From Raw Data to 'Best' Travel Route Computation

LAZAR PENDOV[1], IVAN GANCHEV[1,2,3*],
[1]Faculty of Mathematics and Informatics,
University of Plovdiv "Paisii Hilendarski",
Plovdiv,
BULGARIA

[2]Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences (IMI–BAS),
Sofia,
BULGARIA

[3]Telecommunications Research Center (TRC),
University of Limerick,
Limerick,
IRELAND

Abstract: - This paper presents the processing of raw data for 'best' travel route computation and recommendation to consumers by a specially developed web application operating on top of the IoT platform EMULSION. Three different types of 'best' travel routes are computed for each journey requested by a consumer, namely the healthiest route, the fastest route, and the shortest route. The importance of the presented work lies within the practical applicability of the proposed method for processing raw data for just-on-time 'best' travel route computation and recommendation to consumers.

Key-Words: - Internet of Things (IoT), IoT platform, travel route, computation, recommendation, row data.

Received: June 11, 2024. Revised: January 9, 2025. Accepted: March 12, 2025. Published: May 14, 2025.

1 Introduction

The existing web-based route planners and real-time navigation systems, accessed by fixed or mobile personal devices, allow to plan and adjust journeys by providing comfort and a sense of safety to consumers, [1]. By supplying dynamic and integrated technological support tools to consumers, along with interactive planning and navigation features for various transportation types, these systems, however, mainly find the shortest, fastest, or cheapest traveling routes. Nowadays, people have begun to pay more attention to their quality of life (QoL) by considering environmental factors affecting their health, including air quality, and this has become their new focus when traveling, [2]. Travel types and routes with relatively low pollutant exposure can not only improve human health but can also benefit social stability and sustained progress. In contrast, path-based long-distance outdoor activities, surrounded by poor air quality, generally have a negative effect on human health, especially when cycling, running, jogging, or walking, [3]. Therefore, an alternative approach is needed in route planning based on health-related optimization criteria, bringing the concept of "healthy route" (a.k.a. "green route" or "clean route") and related concepts of "safe route" and "sustainable route,", [1].

This paper demonstrates the use of a developed web application for the recommendation of 'best' travel routes to consumers. The application utilizes a generic distributed architecture, which is technology-independent w.r.t. hardware, operating system, and programming language. A clear explanation of the computation of 'best' travel routes is provided, along with a demonstration of the way this web application works in reality. The usefulness of the presented study relates to the possible integration of the proposed method for 'best' travel route computation and recommendation into the existing navigation systems and applications alike.

2 Related Work

In general, there are two primary types of methods for computing travel routes [3], i.e., based on (1) the *static* cost of paths, serving as an input to the standard Dijkstra algorithm [4] and (2) the *dynamic* cost of paths, varying over space and time, and depending on the travel time (as a function of the path infrastructure's condition and traffic flow) and the Air Quality Index (AQI) value, for instance. AQI was developed by the US EPA, based on the following six major pollutants: ground-level ozone (O3), particulate matter (PM2.5 and PM10), carbon monoxide (CO), sulfur dioxide (SO2), and nitrogen dioxide (NO2). The AQI values range from 0 to 500, divided into six levels of health-related concern. Being more advanced, multiple methods of the second type have been proposed, e.g., utilizing adaptive decision rules [5], genetic algorithms [6], [7], [8], probabilistic models [9], uncertainty [10], etc.

Regarding the computation of healthy routes, previous research could be divided into two groups [2], i.e., using: (1) monitoring stations to measure pollutant exposure on various types of roads, followed by classification of roads as healthy or unhealthy, based on the exposure levels; and (2) pollution distribution data obtained by different means, such as the land use regression [11], [12], the operational street pollution model [13], the interpolation method [14], [3], etc., as an input to the (dynamic) Dijkstra algorithm for generating healthy routes [15], using different indicators (e.g., traffic volume, AQI value, potential pollutant dose taken, etc.) as road network's weights. If taking full advantage of modern pollutant retrieval technologies, the trustworthiness of the generated healthy routes could be significantly increased. For this, [2] proposes a short-distance healthy route planning approach, utilizing fine spatial resolution images and meteorological and socioeconomic data to retrieve the spatial distribution of $PM_{2.5}$ concentration in hourly intervals via a back-propagation neural network. The effectiveness of the approach is verified by comparing the $PM_{2.5}$ potential dose reduction rate of the generated healthy route with that of the shortest route, reaching up to 20% reduction in some cases. As an important factor affecting the AQI values, $PM_{2.5}$ concentration can be used also to predict the AQI [3].

