

From discrete element simulation data to process insights

Daniel N. Wilke^{1,*}, Paul W. Cleary², and Nicolin Govender^{3,4}

¹Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa

²CSIRO Data61, Private Bag 10, Clayton, South Victoria, 3169, Australia

³Research Center Pharmaceutical Engineering GmbH, Inffeldgasse 13, 8010 Graz, Austria

⁴Univ. Lille, Institut Mines-Télécom, Univ. Artois, Junia, ULR 4515 - LGCgE - Laboratoire de Génie Civil et géoEnvironnement, F-59000 Lille, France

Abstract. Industrial-scale discrete element simulations typically generate Gigabytes of data per time step, which implies that even opening a single file may require 5 - 15 minutes on conventional magnetic storage devices. Data science's inherent multi-disciplinary nature makes the extraction of useful information challenging, often leading to undiscovered details or new insights. This study explores the potential of statistical learning to identify potential regions of interest for large scale discrete element simulations. We demonstrate that our in-house knowledge discovery and data mining system (KDS) can decompose large datasets into i) regions of potential interest to the analyst, ii) multiple decompositions that highlight different aspects of the data, iii) simplify interpretation of DEM generated data by focusing attention on the interpretation of automatically decomposed regions, and iv) streamline the analysis of raw DEM data by letting the analyst control the number of decomposition and the way the decompositions are performed. Multiple decompositions can be automated in parallel and compressed, enabling agile engagement with the analyst's processed data. This study focuses on spatial and not temporal inferences.

1 Introduction

Industrial-scale discrete element simulations typically involve up to 100 million particles simulated with 100 000 to 1 000 000 time steps per process second. It is not uncommon to generate gigabytes (GB) of data per time step, e.g. storing for 100 million particles respectively three double-precision numbers per particle position, particle orientation, translational particle velocity and particle angular velocity requires 9.6 gigabytes (GB) of data. Storing this information every hundred-time steps using $\Delta t = 10^{-5}$ s requires 7.2 TB of data every second of process time. It is evident that sensibly analysing this amount of data is a daunting task. In the end, the value of the simulation relies solely on the quality of the information extracted from the simulation. This is complicated by the inherently multi-disciplinary nature of data analysis, as shown in Figure 1, requiring mastery of statistics, machine learning, pattern recognition, databases, discrete element modeling and the application domain [1, 2] applied to discrete element modeling.

A knowledge discovery and data mining (KDD) system, shown in Figure 1, enable the analyst to discover knowledge without first requiring mastery of statistics, machine learning, pattern recognition and databases. This study investigates our in-house knowledge discovery and data mining system's (KDS's) potential to decompose large datasets into

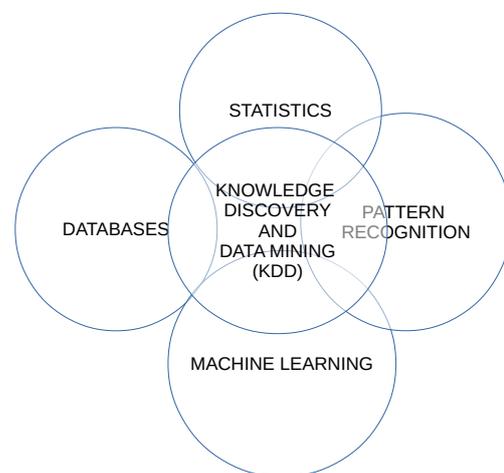


Figure 1: Multi-disciplinary nature of knowledge discovery and data mining (KDD) and data science in general.

1. regions of potential interest to the analyst,
2. multiple decompositions that highlight different aspects of the data,
3. simplify interpretation of DEM generated data by focusing attention on the interpretation of automatically decomposed regions, and
4. streamline the analysis of raw DEM data by letting the analyst control the number of decomposed domains and how the decompositions are performed.

*e-mail: nico.wilke@up.ac.za

A video is available at <https://doi.org/10.48448/qs7x-7e87>

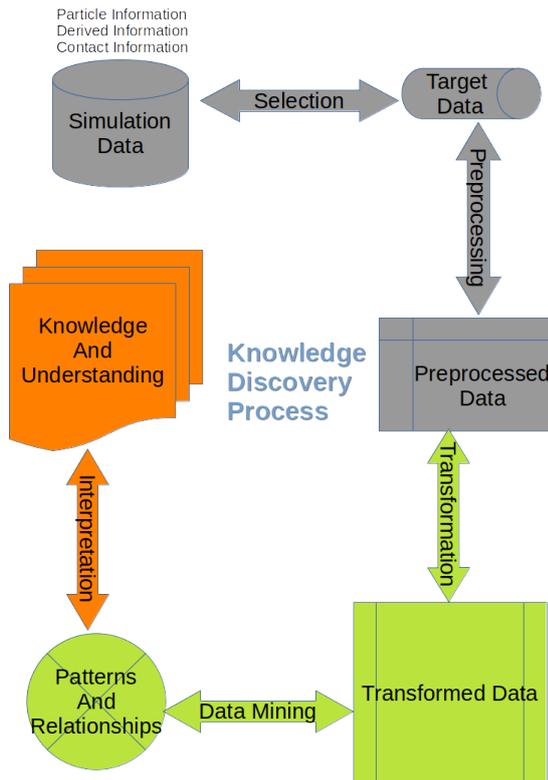


Figure 2: Outline of the Knowledge Discovery Process (KDP).

