

ON JUNTA PROBLEM FOR TABLE-DEFINED FUNCTIONS

P. G. Emelyanov

A (n, k) -junta is a Boolean function of n variables that depends on at most k of these variables. The task of determining whether a given Boolean function is a k -junta and finding these $n - k$ unessential variables is common in machine learning (irrelevant feature reduction) and electronic circuit design (fictitious variable reduction). A natural extension of this idea, which is commonly found in these and other areas, for example, in game theory, is the notion of a variable with little influence on the results of a function's calculation. Of particular interest is the issue of whether insignificant or low-impact variables are considered individually or as an ensemble. The computational efficiency of solving these problems is of significant relevance for practical applications. This depends, in particular, on the representation of the boolean function employed. In this paper Boolean functions are specified by complete or partial truth tables (sample datasets). This paper presents two poly-time algorithms for finding juntas, as well as discussing the case where we approximate a given function making low-impact variables entirely unessential.

Keywords: machine learning, logic design, voting theory, table-defined functions, junta problem, irrelevant features, inessential variables, low-impact variables, cartesian product, SQL JOINS, decomposition of tables.

MSC: 13-04, 68Q32, 68T09, 68W30, 94D10

<https://doi.org/10.21538/0134-4889-2025-31-3-fon-07>

REFERENCES

1. Rong M., Gong D., Gao X. Feature selection and its use in big data: challenges, methods, and trends. *IEEE Access*, 2019, vol. 7, pp. 19709–19725. <https://doi.org/10.1109/ACCESS.2019.2894366>
2. O'Donnell R. *Analysis of Boolean functions*. Cambridge, Cambridge Univ. Press, 2014, 417 p. ISBN: 9781139952507.
3. Mossel E. Probabilistic view of voting, paradoxes, and manipulation. *Bull. Amer. Math. Soc. (New Series)*. 2022, vol. 59, no. 3, pp. 297–330. <https://doi.org/10.1090/bull/1751>
4. Birnbaum Z.W., Esary J.D. Modules of coherent binary systems. *J. Soc. Indust. Appl. Math.*, 1965, vol. 13, no. 2, pp. 444–462. <https://doi.org/10.1137/0113027>
5. Borowik G., Łuba T., Zydek D. Features reduction using logic minimization techniques. *Inter. J. Electron. Telecom.*, 2012, vol. 58, no. 1, pp. 71–76. <https://doi.org/10.2478/v10177-012-0010-x>
6. Amgoud L., Doder D. Compilation of logical arguments. In: *Proc. Twenty-Eighth Inter. Joint Conf. Artificial Intelligence (IJCAI-19)*, 2019, pp. 1502–1508. <https://doi.org/10.24963/ijcai.2019/208>
7. Blum A.L., Langley P. Selection of relevant features and examples in machine learning. *Artific. Intellig.*, 1997, vol. 97, no. 1–2, pp. 245–271. [https://doi.org/10.1016/S0004-3702\(97\)00063-5](https://doi.org/10.1016/S0004-3702(97)00063-5)
8. Crama Y., Hammer P.L. *Boolean models and methods in mathematics, computer science, and engineering*. Ser. Encyclopedia of mathematics and its applications, vol. 134. Cambridge, Cambridge Univ. Press, 2011, pp. 687. ISBN-13: 9780521847520.
9. Tolo S., Andrews J. Fault tree analysis including component dependencies. *IEEE Transactions on Reliability*. 2024, vol. 73, no. 1, pp. 413–421. <https://doi.org/10.1109/TR.2023.3264943>
10. Biswas A., Sarkar P. Influence of a set of variables on a Boolean function. *SIAM J. Discr. Math.*, 2023, vol. 37, no. 3, pp. 2148–2171. <https://doi.org/10.1137/22M1503531>
11. Harder H., Jantsch S., Baier C., Dubsloff C. A unifying formal approach to importance values in Boolean functions. In: *Proc. Thirty-Second Inter. Joint Conf. Artificial Intelligence (IJCAI-23)*, 2023, pp. 2728–2737. <https://doi.org/10.24963/ijcai.2023/304>
12. Emelyanov P., Ponomaryov D. On tractability of disjoint AND-decomposition of Boolean formulas. In: Voronkov A., Virbitskaite I. (eds.) *Perspectives of system informatics (PSI 2014)*. Lecture notes in computer science. Berlin, Heidelberg, Springer, 2015, vol. 8974, pp. 92–101. https://doi.org/10.1007/978-3-662-46823-4_8

13. Emelyanov P., Ponomaryov D. Cartesian decomposition in data analysis. In: *Proc. Siberian Symp. Data Science and Engineering (SSDSE 2017)*, 2017, pp. 55–60. <https://doi.org/10.1109/SSDSE.2017.8071964>
14. Emelyanov P. On two kinds of dataset decomposition. In: Shi Y., et al. *Proc. 18th Inter. Conf. Computational Science (ICCS 2018)*. Lecture notes in computer science, part II. Cham, Springer, 2018, vol. 10861, pp. 171–183. https://doi.org/10.1007/978-3-319-93701-4_13
15. Emelyanov P., Ponomaryov D. The complexity of AND-decomposition of Boolean functions. *Discrete Appl. Math.*, 2020, vol. 280, pp. 113–132. <https://doi.org/10.1016/j.dam.2019.07.005>
16. Knuth D.E. The art of computer programming. Boston, AddisonWesley Prof., 2009, vol. 4, Fascicle 1, 260 p. ISBN-10: 0321580508.
17. Emelyanov P. Factorization of multilinear polynomials over GF(2) by the reduced sorting polynomial: v. 1.1, Python (2025). <https://disk.yandex.ru/d/Urv4x0TCcPAqXQ> (accessed 2025-11-05).
18. Aziz R.A., Chu G., Muise C., Stuckey P. # \exists SAT: Projected model counting. In: *Proc. 18th Inter. Conf. Theory and Applications of Satisfiability Testing (SAT 2015)*. Lecture notes in computer science, Cham, Springer, 2015, vol. 11077, pp. 121–137. https://doi.org/10.1007/978-3-319-24318-4_10
19. Liu Zh., Chen X., Servedio R.A., Sheng Y., Xie J. Distribution-free Junta testing. *ACM Transactions on Algorithms*, 2019, vol. 15, no. 1, art. 1, p. 1–23. <https://doi.org/10.1145/3264434>

Received May 11, 2025

Revised June 20, 2025

Accepted June 27, 2025

Published online June 30, 2025

Pavel Gennadievich Emelyanov, Cand. Sci. (Phys.-Math.), A.P. Ershov Institute of Informatics Systems of the Siberia Branch of the Russian Academy of Sciences, Novosibirsk, 630090 Russia, e-mail: emelyanov@iis.nsk.su.

Cite this article as: P. G. Emelyanov. On junta problem for table-defined functions. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2025, vol. 31, no. 3, pp. 105–120.