The current paper utilizes an interpolation method from the second group for the computation of the healthiest route, computed in near real-time by applying the dynamic Dijkstra algorithm, whereby the potential exposure rate is computed based on collected raw AQI data, whereby the desired values of the main air pollutants (O_3, $PM_{2.5}$, PM_{10}, CO, SO_2, NO_2) serve as upper bounds for reducing the number of possible routes.

3 Data

Before initiating the computation of the 'best' travel route, some preliminary data preparation is needed. These data are divided into three categories: permanent, refreshable, and computable data. The permanent data are created once and change very rarely if changed at all, while the other two categories are updated periodically as new data become available.

3.1 Permanent Data

This category includes data related to the travel map of the target area, including roads, streets, points, and other details. Locations of the AQI monitoring stations are also included here, along with data of the nearest AQI stations to each respective point to enable faster computations.

The permanent data are generated from the OpenStreetMap (OSM) [16], which includes three main types of objects: a road (a 'way' in the OSM terminology), a node, and a relation. Paths are composed of a series of nodes, and relations describe more complex structures. Here, we will focus on the roads/ways and the corresponding nodes that are part of these roads/ways.

First, a specific POST request is made to the OSM application programming interface (API) for selecting, as a locality, a rectangular area bounded by the GPS coordinates (36.846000, 114.345000) and (37.655772, 115.684369), covering an area of 10,735.78 km² and encompassing nearly all of Hebei (China), as follows:

API URL: https://overpass-api.de/api/interpreter
POST (without urlencode):
data=[out:json];(way["highway"](36.846000,114.345000,37.655772,115.684369);>;);out;
POST (with urlencode):
data=%5Bout%3Ajson%5D%3B%28way%5B%22highway%22%5D%2836.846000%2C114.345000%2C37.655772%2C115.684369%29%3B%3E%3B%29%3Bout%3B

The returned result is in a JavaScript Object Notation (JSON) format. Figure 1 displays a portion of this structure formatted as a Hypertext Preprocessor (PHP) jagged array, visualized using the *print_r* PHP function (for compactness, some of the content in Figure 1 is omitted and replaced with multiple dots). The structure contains two types of elements – ways, and nodes (nodes are usually arranged at the beginning). Each node has a unique identifier (id) and GPS coordinates – latitude (lat) and longitude (lon). Each way has a unique identifier (id), a list of included nodes, and a 'tags' field that contains the way type (*highway*) as well as other characteristics, e.g., one-way (*oneway*=*yes*), the name of the corresponding street or boulevard, etc.


```
Array
(
    [version] => 0.6
    [generator] => Overpass API 0.7.62.1 084b4234
    [osm3s] => Array
        (
            [timestamp_osm_base] => 2024-11-05T13:59:45Z
            [copyright] => The data included in this document is from
www.openstreetmap.org. The data is made available under ODbL.
        )

    [elements] => Array
        (
            [0] => Array
                (
                    [type] => node
                    [id] => 567683623
                    [lat] => 36.8304488
                    [lon] => 114.541854
                )
.......
            [250098] => Array
                (
                    [type] => way
                    [id] => 356726107
                    [nodes] => Array
                        (
                            [0] => 3621366680
                            [1] => 3621366681
                            [2] => 3621366682
                            [3] => 3621366678
                            [4] => 3621366676
                            [5] => 3621366673
                            [6] => 3621366667
                            [7] => 3621366663
                        )

                    [tags] => Array
                        (
                            [bridge] => yes
                            [highway] => motorway_link
                            [layer] => 1
                            [oneway] => yes
                        )

                )
.......
        )
)
```