Presenting the information in a reduced dimension compresses the data, which allows for agile interactions with the data. Multiple decompositions can be performed in parallel and independent of the analyst. After that, the analyst can interpret the various compressed results for several decompositions to gain deeper understanding and insight from the data. This enables researchers, scientists, engineers and analysts to focus their attention on interpreting domains of interest.

The application of statistical and machine learning in this study is distinct from the conventional discrete element simulation acceleration studies that have been applied to powder spreading of metal additive manufacturing [3], modelling of bulldozer shovelling [4], and finite element applications [5]. Acceleration studies are based on supervised learning for which a regression or interpolation function of the response manifold is constructed. The advantage of acceleration studies is that many strategies are available, while model training is significantly simplified due to training’s supervised nature. A supervised learning approach’s disadvantage is that the models’ training has to be redone every time the application changes.

We developed an initial KDD framework for existing and new granular processes, where we do not opt for conventional supervised learning approaches. The KDD framework’s value is to identify various data interpretations that may enable new knowledge and insights to be uncovered. This is inherently an unsupervised learning problem. Besides, numerous applied science domains can significantly benefit from such a KDD framework as it could easily be adapted and utilised to better inform practi-

tioners within those domains. Therefore, this study solely employs unsupervised learning approaches for utilisation independent of the simulated application.

First, we describe KDD conceptually, without paying much attention to the algorithmic details as we give an overview of the proposed process in Section 2. A numerical study and sample decomposition are performed in Section 3, for the ball mill simulation. Finally, conclusions are offered in Section 4.

2 Knowledge discovery and data mining (KDD)

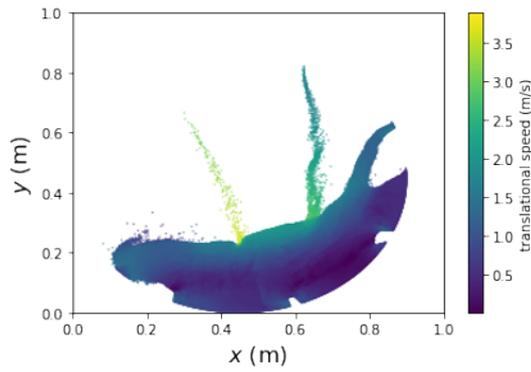
Knowledge discovery and data mining (KDD) is a recent sub-field of statistics which was established in 1995 with the First International Conference on Data Mining and Knowledge Discovery (KDD-95) in Montreal under the sponsorship of the Association for the Advancement of Artificial Intelligence (AAAI). The ultimate goal of KDD is to extract high-level knowledge from low-level data. This study develops a KDD framework to achieve the data mining objective outlined in Definition 1.

Definition 1. Data mining objective is to extract likely high-level knowledge for discrete element simulation data requiring minimal user input.

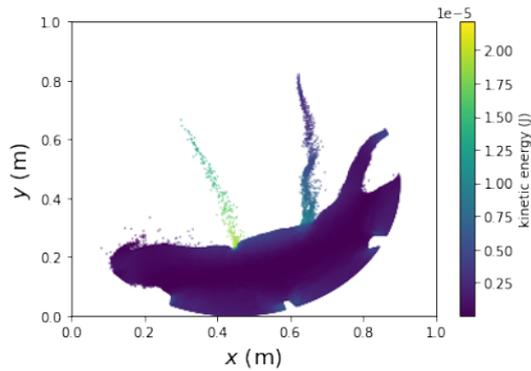
To achieve this, we ensure we conduct KDD within the context of discrete element simulation data that allows for high-level abstraction to be embedded into the framework. Given the raw simulation data that include particle information (geometric information, constitutive information, position, orientation, translational velocity, rotational velocity, damage, etc.), derived information (translational kinetic energy, rotational kinetic energy, coordination number, etc.) and contact information (shear contact energy, normal contact energy, force chains, etc.) several target datasets are selected, with the focus of this study being spatial data, i.e. data at some time. This study aims to understand and explore the spatial distribution of information by employing statistical approaches, pattern recognition, machine learning (ML) on these target datasets, as outlined in Figure 2. Although KDD generally relies on several data mining techniques that include inductive learning, Bayesian statistics, semantic query optimisation, knowledge acquisition for expert systems and information theory, we limit ourselves to unsupervised strategies. We develop a knowledge discovery system (KDS) that

1. decomposes a spatial domain into informative domains, that
2. can find multiple decompositions.

