Annals of the Academy of Romanian Scientists
Series on Engineering Sciences
ISSN 2066-8570 Volume **13**, Number **1**/**2021**

MANAGEMENT OF BIG DATA IN SCIENTIFIC COMPUTING

Marcel ILIE[1], Augustin SEMENESCU[2]

Rezumat. *Dezvoltarile recente in sistemul de procesare al calculatoarelor au condus la generarea unor cantitati mare de date. Procesarea acestor large baze de data reprezinta adevarate provocari pentru comunintatea stiintifica. O problema particulara este interpretarea datelor, in particular seprarea datelor utile de cele inutile. Mai mult decat atat clasificarea datelor in subgrupuri de date aduce provocari suplimentare. In general este acceptat ca analiza acestor date este foarte dificila. De aceea modele eficiente si acurate de data-mining ar inlesni analiza acestor mari baze de date. In acest studiu propunem un odel computational basat pe algoritmul k-means, identificat ca algoritmul de fuzzy-clustering. Folosirea algoritmului de fuzzy-clustering reduce timpul computational cu 72% comparat cu metodele computationale normale.*

Abstract. *The recent developments in computer processing has led to the generation of significant amount of data. However, the post-processing of this large amount of data poses significant challenges for the scientific community. A particular issues is the data interpretation, particularly regarding the segregation of valuable data from arbitrary data. Moreover, data classification in subgroups poses further challenges. It has been largely accepted that the analysis of these data sets may be cumbersome. Therefore, efficient and accurate data mining models would enable the analysis if large data-sets. In this research we propose a computational model based on the k-means algorithm, identified as fuzzy clustering algorithm. The use of the fuzzy clustering algorithm reduces the computational time by 72%, when compared with regular computational approaches.*

Keywords: data management, numerical modeling, adaptive mesh refinement, k-means algorithm

1. Introduction

The rapid progress in the developments of computing technologies and software such as high-performance computing (HPC) have generated large amounts of data that require further post-processing, interpretation and dissemination [1-5]. Post-processing and visualization of this large amount of data pose significant challenges, particularly when the data needs to be assembled/coupled from various instantaneous time-frames [5, 6]. Some of these cases are encountered in the medical field, biology research, engineering applications, etc. Out of all these

[1]PhD, Assistant Professor, Dept. of Mechanical Engineering, Georgia Southern University, Statesboro, GA 30458, USA, e-mail: milie@georgiasouthern.edu
[2]PhD, Professor, Dept. of Material Sciences, University Politehnica Bucharest, Bucharest, Romania, augustin.semenescu@upb.ro