Approximation of the road segments travel time using Levy distributions in the reliable shortest path problem

A A Agafonov¹, V V Myasnikov¹,² and A I Maksimov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086
²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

e-mail: ant.agafonov@gmail.com

Abstract. The current trend towards an increase in the number of vehicles, especially in large cities, as well as the unavailability of the existing road infrastructure to distribute modern traffic flows leads to a higher congestion level in transportation networks. This problem emphasized the relevance of navigation problems. Despite the popularity of these problems, many existing commercial systems consider only deterministic networks, not taking into account the time-dependent and stochastic properties of traffic flows. In this paper, we consider the reliable shortest path problem in a time-dependent stochastic transportation network. The considered criterion is maximizing the probability of arriving at the destination point on time. We consider the base algorithm for the stochastic on-time arrival problem, which has a computationally complex convolution operation for calculating the arrival probability. We propose to use parametrically defined Levy stable probability distributions to describe the travel time of road segments. We show, that the use of stable distributions allows us to replace the convolution operation with the distribution value, and significantly reduces the execution time of the algorithm. Experimental analysis has shown that the use of stable distributions allows approximating the exact value of the arrival probability at a destination with a low approximation error.

1. Introduction

The navigation task continues to be one of the most crucial problems in transportation systems. Although existing works investigate this problem in various formulations, including considering time-dependent and stochastic transportation networks, commercial systems still work with deterministic networks. Taking into consideration not only the expected travel time but also the variance of time, i.e., route reliability, makes the optimal routing task computationally difficult.

This paper considers the problem of finding a reliable shortest path in a stochastic transportation network. The formulation is the following: we determine the optimal routing policy, which maximizes the arrival probability at the destination within a predetermined time period. In this paper, we propose to use stable probability distributions to describe the travel time of road network segments, which allows us replacing the convolution operation with recalculation of the density distribution parameters.
2. Related works

The road segment travel time in time-dependent stochastic transportation networks is represented as a random variable with a time-dependent distribution function [1].

A priori shortest path [2] or an adaptive routing policy [1, 3] can be considered as a solution of the routing problem. The route optimality can also be formulated depending on the objective function used:

- minimization of the least expected time (LET-problem) [4, 2, 1],
- maximization of the probability of arrival within a specified time period (Stochastic On-Time Arrival - SOTA-problem) [5, 6, 3],
- minimization of the travel time budget for a specified on-time arrival probability [7, 8, 9].

The paper solves the problem of finding a routing policy that maximizes the probability of arrival within a given time budget.

In [5] authors formulated the SOTA-problem as a stochastic dynamic programming problem and used the standard method of successive approximation to solve it. However, this method has poor convergence. Alternatively, in [6] a discrete approximation algorithm was proposed for the SOTA problem, which converges in a finite number of steps and works in pseudopolynomial time.

In [3] an exact solution to the SOTA problem for networks in which the travel time of the segments is a positive value was proposed. As in [5], one of the steps of the algorithm is the convolution calculation, which is the main computationally complex task. In general terms, this convolution cannot be calculated analytically, and therefore a discrete approximation scheme is required. The solution presented in [3] performs batch convolution calculations, which is more efficient than using the standard algorithm used in [6].

Improving the quality of the navigation algorithm and reducing the error in forecasting the travel time can be achieved by predicting traffic flows on the road network segments [10, 11]. The paper [12] presented a modification of the [3] solution, which takes into account current and forecast information about the parameters of traffic flows in the network.

Several studies have been devoted to accelerating the SOTA problem solving algorithm. In the article [13], the authors presented several methods for accelerating the algorithm for solving the SOTA problem, including improved convolution calculation algorithms using the fast Fourier transform and algorithms for calculating convolution with zero delay [14], as well as methods for determining the optimal order of routing policy calculation. The article [15] presented a heuristic for finding an adaptive traffic route in a stochastic network, the proposed method provides the most computationally efficient strategy for finding a path for general probability distributions. The paper [16] presented stochastic variants of two graph pre-processing methods for solving the problem of finding a deterministic shortest path that can be adapted to the SOTA problem. The parallelization strategy using the graphics accelerator was proposed in [17].

