Machine Learning and Data Analysis Using Posets: A Survey

Arnauld Mesinga Mwafise

Abstract

Posets are discrete mathematical structures which are ubiquitous in a broad range of data analysis and machine learning applications. Research connecting posets to the data science domain has been ongoing for many years. In this paper, a comprehensive review of a wide range of studies on data analysis and machine learning using posets are examined in terms of their theory, algorithms and applications. In addition, the applied lattice theory domain of formal concept analysis will also be highlighted in terms of its machine learning applications.

Keywords— posets, machine learning, data analysis, formal concept analysis, lattice theory

1 Introduction

Data analysis plays a pivotal role in today's data-driven world. Machine learning is used to automate the data analysis process, and enables the extraction of valuable insights by making accurate predictions from large complex datasets. Data often have complex structure which can usually be endowed with a natural order. Such complex data can be derived from text, social and behavioral sciences, chemical structures, biological structures, images, videos, governmental and business settings, and so forth. Mathematical techniques and concepts serve as powerful tools in the development of new algorithms and methods for machine learning and data analysis. For some time now, many studies based on the use of partially ordered sets(posets) for machine learning and data analysis have emerged. Furthermore, over this time the literature on these topics has significantly expanded in volume, scope, and the range of its applications in various fields. However, no recent surveys exist to collect and organize this knowledge, thereby impeding the ability of researchers and machine learning scientists alike to utilize it for various data science or artificial intelligence tasks. In order to address this gap, we present an up-to-date overview with an emphasis on the most prominent and currently relevant works on data analysis and machine learning methods in which the role of posets has been established. By covering earlier works as well as more recent advances, this survey aims to provide researchers, applied mathematicians, data scientists, statisticians, social scientists and practitioners new to the field as well as all those aiming at keeping pace with innovations in this interesting, growing and promising research field with a solid understanding of the different conceptual and methodological approaches which incorporates posets into data analytics or machine learning applications.

The rest of this survey is organized as follows. Section 2 introduces and outlines the key concepts on poset theory and lattice theory. Section 3 presents a summary of research studies on machine learning and deep learning with respect to poset theory, lattice theory and formal concept analysis. Section 4 focuses on cluster analysis using posets and formal concept analysis methods. Sections 5-7 presents a summary on data analysis from a poset theory perspective covering its multidimensional, exploratory and descriptive aspects respectively. Section 8 presents a collection of applications across various domains including software packages, datasets and algorithms selected for this purpose. Section 9 discusses possible future research directions. Section 10 summarizes the paper.

2 Basics of Partial Order Theory

Order theory is a fundamental branch of mathematics which focuses on the arrangement of elements within various structures based on certain rules that can intuitively be captured using binary relations. It plays a crucial role in understanding mathematical sequences, hierarchies, and the organisation of data in fields such as computer science, economics, social sciences, environmental sciences, biomedical sciences etc. Generally, order theory deals

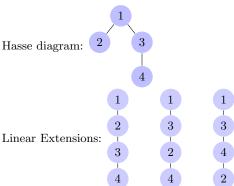
with the structure and properties of partial orders which occurs naturally in subset relations and integer relations. The analysis of partially ordered sets is a well studied topic in discrete mathematics particularly in combinatorics [247]. The key concept in partial order theory is the 'concept of comparison'. Objects are mutually compared, and transitivity is assumed [83]. That is if object x is better than object y, and object y is better than object z then object x is better than object z. However, comparison can be independently possible without the requirement of transitivity in some scenarios. For instance, in the theory of tournaments in sports, it is possible for team A to beat team B and team B beats team C without necessarily implying that team A beats team C. Mathematically, we can summarize the definition of partial order sets as follows. A nonempty set X of size n which is endowed with the order relation "<" (comparison operator) is called a partially ordered set or simply a poset, denoted by P = (X, <), if the relation < satisfies the following conditions:

- 1. reflexive i.e. $x \in X, x \leq x$;
- 2. antisymmetric i.e. $x, y \in X$ if $x \le y$ and $y \le x \implies x = y$;
- 3. transitive i.e. $x, y, z \in X$ if $x \le y$ and $y \le z \implies x \le z$.

