DOI 10.18699/SBB-2022-676

Mathematical modeling of information diffusion in Twitter

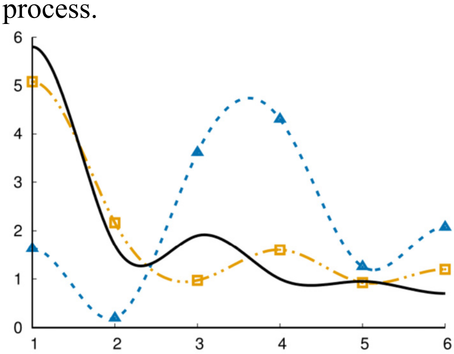
Zvonareva T.[1]*, Krivorotko O.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*
* *t.zvonareva@g.nsu.ru*

Key words: diffusive logistic model, source problem, inverse problem, data analysis, optimization, regularization, tensor train, multilevel gradient method, high performance computing

Motivation and Aim: The social media allow for the rapid diffusion of various information around the world and this information is not always neutral. Such processes could be described with the mean-field games that consists in Kolmogorov-Fokker-Planck (KFP) and Hamilton-Jacobi-Bellman (HJB) equations [1]. KFP characterizes the information diffusion in social media as well as HJB describes the control of such diffusion. In this work we investigate the information diffusion in Twitter represented by the simplified KFP equation which is characterized by its coefficients and the initial condition [2]. These parameters describe the main features of information diffusion in a social network. The aim of the work is to clarify the source of information propagation for formulation the control problem of such diffusion [3].
Methods and Algorithms: The problem of determining the initial condition is reduced to the problem of minimization of the misfit function (inverse problem) which was solved by the global, local and hybrid optimization methods:

- particle swarm optimization (PSO) is a stochastic method for finding the global minimum region;
- multilevel gradient method (MGM) [4] is a modification of the local gradient descent method which has a higher rate of convergence;
- tensor train (TT) [5] is a global optimization method based on tensor decomposition;
- combination of TT and MGM as a hybrid optimization approach.

Results: For the benchmark problem the source (solid black line) of information propagation was specified from additional information about the process at 21 points in time for each hour. The figure shows the solutions of the inverse problem using combinations of PSO + MGM (blue dashed line with triangles) and TT + MGM (orange dashed line with squares). The x-axis is the distance from the source measured in the minimum number of links between the source and the user. The y-axis shows the density of users involved in the process.



Conclusion: For additional information at different points in time without spatial separation the solution of the source problem is nonunique and unstable. The TT+MGM combination could be found the solution closer to the exact one in sense of minimization of the relative accuracy of model and benchmark solutions of inverse problem than the PSO + MGM. However, TT + MGM requires a large computational cost, i.e., to obtain an accuracy of about 10^{-2} it requires 17 hours of computation on the SSCC SB RAS node with 2 Intel Xeon Gold 6248R processors (3 GHz, 24 cores) and 384 GB of main memory. This algorithm will be tested for popular news on Twitter in order to analyze social influence.

Acknowledgements: The study is supported by Russian Science Foundation (project No. 18-71-10044-П).

References

1. Bauso D., Tembine H., Başar T. Opinion Dynamics in Social Networks through Mean-Field Games. *SIAM J Control Optim.* 2016;54(6):3225-3257.
2. Wang F., Wang H., Xu K., Wu J., Jia X. Characterizing Information Diffusion in Online Social Networks with Linear Diffusive Model. *Proc ICDCS.* 2013;307-316.
3. Krivorotko O., ZvonarevaT., Zyatkov N. Numerical solution of the inverse problem for diffusion-logistic model arising in online social networks. *Commun Comput Info Sci.* 2021;1476:444-459.
4. Zvonareva T.A., Krivorot'ko O.I. Comparative analysis of gradient methods for source identification in a diffusion-logistic model. *Zh Vychisl Ma. Mat Fiz.* 2022;62(4):694-704.
5. Oseledets I.V. Tensor-train decomposition. *SIAM J Sci Comput*. 2011;33(5):2295-2317.