Fig. 1: An OSM JSON output in *print_r* format

The most important elements are the nodes, along with the information about their linking to each other, including the type of corresponding road/way. For example, the road/way with id=356726107 is a one-way road, including nodes linked in a sequence, as follows: 3621366680 to 3621366681, 3621366681 to 3621366682, 3621366682 to 3621366678, and so on, up to the last node in the list. If the road/way is a two-way one (the more common case), in addition to the links in the forward direction, one must also add those in the reverse direction. The type of linking between two nodes depends on the type of road/way they are part of.

The links can be recorded in a relational database (e.g., MySQL), as shown in Table 1 for the example of a *motorway_link* marked with *way_type_id=3*. The data for the nodes participating in this link are presented in Table 2.

Table 1. Sample link records

id	node1_id	node2_id	way_type_id
1	3621366680	3621366681	3
2	3621366681	3621366682	3
3	3621366682	3621366678	3
4	3621366678	3621366676	3
5	3621366676	3621366673	3
6	3621366673	3621366667	3
7	3621366667	3621366663	3

Table 2. Sample node records

id	lat	lon
3621366680	37.38821060	115.27720750
3621366681	37.38827570	115.27676360
3621366682	37.38827840	115.27635720
3621366678	37.38820520	115.27598150
3621366676	37.38810480	115.2756298
3621366673	37.38793930	115.27527120
3621366667	37.38774930	115.27500820
3621366663	37.3875105	115.27475890

In addition to the existing data, some preliminary computations need to be performed as well.

The first computation relates to the distances between the corresponding nodes in Table 1, which should be entered in an additional distance column. The distance between two points on the surface of Earth could be computed using the haversine formula, as follows:

$$\begin{aligned} a &= \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right) \\ c &= 2 \cdot \operatorname{atan2}\left(\sqrt{a}, \sqrt{1-a}\right) \\ d &= R \cdot c \end{aligned} \tag{1}$$

where $\varphi 1$ and $\varphi 2$ are the latitudes of the two points (in radians), $\Delta\varphi$ is the difference between the latitudes of the two points (in radians), $\Delta\lambda$ is the difference between the longitudes of the two points (in radians), and R is the Earth's radius.

The second computation uses formulae (1) to determine the distances from each node to every AQI station deployed in the area, helping to identify the three closest AQI stations and their respective distances to that node. The relationship between each node and the three nearest AQI stations is cached for future use.

3.2 Refreshable Data

These data are refreshed periodically and most often come from the AQI stations. These data are typically obtained through GET or POST requests to the respective API providing data from the AQI stations. It is also possible to combine the data obtained from different providers.

3.3 Computable Data

Each time the refreshable data are updated, all computable data must be recomputed. In the case of computing the healthiest routes, all AQI values reported by the monitoring stations need to be distributed across the nodes. For this purpose, we use the three closest AQI monitoring stations already selected for each node and the pre-computed distance from this node to each of these stations.

Obtaining the AQI value for each node can be done using the following formula:

$$\mathrm{AirQ} = \frac{\frac{\mathrm{AirQ}_1}{d_1} + \frac{\mathrm{AirQ}_2}{d_2} + \frac{\mathrm{AirQ}_3}{d_3}}{\frac{1}{d_1} + \frac{1}{d_2} + \frac{1}{d_3}} \tag{2}$$

where AirQ, AirQ_1, AirQ_2, and AirQ_3 are the air quality values for this node and the three closest AQI monitoring stations, respectively, and d_1, d_2, and d_3 represent the distance from this node to each of these AQI stations, respectively.

Computable data, along with the link from each node to its neighboring nodes, are cached in a format optimized for fast reading. The current implementation is done in PHP using the APCu module for caching in RAM. The first step is to create a PHP jagged array, as shown in Figure 2, visualized with the *print_r* PHP function (for compactness, some of the content in Figure 2 is omitted and replaced with multiple dots).