Moreover, reflexivity means that a given object can be compared with itself. Anti-symmetry means that if both comparisons are valid, i.e., y is better than x and at the same time, x is better than y, then this axiom requires that x is identical to y. Transitivity means that if the objects are characterized by properties which are at least ordinal scaled, then any measurable quantity like height, length, price, tournament team rankings, order of product quality, order of agreement or satisfaction etc., implies transitivity [84]. If either $x \le y$ or $y \le x$ then x and y are comparable. For partially ordered sets it is not required that all pairs $x, y \in X$ are comparable either as $x \le y$ or $y \le x$. In the case where all pairs $x, y \in X$ are comparable, then the set X is referred to as a totally ordered set or an ordered set. The numerical comparison operator ≤ is used to denote the relationship between any two pairs of elements(objects) within an ordered set. With an ordered set, ranking is always possible and any sorting algorithm may be applied using \leq as the comparison operator for choosing whether to swap the objects in an ordered list. Classical problems of sorting and searching assume an underlying linear ordering of the objects being compared. An important result in order theory is that every partial order can be extended to a linear ordering or a total order. In order theory, a linear extension of a partial order is a total order (or linear order) that is compatible with the partial order. More formally, a linear extension [208, 250] of a partially ordered set P is a permutation of the elements $p_1, p_2, p_3, ...$ of P such that $p_i < p_j$ implies i < j. For example, the linear extensions of the partially ordered set ((1,2),(3,4)) are 1234,1324,1342,3124,3142, and 3412, all of which have 1 before 2 and 3 before 4. Some key concepts on partially ordered sets $P = (X, \leq)$ are described as follows:

- Maximal elements (or objects) of the poset P are the set of elements $x \in X$ in which no other element $y \in X$ satisfies the relation y > x. If x is the only maximal element then it is referred to as the "greatest" element. In a totally ordered set, the terms maximal element and greatest element coincide.
- Minimal elements (or objects) of the poset P are the set of elements $x \in X$ in which no other element $y \in X$ exist such that y < x. If x is the only minimal element, then x is referred to as the "least" element. In a totally ordered set, the terms minimal element and least element coincide.
- Chain is a subset C of X, in which any object (or element) is mutually comparable with all other elements of C. That is, a chain C is a poset such that no element can be added between any two of its elements without losing the property of being totally ordered.
- Antichain is a subset C' of X, in which each object (or element) of C' is mutually incomparable with all other elements in C'. That is, all the elements of the set C' are pairwise incomparable and as such they can never have the property of being totally ordered.
- A cover relation for $\mathcal{P} = (X, \leq)$ is the set of pairs (x, y) such that $x, y \in X$, and y covers x whenever there exist no element $z \in X$ such that x < z < y.

Partially ordered sets are graphically visualized using Hasse diagrams. A Hasse diagram can be derived from a directed acyclic graph, where the vertices are representing the objects and a line relates object x with y whenever x < y. In the case of a transitivity relation whenever x < y and y < z then it suffices to draw a line only for x < y and y < z and not for x < z. Hasse diagrams can be viewed as a powerful tool for unifying ideas and concepts. Hasse diagrams reveal an intricate network of comparabilities and incomparabilities, maximal and minimal elements. In addition, Hasse diagrams are also characterized with a structure that comprises levels, chains and antichains.



Finite posets can also be represented by square matrices which comprise: incidence [98, 223], cover [7], and mutual ranking probability matrices [131]. The mutual ranking probability matrices are a class of matrices which comprise and convey information on the dominance among statistical units[131]. All the three main types of poset matrices are useful in conveying essential information on the structure of the order relation from different perspectives. Posets can also be subdivided into variour classes such as series-parallel posets, semantic posets, and ranking posets etc. Series-parallel posets are characterized by the N-free partial order structure. Series-parallel partial orders have been applied to machine learning of event sequencing in time series data [212]. Semantic posets are used for compositional generalization in semantic parsing [150]. The ranking poset \mathfrak{M}_n [204] is the poset in which the elements are all possible cosets $\mathfrak{G}_{\lambda}\pi$, where λ is an ordered partition of n and $\pi \in \mathfrak{G}_n$. The ranking poset is useful in model selection and validation in machine learning.

Posets are natural models for many statistical applications [38, 272, 235, 270]. Partial orders are also the natural mathematical structure for comparing multivariate data that lacks a natural order [234]. For instance, multivariate data on colours has no natural ordering and therefore the rank features in colour spaces can be modelled as partial orderings, and this approach is generalizable in similar contexts [257]. It also plays a key role in decision making in environmental informatics and computational chemistry [52, 156, 192, 194, 193, 55, 63], social sciences [28, 67] etc. Moreover, in data analysis a typical binary relation is defined for a set of objects and a set of attributes. For instance, objects are transactions in a supermarket, attributes are items in the supermarket, and the relation consists of pairs (transaction, item) such that the item occurs in the transaction [203].