The solution of the reliable routing problem can be used to solve other problems, for example, autonomous routing [18].

In this paper, we propose to use the Levy probability distribution to describe the travel time of road segments, which allows us to replace the computationally complicated convolution operation by recalculation the distribution function parameters.

3. Basic notation and problem formulation

We consider a time-dependent stochastic road network as an oriented graph $G = (N, A, P)$, where $N$ is the set of graph vertices, $A$ is the set of graph edges, $P$ is the probabilistic description of the segments travel times.

In time-dependent stochastic networks, the weight of each segment $(i, j) \in A$ is usually represented as a random variable $T_{ij}(\tau)$ with a time dependent probability density $p_{ij}(t)$. 
An optimal routing policy is defined as a strategy of maximizing the arrival probability at the destination vertex \(d \in N\) within a time budget \(T\). Let \(u_i(t)\) be the probability of reaching the vertex \(d\) from the vertex \(i\) in less than time \(t\) when following the optimal policy. Then the optimal routing policy is formulated as [3]:

\[
\begin{align*}
\tau^*_i(t) &= \max_{j \in N \setminus \{d\}, (i,j) \in A} \int_0^t p_{ij}(\theta) u_j^{\tau^*_j}(t - \theta) d\theta, \quad \forall i \in N \setminus \{d\}, t \in [0, T], \tau \geq 0 \\
u^*_d(t) &= 1, \quad t \in [0, T], \tau \geq 0
\end{align*}
\]

(1)

The following discrete algorithm is used to solve the problem (1):

**Algorithm 1:** Discrete SOTA algorithm

**Step 0.** Initializing

\[
k = 0 \\
u^*_i(x) = 0, \quad \forall i \in N, i \neq d, x \in \mathbb{N}, 0 \leq x \leq \frac{T}{\Delta t} \\
u^*_d(x) = 1, \quad x \in \mathbb{N}, 0 \leq x \leq \frac{T}{\Delta t}
\]

**Step 1.** Updating

\[
k = 1, 2, \ldots, L \\
u^*_d(x) = 1, \quad x \in \mathbb{N}, 0 \leq x \leq \frac{T}{\Delta t} \\
u^*_i(x) = u^{k-1}_i(x), \quad \forall i \in N, i \neq d, (i, j) \in A, x \in \mathbb{N}, 0 \leq x \leq \frac{T}{\Delta t} - 1 \\
u^*_i(x) = \max_j \sum_{h=0}^x p_{ij}(h) u^{k-1}_j(x-h) \\
\forall i \in N, i \neq d, (i, j) \in A, x \in \mathbb{N}, \frac{T}{\Delta t} - 1 \leq x \leq \frac{T}{\Delta t}
\]

In the algorithm \(\Delta t\) is the discretization interval, \(\delta\) is the minimum realizable travel time of the road segments across the entire network.

The next vertex \(j\) using the time budget \(t\) and probabilities \(u_i(x)\) is chosen by the following rule:

\[
j = \arg \max_{i \in N} u_i(t).
\]

(2)

The most computationally complex stage of the algorithm is the convolution calculation \(\sum_{h=0}^x p_{ij}(h) u^{k-1}_j(x-h)\). In the work [12], the log-normal distribution was used as the probability density distribution \(p_{ij}(t)\) of the segments travel times. In this paper, we propose to use a Levy stable distribution that allows us replacing the convolution operation with a recalculation of the distribution function parameters.

4. Proposed model

4.1. Levy stable distribution

A distribution is said to be stable if a linear combination of two independent random variables with this distribution has the same distribution, up to location and scale parameters.

In this paper, we use the Levy stable distribution. The probability density function of the Levy distribution over the domain \(x \geq \mu\) is

\[
f(x; \mu, c) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/2(x-\mu)}}{(x-\mu)^{3/2}},
\]

(3)
where $\mu$ is the location parameter, $c$ is the scale parameter.