This jagged array is first saved as a PHP file suitable for inclusion via the *include* PHP function, and then a cache is created in RAM using APCu, making it accessible to all PHP instances.

```
Array
(

.......

  [567683685] => Array
    (
      [lat] => 37.1439951
      [lng] => 114.5781211
      [pm2_5] => 63.58097790827
      [pm10] => 137.40262105808
      [o3] => 180.98915786492
      [co] => 1.3228594055887
      [so2] => 21.429427890673
      [no2] => 17.216362196162
      [aqi] => 99.265806687428
      [nodes] => Array
        (
          [1949807523] => Array
            (
              [distance] => 338.4824
              [way_type] => 1
            )

          [1949807522] => Array
            (
              [distance] => 404.9615
              [way_type] => 1
            )

        )

    )

.......

)
```

Fig. 2: Node data in *print_r* format

4 Web Application

The corresponding web application, which has been developed for the 'best' travel routes recommendation as part of the client tier of the IoT platform EMULSION [17], can be tested as per [18].

The sliders, depicted in Figure 3, allow consumers to set their maximum preferable values for air quality indicators ($PM_{2.5}$, PM_{10}, O_3, CO, SO_2, NO_2, and AQI). During the route computation, none of these values for the respective indicator can be exceeded.

A dropdown menu allows setting the transportation type. The sliders, shown in Figure 4, allow consumers to set their average speed for different road types. If a consumer feels confused with the numerous sliders or s/he is unsure of their average travel speed, s/he can use the default values. Different transportation types offer various options for average speed and different protection from polluted air.

Figure 5 (Appendix) shows the lower section of the developed app's screen, with most of the space occupied by the map of the target area. To visualize the map and perform operations on it, the JavaScript library Leaflet [19] was used. A left click on the map sets the starting point of the travel route, while a right-click sets the endpoint. If a selected point does not match a specific node, it is replaced with the nearest node.

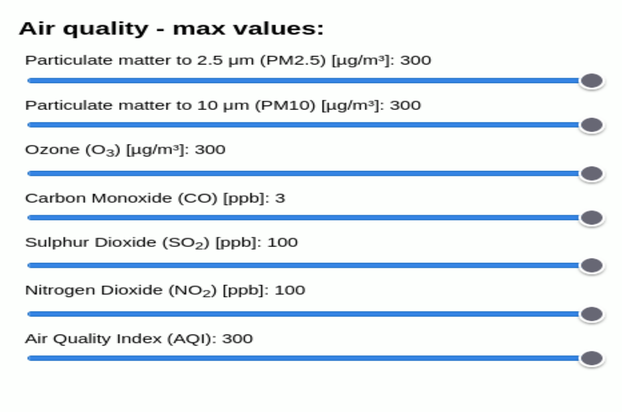

Fig. 3: Air quality indicators' limits

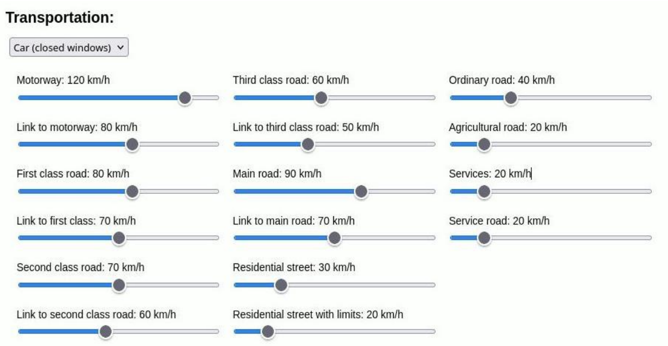

Fig. 4: Transportation options

In the end, the consumer selects the types of travel route s/he wants (i.e., the healthiest, the fastest, or the shortest one), as shown on the right-hand side in Figure 5 (Appendix). More than one route type can be chosen. After completing the settings, the consumer clicks on the "Calculate route(s)" button, and the route computation process begins.