This enables us to uncover multiple viewpoints on the same discrete element simulation data, allowing practical knowledge and insight discovery about a process. An ensemble of unsupervised approaches is utilised in our KDS to distil information presented for the analyst’s ease of interpretation.

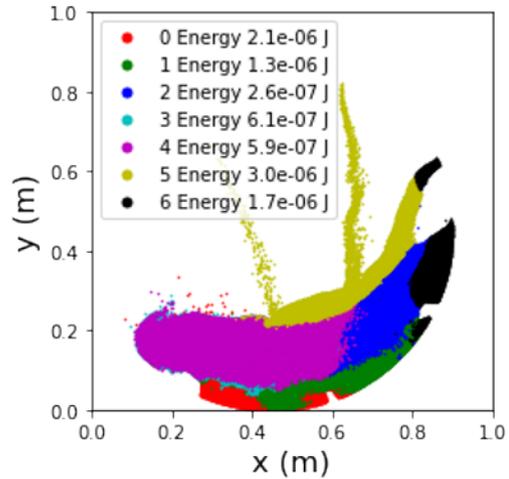


(a) Translational speed (m/s)

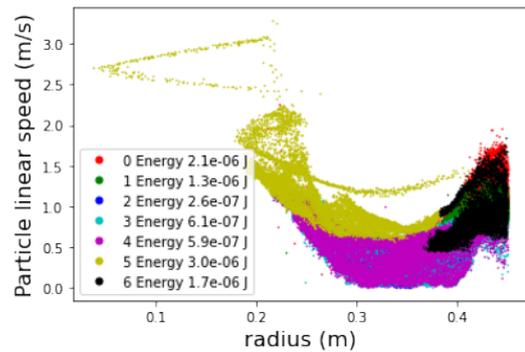


(b) Kinetic energy (J)

Figure 3: Standard spatial visualisation of raw information.



(a) Conventional spatial visualisation



(b) Non-conventional spatial visualisation

Figure 5: (a) Conventional spatial visualisation and (b) non-conventional phase-space visualisation of standard or corroborative information. Total kinetic energy for each spatial domain is indicated.

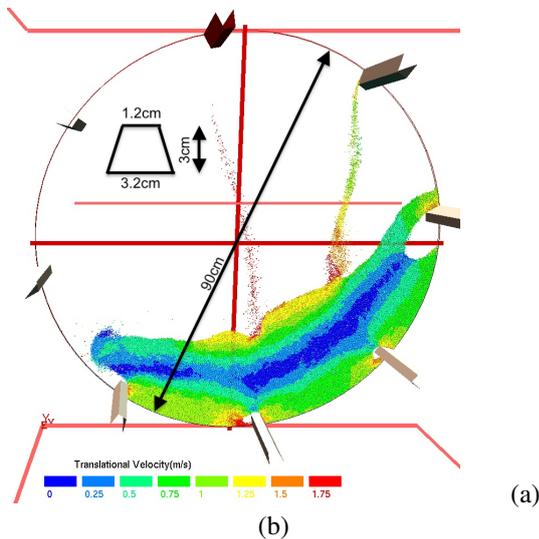


Figure 4: Ball mill modelled processes for this study with translation speed (m/s) indicated in colour.

3 Numerical Study

To demonstrate knowledge discovery from simulation data, we consider the simulation of a 90cm ball mill rotating at 36 revolutions per minute (RPM), shown in Figure 4. Only particle positions, translational velocities and rota-

tional velocities at a single time step are considered. For compactness, only a sample decomposition is presented using several visualisations. The ball mill simulation particles form a unified particle domain, which allows us to interrogate the KDS for dense particle systems.

The 90cm ball mill slice modelled with 2mm length, using eight trapezoidal lifters as indicated in Figure 4. The mill is filled with 1 million spherical particles of 1mm diameter. The spherical particles are modelled as steel particles with a density of $7850 \frac{\text{kg}}{\text{m}^3}$, effective particle-particle static and dynamic friction of 0.5 and a coefficient of restitution (COR) of 0.4. The particle-lifter static and dynamic friction are modelled using 0.55 and 0.65 for the COR, while the particle-shell were modelled using static and kinetic friction of 0.45 and 0.7 for the COR. The simulation was conducted using a time-step of 5×10^{-5} s.

