلوتوث، گوگل، شناساگر حلقه ۱ و شناساگر حلقه ۲ که به صورت ۱ ساعته همفزون شده اند، محاسبه شده و خطای منابع مختلف نسبت به یکدیگر در جدول ۳ نشان داده شده است: زمان سفر به دست آمده توسط حسگرهای جداگانه و همچنین همجوشی شده برای روز ۹۵/۹/۲۳ به صورت شکل ۴ می باشد:

لازم به ذکر است برای انتخاب روز یکی از موضوعات موجود بودن حداکثری داده های حسگرها بوده است، با وجود نقص داده های گوگل یکی از بهترین روزها از نظر موجود بودن اطلاعات روز ۲۳ آذر بوده است. همانطور که در این شکل نشان داده شده است نمودار زمان سفر





شکل ۳. نمودار اساسی ترافیک تونل جنوبی  
Fig. 3. Basic chart of southern tunnel traffic

دست آمده از شناساگر حلقه به رسم نمودار اساسی ترافیک بر حسب سرعت-جریان در ساعات مختلف روزهای چهار ماه شهریور، مهر، آبان و آذر ۱۳۹۵ پرداخته شد. این نقاط که مجموعاً پانزده میلیون نقطه بوده‌اند در شکل ۳ زیر قابل مشاهده هستند.

برای استفاده از این نمودار در تخمین سرعت، لازم است این نقاط به یک منحنی برازش داده شود. یکی از نمودارهای رایج برای رسم نمودار اساسی ترافیک، نمودار مثلثی می‌باشد. در اینجا نیز سعی بر آن شد که بر داده‌های موجود، نمودار مثلثی را برازش کنیم که معادله خط آن به در شکل ۵ زیر نمایش داده شده است.

داده‌های جریان به شرح زیر تفکیک شد (جزئیات فرضیات این تقسیم‌بندی در پیوست موجود است):

• جریان کمتر از ۲۵۰۰ وسیله نقلیه بر ساعت

برای محاسبه سرعت از معادله زیر استفاده می‌شود:

$$speed = -0.0035 \times flow + 66$$

• جریان بیشتر از ۲۵۰۰ وسیله نقلیه بر ساعت

ساعات ۸ تا ۱۲ و ۲۲ و ۲۳

$$speed = -0.0035 \times flow + 66$$

ساعات ۶ تا ۷ و ۱۳ تا ۲۱

$$speed = 0.011 \times (flow - 4548) + 50.1$$

می‌باشد. دسته دوم از ترسیم اطلاعات جریان و سرعت وسایل نقلیه در نمودار اساسی ترافیک و سپس با استفاده از داده‌های جریان حاصل از شناساگر حلقه سرعت متوسط مکانی حاصل شد، که با استفاده از سرعت، زمان سفر به دست آمد.

(۱) تبدیل سرعت متوسط زمانی به سرعت متوسط مکانی

$$\bar{u}_{space\ mean\ speed} \approx \bar{u}_{time\ mean\ speed} - \frac{\sigma_{time\ mean\ speed}^2}{\bar{u}_{time\ mean\ speed}}$$

در این رابطه با در نظرگیری توزیع سرعت از پارامترهای این توزیع استفاده می‌شود؛

$\bar{u}_{space\ mean\ speed}$ : میانگین داده‌های سرعت متوسط مکانی

$\bar{u}_{time\ mean\ speed}$ : میانگین داده‌های سرعت متوسط زمانی

$\sigma_{time\ mean\ speed}^2$ : واریانس داده‌های سرعت متوسط زمانی

در بیشتر مشاهدات سرعت، انحراف معیار توزیع تقریباً عددی نزدیک به ۸ km/h است.

(۲) تخمین سرعت متوسط مکانی با استفاده از نمودار اساسی ترافیک

تحقیقات و بررسی‌های زیادی برای به دست آوردن رابطه بین جریان و سرعت انجام شده است؛ از این رو با تبدیل سرعت متوسط زمانی به سرعت متوسط مکانی از قسمت قبل و با استفاده از حجم به