It is important to note that the web application itself does not perform any route computations. It makes a call to a proxy server, which forwards the request to one of the utilized computation servers to generate a route. If all three types of routes are selected, three separate requests are made to different computation servers. The generated routes are visualized on the map, and information about these appears below the buttons shown on the right-hand side in Figure 5 (Appendix). By pressing the "Details" button, which then appears, detailed information about the route can be obtained, as depicted in Figure 6 (Appendix).

5 'Best' Travel Route Computation

The computation of the 'best' travel route is performed by the selected computation server using the dynamic Dijkstra algorithm, with a different metric depending on the chosen route type. If computing the shortest route, the metric used is the route length (in meters). If computing the fastest route, the metric used is the time required to travel on the route.

For computing the healthiest route, a metric needs to be defined that properly describes exposure to polluted air. There are many options for doing this, such as considering only some of the air pollutants ($PM_{2.5}$, PM_{10}, O_3 , CO, SO_2 , NO_2 , and AQI) or a combination of them, weighted differently. In the developed application, it was decided to use the overall air quality indicator (AQI). The exposure is computed using the following formula:

$$\text{Exposure} = \frac{\text{AQI}_1 + \text{AQI}_2}{2} \times \text{time} \times C \tag{3}$$

where AQI_1 и AQI_2 are the AQI values at two specific sequential nodes of a road/way on the route, *time* is the time needed to travel the distance between these two nodes, and *C* is a protection coefficient, whose value (in the range of 0 to 1) is determined based on the transportation type, as shown in Table 3.

Table 3. Values of the protection coefficient *C*

Transportation type	***C***
Car (closed windows)	0.200
Car (open windows)	0.229
Motorbike	0.286
Bicycle	1.000
Walk	0.429

6 Conclusion

This paper has examined the process of computing and recommending the 'best' travel route to consumers, describing the steps from receiving raw data to generating and recommending the route itself. Preliminary data preparation, divided into permanent, refreshable, and computable data, has been described, along with the implementation of a specially developed web application providing multiple options for computing the 'best' travel route using a distributed server architecture.

Future directions for research and development will include adding traffic load data for introducing a speed-reducing coefficient and inclusion of expanding options for computing the routes based on extra data, e.g., supplied by weather forecasting and AQI prediction algorithms. With different input data, the application could be modified to recommend also the safest route (e.g., through areas with the lowest crime rate depending on the time of the day, day of the week, etc.), the most comfortable route (with the best road conditions), and other types of routes.

References:

[1] P. Ribeiro and J. F. G. Mendes, "Healthy routes for active modes in school journeys," *International Journal of Sustainable Development and Planning,* vol. 8, no. 4, pp. 591-602, 2013. DOI: 10.2495/sdp-v8-n4-591-602.

[2] L.-N. Gao *et al.*, "A short-distance healthy route planning approach," *Journal of Transport & Health,* vol. 24, 101314, 2022. DOI: 10.1016/j.jth.2021.101314.

[3] Z. Zou, T. Cai, and K. Cao, "An urban big data-based air quality index prediction: A case study of routes planning for outdoor activities in Beijing," *Environment and Planning B: Urban Analytics and City Science,* vol. 47, no. 6, pp. 948-963, 2020. DOI: 10.1177/2399808319862292.

[4] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik,* vol. 1, no. 1, pp. 269-271, 1959. DOI: 10.1007/bf01386390.

[5] R. W. Hall, "The Fastest Path through a Network with Random Time-Dependent Travel-Times," *Transportation Science,* vol. 20, no. 3, pp. 182-188, 1986. DOI: 10.1287/trsc.20.3.182.

[6] M. B. Pellazar, "Vehicle route planning with constraints using genetic algorithms," in *Proc. of National Aerospace and Electronics Conference (NAECON'94)*, vol. 1, pp. 111-118. Dayton, OH, USA, 1994. DOI: 10.1109/NAECON.1994.333010.

