CSCE 5300 – Introduction to Big Data and Data Science

Project "Coding, Simulation and Comparison of Big Data Related Method/Algorithm"

Name: _____

Problem Statement

Create project group consisting of approximately 5 students depending on project complexity.

Choose own original up to date an official non open access journal/conference paper (e.g., IEEE paper) not older than 5 years in Big Data related field. Use for this purpose UNT Library Internet tools. In particular, the following options can be chosen:

1. Template Examples 1 and 4 and corresponding k-WTA neural network model (4), (5) described in [1].
2. Template Examples 1 and 2 and corresponding k-WTA neural network model (8) presented in [2].
3. Template Examples 1 and 2 and corresponding k-WTA neural network model (9) , (10) presented in [2].
4. Template Examples 1 and 3 and corresponding k-WTA neural network model (4), (5) presented in [3].
5. Template Examples 1 and 4 described in [1] and corresponding k-WTA neural network model (3), (4) presented in [4].
6. Template Examples 1 and 4 described in [1] and corresponding k-WTA neural network model (6), (7) presented in [4].
7. Template Examples 1 (Fig. 3) and 2 (Fig. 5) and corresponding k-WTA neural network model (9) presented in [5].
8. Template Examples 1 and 4 described in [1] and corresponding k-WTA neural network model (3) presented in [6].

Send chosen paper to TA for approval by February 17, 23:59 PM.

After the approval, perform computer simulations of the method/algorithm presented in the chosen paper. For this purpose, write in any programming language different form one used in the paper and execute a corresponding code for the same and different input big data consisting not less than 1000 items. Use for this purpose Python/Java/C/C++ programming language. Using free data processing systems such as Hadoop and/or Dryad and/or Pregel and/or Storm and/or Spark is strongly encouraged. Compare obtained results with ones described in the paper as qualitatively as quantitatively in the form of figures and/or numerical values. Describe the following: the problem statement studied in the paper, existing methods/algorithms to solve the problem, qualitative and quantitative comparative analysis of such methods/algorithms, method/algorithm presented in the paper, results of its quantitative and qualitative comparison, advantages and limitations of the method/algorithm, possible applications of the method/algorithm, the results of own simulations and their comparison with ones presented in the paper, literature citations in square brackets and numbered in square brackets list of corresponding literature references in IEEE style [6].

Design based on obtained description related report in standard form presented, e.g., in [7], [8] or prepare corresponding presentation in PowerPoint in the form of Instructor's lecture presentations of CSCE 5300 course.

The project report/PowerPoint presentation content should have all parts indicated above.

Verify non exceeding 30% threshold of plagiarism in the project report/PowerPoint presentation by UNT Canvas Turnitin tool.

Send a draft of the project report/PowerPoint presentation and PDF files named by project group number to TAs for possible comments/suggestions by March 17, 23:59 PM.

Upload the corrected/refined final project report/PowerPoint presentation and PDF files named by project group number to the UNT CSCE 5300 Canvas environment by April 17, 23:59 PM.

Fulfill presentation online timely according to the CSCE 5300 Syllabus schedule. Presentation groups are chosen randomly by TA at the start of presentations. Each presentation should have a part related to code execution demonstration from students' laptops via Zoom on screen. The presentations can be performed in the form of videos as well. In his case, all group members should be present as in the videos as physically in the class. Each group member should present own part of the results obtained. One presentation should last approximately 15-20 min including questions and answers.

Tools to Fulfill the Project

A description of necessary tools can be found in a chosen paper, related references, PowerPoint presentations of Instructor's CSCE 5300 course and cited literature references. Additional instructions related to using Hadoop and/or Dryad and/or Pregel and/or Storm and/or Spark are provided by TA.

Remarks: additional details related to the Project including penalties for late submissions can be found in the CSCE 5300 syllabus uploaded to the UNT Canvas environment.

Final project reports/PowerPoint presentations which do not satisfy the requirements described above are accepted with reduced grades.

TA select groups for presentations by random choice at the start of presentation classes.

Students of each group are encouraged to delve into presentations of other groups, ask related questions, make reasonable comments and suggestions to get additional activity grades from TA.

References

[1] P. Tymoshchuk and D. Wunsch, "Design of a K-winners-take-all model with a binary spike train," *IEEE Trans. Syst. Man. Cyber. B, Cyber.*, vol. 49, no. 8, pp. 3131-3140, Aug. 2019.

[2] P. V. Tymoshchuk, "A model of analogue K-winners-take-all neural circuit," *Neural Networks*, vol. 42, pp. 44-61, June 2013.

[3] P. V. Tymoshchuk," A fast analogue *K*-winners-take-all neural circuit", in *Proc. Int. Joint Conf. Neural Networks*, Dallas, TX, 2013, pp. 882-889.

[4] Z. Guo and J. Wang, "Information retrieval from large data sets via multiple-winners-take-all," in *Proc. IEEE Int. Sym. Circuits Syst. (ISCAS)*, Rio de Janeiro, Brazil, 2011, pp. 2669–2672.

[5] P. V. Tymoshchuk, "A discrete-time dynamic K-winners-take-all neural circuit", Neurocomputing, vol. 72, 2009, pp. 3191-3202.

[6] L. I. Ferreira, E. Kaszkurewicz and A. Bhaya," Synthesis of a k-winners-take-all neural network using linear programming with bounded variables", in Proc. Int. Joint Conf. Neural Networks, vol. II, Portland, OR, 2003, pp. 2360 - 2365.

[7] IEEE Reference Guide, IEEE Periodicals Transactions/Journals Department, Piscataway, NJ, 2018.

[8] R. P. Lippmann, B. Gold, and M. L. Malpass, "A comparison of hamming and Hopfield neural nets for pattern classification," Technical Report TR-769, MIT Lincoln Laboratory, 1987.

[9] P. Tymoshchuk and E. Kaszkurewicz, "A winner-take-all circuit using second order neural networks as building blocks," Technical report TR-16, UFRJ NACAD Laboratory, 2002.