Information presented as conventional simulation data informs the analyst for corroboration, interrogation, verification and validation. Consider the visualisation of aspects of raw particle simulation information in Figure 3(a) and derived information in Figure 3(b). For a 90cm diame-

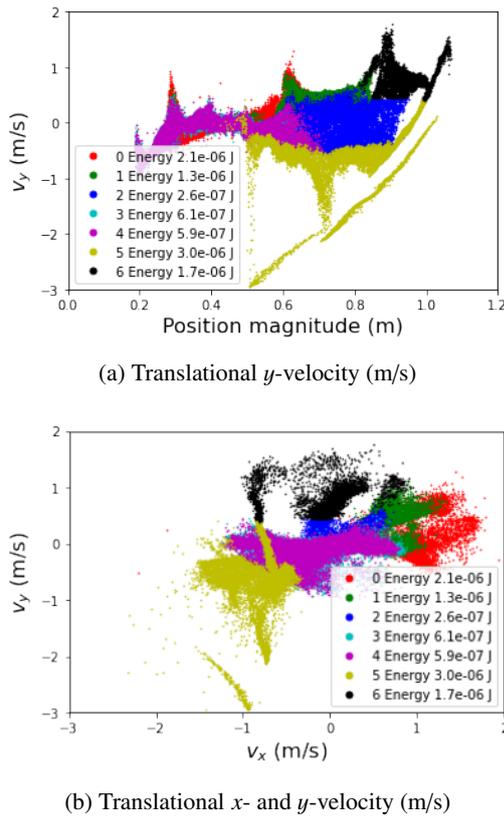


Figure 6: Non-conventional spatial visualisation of standard or corroborative information. Total kinetic energy for each spatial domain is indicated.

ter ball mill travelling at 36 rpm or 3.8 radians per second anti-clockwise, the translational (or linear) speed of the shell is 1.7m/s, which is easily corroborated.

The KDS offers multiple decompositions of information to identify several informative viewpoints on the same data. A sample decomposition into seven domains is shown in Figure 5(a) that the KDS automatically identified and the associated kinetic energy for each domain shown. The information is further explored in a basic phase-space visualisation in Figure 5(b). The phase-space's granularity and visualisation are extended in Figure 6(a)-(b). Here, for the low kinetic energy recirculation zone, changes in flow directions are visible.

Let's consider an obvious but illustrative example. Consider the spatial domain 5, that has the highest total kinetic energy, one decade higher than domain 2. Information can be quickly corroborated, cross-referenced and explained. Domain 5 is the result of cataracting media, the translational kinetic energy is primarily concentrated in a downward vertical velocity and supported by a horizontal velocity component to the left, as derived from Figures 6(a) and (b). In turn, domain 2 has a horizontal

velocity component to the right due to the anti-clockwise rotation with a smaller horizontal velocity component to the left due to recirculation at the shoulder. The lowest translational speeds are domains 2 and 4, also having the lowest x and y speed components as it predominantly captures the recirculation zone. Zones 0, 1 and 7 has a maximum translational speeds equal to the shell speed, with an apparent linear reduction in maximum translational speed with a decrease in radius.

The KDS is designed to be a flexible environment. It allows the user to change the target data, the number of decompositions and decomposing strategies that can be mined automatically and in parallel. The analyst can cycle through the various decompositions to verify simulations and gain additional insight into the process. Lazy compression is employed through sub-sampling, with current research being done to automate domain descriptions for efficient compression of decompositions.

4 Conclusion

This study demonstrated the benefits of having a knowledge discovery system (KDS) to explore simulation data, specifically, discrete element simulation data. Qualitative and quantitative information is presented in standard and non-standard ways allowing the analyst to examine information over the spatial or phase-space domains.

We demonstrated that our in-house KDS could decompose large datasets into i) regions of potential interest to the analyst, ii) multiple decompositions that highlight different aspects of the data, iii) simplify interpretation of DEM generated data by focusing attention on the interpretation of automatically decomposed regions, and iv) streamline the analysis of raw DEM data by letting the analyst control the number of decomposition and the way the decompositions are performed.

In the end, it is an important reminder that the value of a simulation relies solely on the quality of the information extracted from that simulation.

References

- [1] F. Fleissner, T. Gaugele, P. Eberhard, *Multibody System Dynamics* **18**, 81-94, (2007)
- [2] P.W. Cleary, *Particuology* **8**, 106-118, (2010)
- [3] W. Zhang, A. Mehta, P.S. Desai, C.F. Higgs, *International Solid Free Form Fabrication Symposium*, Austin, TX. pp. 1235-1249, (2017)
- [4] S. Ulin, Ph.D. thesis, Umea University (2018), <http://umu.diva-portal.org/smash/get/diva2:1254164/FULLTEXT01.pdf>
- [5] B. Bohn, J. Garcke, R. Iza-Teran, A. Paprotny, B. Pehrerstorfer, U. Schepsmeier, C.A. Thole, *Procedia Computer Science* **18**, 621-630, (2013)