The cumulative distribution function is

$$F(x; \mu, c) = \text{erfc}\left(\sqrt{c/\left(2(x - \mu)\right)}\right), \quad \text{(4)}$$

where $\text{erfc}(z)$ is the complementary error function.

If $X_1 \sim \text{Levy}(\mu_1, c_1)$, $X_2 \sim \text{Levy}(\mu_2, c_2)$, then $X_1 + X_2 \sim \text{Levy}(\mu, c)$, where

$$\mu = \mu_1 + \mu_2,$$

$$|c| = (\sqrt{c_1^2 + c_2^2})^2. \quad \text{(5)}$$

4.2. Convolution calculation

Consider the convolution operation in the algorithm 1. We introduce the following notation

$$u_{ij}^k(x) = \sum_{h=0}^{x} p_{ij}(h) u_{j}^{k-1}(x - h). \quad \text{(6)}$$

Then the convolution operation can be written as

$$u_i^k(x) = \max_j \sum_{h=0}^{x} p_{ij}(h) u_j^{k-1}(x - h) = \max_j u_{ij}^k(x). \quad \text{(7)}$$

Consider the expression $u_{ij}^k(x)$ first. Having $u_{ij}^1(x) = 1$, we can obtain the following expression for the edges incoming in the destination vertex $d$:

$$u_{md}^k(x) = \sum_{h=0}^{x} p_{md}(h) u_d^{k-1}(x - h) = \sum_{h=0}^{x} p_{md}(h) = P_{md}(x), \quad \forall m \in V : \exists (m, d) \in E, \quad \text{(8)}$$

where $P_{md}(x)$ is the cumulative distribution function.

Further on, for previous graph vertices $i : (i, m) \in E$ we obtain:

$$u_{im}^k(x) = \sum_{h=0}^{x} p_{im}(h) u_m^{k-1}(x - h) = \sum_{h=0}^{x} p_{im}(h) \sum_{s=0}^{x-h} p_{md}(s) = \sum_{s=0}^{x-h} p_{md}(s) + \sum_{s=0}^{x-h} p_{im}(s) + \cdots + \sum_{s=0}^{x-h} p_{im}(x) p_{md}(0) = \sum_{l=0}^{x} p_{im}(l) p_{md}(x - l) + \sum_{l=0}^{x-1} p_{im}(l) p_{md}(x - 1 - l) + \cdots + \sum_{l=0}^{x-x} p_{im}(l) p_{md}(x - x - l) = \sum_{n=0}^{x-x} p_{im}(l) p_{md}(x - n - l) = \sum_{n=0}^{x} p_{im+md}(x - n) = P_{im+md}(x), \quad \text{(9)}$$

where $p_{im+md}(t)$ is the probability density of the sum of random variables.

The values $u_{ij}^k(x) \forall i, j \in V$ can be calculated in the same way, which allows replacing convolution calculation in the algorithm 1 with the calculation of distribution function value. Scale and location parameters can be obtained using the equation (5).

Next, we have to obtain the estimation of function $u_{ij}^k(x) = \max_j u_{ij}^k(x)$. We approximate the value of $u_{ij}^k(x)$ by the Levy distribution function. Denote the approximable function as $F^*(x)$ and the exact function as $F(x; \hat{\mu}, \hat{c})$. To estimate the required parameters $\alpha, \hat{\mu}, \hat{c}$ it is necessary to minimize the following error:

$$J = e^2 = \frac{1}{2} \sum_j \left(F^*(x_j) - \alpha F(x; \hat{\mu}, \hat{c})\right)^2 \to \min_{\alpha, \hat{\mu}, \hat{c}} \quad \text{(9)}$$
5. Experimental study

The purpose of the experimental studies is to compare the estimated probability of arrival at the destination obtained by calculating the convolution using the equation (7) and by using the approximation by the Levy distribution function $F(x; \hat{\mu}, \hat{c})$.