[7] S. B. Pattnaik, S. Mohan, and V. M. Tom, "Urban bus transit route network design using genetic algorithm," *J Transp Eng-Asce,* vol. 124, no. 4, pp. 368-375, 1998. DOI: 10.1061/(Asce)0733-947x(1998)124:4(368).

[8] S. Chien, Z. W. Yan, and E. Hou, "Genetic algorithm approach for transit route planning and design," *J Transp Eng-Asce,* vol. 127, no. 3, pp. 200-207, 2001. DOI: 10.1061/(Asce)0733-947x(2001)127:3(200).

[9] J. Boyan and M. Mitzenmacher, "Improved results for route planning in stochastic transportation," *Proc. of 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, Washington, D.C., USA, 2001. DOI: 10.5555/365411.365803.

[10] E. Nikolova, M. Brand, and D. R. Karger, "Optimal route planning under uncertainty," *Proc. of 16th International Conference on Automated Planning and Scheduling*, Cumbria, UK, 2006. DOI: 10.5555/3037104.3037122.

[11] M. Hatzopoulou *et al.*, "A web-based route planning tool to reduce cyclists' exposures to traffic pollution: a case study in Montreal, Canada," *Environ Res,* vol. 123, pp. 58-61, 2013. DOI: 10.1016/j.envres.2013.03.004.

[12] B. Zou, S. Li, Z. Zheng, B. F. Zhan, Z. Yang, and N. Wan, "Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk," *Computers, Environment and Urban Systems,* vol. 80, 101456, 2020. DOI: 10.1016/j.compenvurbsys.2019.101456.

[13] O. Hertel, M. Hvidberg, M. Ketzel, L. Storm, and L. Stausgaard, "A proper choice of route significantly reduces air pollution exposure--a study on bicycle and bus trips in urban streets," *Sci Total Environ,* vol. 389, no. 1, pp. 58-70, 2008. DOI: 10.1016/j.scitotenv.2007.08.058.

[14] H. Zahmatkesh, M. Saber, and M. Malekpour, "A New Method for Urban Travel Rout Planning Based on Air Pollution Sensor Data," *Current World Environment,* vol. 10, no. Special-Issue1, pp. 699-704, 2015. DOI: 10.12944/CWE.10.Special-Issue1.83.

[15] S. Mahajan, Y.-S. Tang, D.-Y. Wu, T.-C. Tsai, and L.-J. Chen, "CAR: The Clean Air Routing Algorithm for Path Navigation With Minimal PM2.5 Exposure on the Move," *IEEE Access,* vol. 7, pp. 147373-

147382, 2019. DOI: 10.1109/access.2019.2946419.

[16] J. J. Arsanjani, A. Zipf, P. Mooney, and M. Helbich, *OpenStreetMap in GIScience: Experiences, Research, and Applications*, 1st ed. Cham: Springer International Publishing: Imprint: Springer, 2015.

[17] I. Ganchev, Z. Ji, and M. O'Droma, "Horizontal IoT Platform EMULSION," *Electronics*, vol. 12, no. 8, 2023. DOI: 10.3390/electronics12081864.

[18] Healthy Route Discovery, [Online]. http://emulsion.science/en/HealthyRoute (Accessed Date: February 16, 2025).

[19] D. Edler and M. Vetter, "The Simplicity of Modern Audiovisual Web Cartography: An Example with the Open-Source JavaScript Library leaflet.js," *KN - Journal of Cartography and Geographic Information*, vol. 69, no. 1, pp. 51-62, 2019. DOI: 10.1007/s42489-019-00006-2.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the presented research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This publication has emanated from research conducted with the financial support of the Bulgarian National Science Fund (BNSF) under the Grant No. KP-06-IP-CHINA/1 (КП-06-ИП-КИТАЙ/1).

Conflict of Interest

The authors have no conflicts of interest to declare.