## ۶- نتیجه گیری

نتایج نشان می‌دهد اولاً همجوشی در صورت دسترس بودن داده از دو منبع (گوگل و بلوتوث)، سه منبع (گوگل، بلوتوث و شناساگر حلقه ۱)، و یا چهار منبع (گوگل، بلوتوث، شناساگر حلقه ۱ و شناساگر حلقه ۲) به راحتی قابل انجام است. ثانیاً جواب‌های به دست آمده به منبع بلوتوث که دارای خطای کمتر است نزدیک هستند. در حالتی که از دو دسته داده شناساگر حلقه برای به دست آوردن داده استفاده شده است، به منظور مقایسه با داده‌های برداشت شده ۱۵ دقیقه‌ای گوگل و بلوتوث، به صورت همفزون ۱ ساعته و ۱۵ دقیقه‌ای در نظر گرفته شده است. از این رو با تبدیل دو دسته داده حاصل از شناساگر حلقه از حالت ۱ ساعته به ۱۵ دقیقه‌ای به آن خطای بیشتری اعمال شده است. بنابراین فاصله‌گیری الگوریتم‌های همجوشی در حالت ۱ ساعته کمتر از ۱۵ دقیقه‌ای نسبت به داده با خطای کمتر (بلوتوث) بوده است. به دلیل اینکه بلوتوث در تمامی زمان‌ها موجود نیست و به طور مثال در ساعات نیمه شب، داده‌هایی از بلوتوث در دسترس نیست، این حسگر دارای پوشش زمانی محدود است، اما دارای پوشش مکانی کاملی می‌باشد. از طرف دیگر، به دلیل اینکه شناساگر حلقه تنها در ابتدا و انتهای تونل وجود دارد، اطلاعات تنها در نقاط محدودی از مسیر موجود است و در نتیجه این اطلاعات پوشش مکانی محدودی دارد، اما دارای پوشش زمانی کامل می‌باشد. ترکیب این دو حسگر به وسیله مدل‌های همجوشی، باعث می‌شود نتایج نهایی دارای پوشش مکانی و زمانی کاملی باشند. نتایج حاصل از داده‌های فرضی نشان می‌دهد زمان سفر خروجی از مدل‌های همجوشی، دارای دقت بالاتری نسبت به زمان سفرهای به دست آمده از هر کدام از آن حسگرها، به صورت جداگانه می‌باشد. این مقایسه به صورت جذر میانگین مربع خطاها (RMSE) بررسی شده است و نتایج آن در جدول ۳ نشان داده شده است. از آنجا که زمان سفر همجوشی شده دارای دقت بالاتری نسبت به تک‌تک حسگرهای موجود است، می‌تواند به بهبود همه کاربردهای زمان سفر در بهره‌برداری از شبکه حمل و نقل کمک کند. برخی از این کاربردها عبارتند از: شرکت‌های حمل و نقل بار با در دست داشتن اطلاعات دقیق زمان سفر می‌توانند بهترین مسیر و بهترین زمان را برای خدمت‌رسانی و تحویل بار انتخاب نمایند. اگر اطلاعات نرم‌افزارهای مسیریابی دارای دقت بالایی باشد، کاربران

می‌توانند از برنامه‌ریزی روزمره دقیق‌تری بهره‌مند شوند. در حال حاضر زمان سفر نمایش داده شده بر روی تابلوهای پیام متغیر (VMS)<sup>۱</sup> تهران تنها به وسیله یک حسگر به دست آمده است. با نمایش زمان سفر به دست آمده از مدل‌های همجوشی به جای زمان سفر به دست آمده از تک حسگر، می‌توان زمان سفری با دقت بالاتر در اختیار مسافران قرار داد و در نتیجه کاربران می‌توانند تصمیمات مناسب‌تری را اتخاذ کنند. آژانس‌های اتومبیل و تاکسی‌های اینترنتی با در دست داشتن اطلاعات دقیق‌تر از زمان سفر می‌توانند بهترین مسیر را برای رسیدن به مقصد انتخاب کنند و در هزینه‌ها صرفه جویی قابل توجهی داشته باشند. به طور کلی شرکت کنترل ترافیک اطلاعات ترافیکی را در اختیار بسیاری از اپراتورها قرار می‌دهد. در نتیجه هر چه این اطلاعات دقیق‌تر باشد، عملکرد اپراتورها بهتر می‌شود. برای مطالعات آینده همچنین پیشنهاد می‌شود روش همجوشی توسعه یافته در این مقاله در سطح گسترده به کارگیری شود، در این حالت کمان‌ها و گره‌ها باید به صورت دقیق مشخص شوند، به این معنا که تعداد خطوط هر کمان و اطلاعات ترافیکی دقیق آن، در دسترس باشد تا بتوان زمان سفر را در همه قطعات شبکه با دقت بالایی محاسبه کرد.

## پیشنهادات:

یکی دیگر از منابع داده ورودی مدل‌های همجوشی، زمان سفرهای محاسبه شده از روی دوربین‌های نظارتی در تونل است که دسترسی به فیلم یا اطلاعات پردازش شده آن در این تحقیق ممکن نشد. لذا پیشنهاد می‌شود در صورت امکان نیز به کمک این داده‌ها و مقایسه آن با داده‌های منابع دیگر و یا حتی در روش‌های دیگر این روند انجام شود.

در روابط ذکر شده فرض نرمال بودن در نظر گرفته شده است. پیشنهاد می‌شود در آینده این موضوع با فرض نرمال نبودن نیز مورد بررسی قرار گیرد.