Figure 1 presents the examples of estimated probability of arrival $u_k^i(x)$. The left part of it shows the case when the vertex $i$ is connected with two other vertices (i.e., there are two possible routes from the vertex $i$). The right part of figure 1 shows the case of three possible routes.

![Figure 1. Probability approximation of arrival for the case of two and three connected vertices.](image_url)

Next, experiments to estimate the approximation error were carried out. The approximation procedure was performed for 100 randomly selected parameters $\mu, c$ of the Levi distribution. Then, the root mean squared error was calculated using the following equation:

$$
RMSE = \frac{1}{n} \sqrt{\sum_{t=1}^{n} (x_t - \hat{x}_t)^2}
$$

where $x_t$ is the exact value, $\hat{x}_t$ is the approximated value.

The mean approximation error is

$$
RMSE = 0.0174.
$$

6. Conclusion

In this paper, we investigated the use of the Levy distributions for the road segments travel time description in the reliable shortest path problem in a stochastic transportation network. This approach allows us to reduce the computation time of the reliable routing algorithm. The use of the Levy distributions allows replacing a computationally complex convolution operation with calculation the distribution function value to estimate the probability of arrival at a destination within a specified time budget.

Based on the results of the experimental study, we can conclude that the use of Levy stable distribution allows approximating the exact probability value of arrival from a given vertex to a destination with high accuracy.
7. References

[1] Gao S and Chabini I 2006 Transportation Research Part B: Methodological 40 93-122
[2] Fu L and Rilett L 1998 Transportation Research Part B: Methodological 32 499-516
[3] Samaranayake S, Blandin S and Bayen A 2012 Transportation Research Part C: Emerging Technologies 20 199-217
[4] Hall R W 1986 Transportation Science 20 182-188
[5] Fan Y and Nie Y 2006 Networks and Spatial Economics 6 333-344
[6] Nie Y and Fan Y 2006 Transportation Research Record 193-200
[7] Chen A and Ji Z 2005 Journal of Advanced Transportation 39 19-37
[8] Nie Y and Wu X 2009 Transportation Research Part B: Methodological 43 597-613
[9] Chen B, Lam W, Sumalee A, Li Q, Shao H and Fang Z 2013 Networks and Spatial Economics 13 123-148
[10] Agafonov A and Myasnikov V 2015 Communications in Computer and Information Science 542 163-174
[11] Agafonov A A, Yumaganov A S and Myasnikov V V 2018 Big data analysis in a geoinformatic problem of short-term traffic flow forecasting based on a k nearest neighbors method Computer Optics 42(6) 1101-1111 DOI: 10.18287/2412-6179-2018-42-6-1101-1111
[12] Agafonov A A, Myasnikov V V 2016 Method for the reliable shortest path search in time-dependent stochastic networks and its application to GIS-based traffic control Computer Optics 40(2) 275-283 DOI: 10.18287/2412-6179-2016-40-2-275-283
[13] Samaranayake S, Blandin S and Bayen A 2012 Speedup techniques for the stochastic on-time arrival problem 25 83-96
[14] Gardner W G 1995 AES: Journal of the Audio Engineering Society 43 127-136
[15] Niknami M and Samaranayake S 2016 Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9685 231-245
[16] Sabran G, Samaranayake S and Bayen A 2014 Precomputation techniques for the stochastic on-time arrival problem 138-146
[17] Abeydeera M and Samaranayake S 2014 Gpu parallelization of the stochastic on-time arrival problem
[18] Agafonov A A, Myasnikov V V 2018 Numerical route reservation method in the geoinformatic task of autonomous vehicle routing Computer Optics 42(5) 912-920 DOI: 10.18287/2412-6179-2018-42-5-912-920

Acknowledgments
The work was partly funded by the RFBR under research projects 18-07-00605-A, 18-29-03135-mk ("Basic notation and problem formulation" and "Proposed model") and by the Russian Federation Ministry of Science and Higher Education within a state contract with the "Crystallography and Photonics" Research Center of the RAS under agreement 007-G3/Ch3363/26 ("Experimental study").