APPENDIX

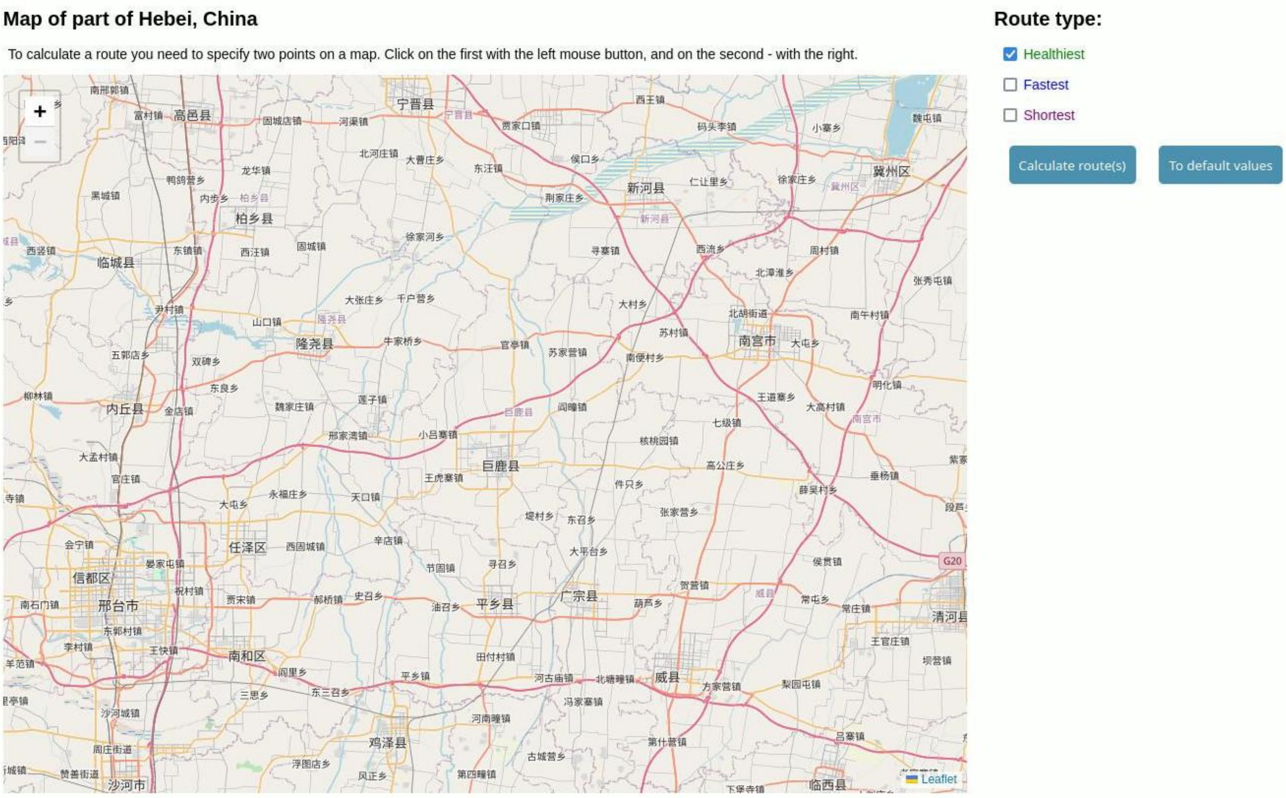

Fig. 5: The map of the target area

Healthiest

Parameter	Start	End	Min	Max	Average	Exposure [Parameter.minutes]
PM2.5 [µg/m³]	23.57	26.17	11.09	56.49	26.51	378.83
PM10 [µg/m³]	58.02	45.42	41.91	81.78	63.52	907.77
O_3 [µg/m³]	34.94	56.63	20.74	57.93	34.86	498.22
CO [ppb]	0.42	0.43	0.3	1.45	0.61	8.77
SO_2 [ppb]	10.43	7.23	4	11.86	8.17	116.75
NO_2 [ppb]	33.24	64.11	18.57	73.85	45.21	646.1
AQI	54.35	53.97	52.65	84.93	61.22	875.01

Fig. 6: Sample healthiest route’s detailed information