## مراجع

- [1] Bachmann, C., et al., A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling.

<sup>1</sup>Variable Message Sign

- fusion. 2004
- [12] Park, T. and S. Lee. A Bayesian approach for estimating link travel time on urban arterial road network. in International Conference on Computational Science and Its Applications. 2004. Springer.
- [13] Bachmann, C., Multi-sensor data fusion for traffic speed and travel time estimation. 2011
- [14] Guo, K., et al., Application of multi-sensor target tracking to multi-station monitoring data fusion in landslide. *Yantu Lixue(Rock and Soil Mechanics)*, 2006 3 (27): p. 481-479
- [15] Martí ,F.S., Highway travel time estimation with data fusion. 2013, Springer.
- [16] Soriguera, F. and F. Robusté, Highway travel time accurate measurement and short-term prediction using multiple data sources. *Transportmetrica*, 1(7).2011: p. 109-85
- [17] Liu, K ,et al., Iterative bayesian estimation of travel times on urban arterials: fusing loop detector and probe vehicle data. *PloS one*, (6)11 .2016
- [18] Mahmoud Mesbah, A.G., Farzaneh Abdollahi, Fatemeh Banani, Zahra Pakdaman, Afarin Kheirati, Mohsen Kheirandish & Amirhossein Alikhani, Research plan for transportation systems data fusion Smart city of Tehran to estimate travel time. *Amirkabir University of Technology (Tehran Polytechnic)*, 2016: p. 149.
- [19] Guo, Y. and L. Yang, Reliable Estimation of Urban Link Travel Time Using Multi-Sensor Data Fusion. *Information*, (5)11 .2020: p. 267
- [20] Mil, S. and M. Piantanakulchai, Modified Bayesian data fusion model for travel time estimation considering spurious data and traffic conditions. *Applied Soft Computing*, 72 .2018: p. 78-65
- [21] Zahra Pakdaman, Fusion of traffic data to determine the network performance evaluation index. *Iran University of Science and Technology*, 2017
- Transportation research part C: emerging technologies, 26 .2013: p. 48-33
- [2] Bachmann, C., et al ,Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. *Journal of Intelligent Transportation Systems*, 2(17) .2013: p. 164-152
- [3] Kwon, J., B. Coifman, and P. Bickel, Day-to-day travel-time trends and travel-time prediction from loop-detector data. *Transportation Research Record*, 2000 1(1717): p. 129-120
- [4] Zhang, X. and J.A. Rice, Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, 4(3)11 .2003): p. 210-187
- [5] El Faouzi, N.-E., L.A. Klein, and O. De Mouzon, Improving travel time estimates from inductive loop and toll collection data with Dempster–Shafer data fusion. *Transportation research record*, 1)2129 .2009): p. 80-73
- [6] Heilmann, B., et al., Predicting motorway traffic performance by data fusion of local sensor data and electronic toll collection data. *Computer-Aided Civil and Infrastructure Engineering*, 6)26 .2011): p. 463-451
- [7] Ivan, J.N., Neural network representations for arterial street incident detection data fusion. *Transportation Research Part C: Emerging Technologies*, 4(3)5 .1997): p. 254-245
- [8] Bhaskar, A., E. Chung, and A.-G. Dumont, Analysis for the use of cumulative plots for travel time estimation on signalized network. *International Journal of Intelligent Transportation Systems Research*, 3(8) .2010: p. 163-151
- [9] Skabardonis, A. and N. Geroliminis, Real-time estimation of travel times on signalized arterials. 2005
- [10] Pueboobpaphan, R. and T. Nakatsuji, Real-time traffic state estimation on urban road network: the application of unscented Kalman filter, in *Applications of Advanced Technology in Transportation*. 2006. p. 547-542
- [11] Chu, L. and W. Recker, Micro-simulation modeling approach to applications of on-line simulation and data

2016-11-05T17:15:00+03:30;15 mins;912
2016-11-05T17:30:00+03:30;18 mins;1055
2016-11-05T17:45:00+03:30;15 mins;899

شکل a. داده‌های خام گوگل  
**Fig. a. Google Raw Data**

2016-11-05T17:15:00+03:30;15 mins;912
2016-11-05T17:30:00+03:30;18 mins;1055
2016-11-05T17:45:00+03:30;15 mins;899

شکل b. داده‌های خام گوگل پس از مرتب‌سازی  
**Fig b. Google raw data after sorting**

چگونه به این مقاله ارجاع دهیم

*N. Family, Fusion of travel time data in Niayesh tunnel using Bayesian inference, Amirkabir J. Civil Eng., 53(10) (2022) 4205-4218.*

DOI: [10.22060/ceej.2020.18214.6803](https://doi.org/10.22060/ceej.2020.18214.6803)