The study of partially ordered sets and lattices has been covered by a significant number of mathematical publications. Birkhoffs (1967) [37] and Grätzer's (1978)[149] books on lattice theory are considered to be classics. Lattices[297] are a broad structured class of posets. Lattice theory has shown its ability to contribute to solving problems in a wide variety of applications. Formally, a lattice[147] is a poset, in which every pair of elements has both a least upper bound and a greatest lower bound. In other words, it is a structure with two binary operations: join and meet. The following sub-definitions summarize a lattice structure. A meet semilattice is a poset for which any two elements a and b have a greatest lower bound denoted $a \wedge b$. The greatest lower bound of a and b is the largest element that is still less than both of them. In a lattice, the greatest lower bound must be unique. The greatest lower bound of a and b is also called the meet or infimum of a and b. A join semilattice is a poset for which any two elements a and b have a least upper bound, denoted $a \vee b$. The least upper bound of a and b is the smallest element that is still greater than both. In a lattice, the least upper bound must be unique. The least upper bound of a and b is also called the join or supremum of a and b. If a poset is both a meet semilattice and a join semilattice, then the poset is also a lattice.



In the example above: The maximal elemnt is 4 and the greatest element is 4. Similarly, the minimal element is 1 and the least element is 1.

3 Machine learning using posets and formal concept analysis

3.1 Overview

In general, machine learning involves the creation of knowledge from a training dataset by using models and algorithms to learn from data in order to make predictions, find patterns, classify data and automate decision making processes. Machine learning has progressed significantly over the past two decades, and this has been driven by the development of new learning algorithms, theory, and high performance computing architecture [145]. There are many different kinds of machine learning algorithms. The most well-known ones are supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning is useful in scenarios for which some relationship between input and output labelled data has been established, unsupervised learning is effective for uncovering hidden patterns in unlabelled data, and semi-supervised learning utilizes a combination of labelled and unlabelled data to train models. The goal of any machine learning algorithm is to optimize the performance of a system when handling new instances of data through user defined programming logic in a given environment [29]. In order to build more efficient machine learning algorithms or models so as to improve on their performance and accuracy in most predictive task, various ensemble learning methods have been developed. Ensemble learning [246] is an approach which aggregates the predictions of two or more models fitted to the same data in order to minimize error. In the area of supervised learning, the most popular ensemble machine learning techniques are bagging, boosting, and stacking, which use multiple learning algorithms or models to produce one optimal predictive model. Ensemble methods can broadly be categorized into sequential ensemble techniques and parallel ensemble techniques. Bagging(bootstrap aggregating) aims to reduce variance by adopting parallel ensemble learning on homogenous weak learners, boosting aims to reduce bias by adopting sequencial ensemble learning on homogenous weak learners, and stacking aims to improve prediction accuracy by adopting parallel ensemble learning on heterogeneous weak learners. Sequential ensemble techniques generate base learners in a sequence that are characterized by the dependence between the base learners. On the other hand the parallel ensemble technique base learners are generated in parallel, and in such a way that independence is enforced between the base learners. Dagging (Disjoint samples aggregating) is another useful type of parallel ensemble technique that was introduced in [280]. Dagging is similar to bagging, but instead of bootstrapping, it uses stratified sampling by creating a number of disjoint groups and stratified data from the original learning data set, with each considered as a subset of learning. In unsupervised learning, ensemble clustering [306, 19] aims at combining the outputs of several clustering algorithms to form a single clustering structure (crisp or fuzzy partition, hierarchy).

In recent years, deep learning [114] has been the most popular computational approach in the field of machine learning and has unique advantages when dealing with high-dimensional nonlinear problems. The performance of machine learning algorithms relies heavily on the representation, and finding a good representation can facilitate the discovery of structure in the input data by the learning algorithm. Deep learning as a model is a specific type of representation learning[33] that uses layered algorithms known as artificial neural networks, which attempts to mimic the human brain through a combination of data inputs, weights, and biases, to learn representations of data. It has succesfully been used in a wide range of disciplines such as cybersecurity, natural language processing, visual recognition, machine translation, robotics, recommendation systems etc. There are several types of deep learning algorithms. These are: Convolutional Neural Network (CNN), Long Short Term Memory, Graph Neural Networks etc. Traditionally, machine-learning and deep learning algorithms have been used on data represented in Euclidean space such as image data which can be represented as a regular grid of pixel values. On the other hand, graph data cannot be represented on Euclidean space [20]. Graph Neural Networks (GNNs)[299] are a class of deep learning methods designed to perform inference on data described by graphs which are non-Euclidean (non-metric) structured. They are highly influenced by CNN which learns features by inspecting neighboring pixels (nodes in GNN) in three dimensional image data for classification and object recognition purposes. Posets are special classes of directed graphs. There is also a growing use of graph neural networks for the modeling and analysis of partially ordered data [132, 116, 150, 295, 304]. More recently, the problem of machine learning on meet/join lattices and posets was extensively studied in [296]. The main focus of the study was on developing methods that are based on generalized convolutions and sparse Fourier transforms algorithms which are capable of learning set functions. The first lattice-based machine learning models were introduced by V.K. Finn[137, 136, 135]. It was based on a closure system which uses the JSM(John Stuart Mille)-method of automated hypothesis generation. In this model, positive hypotheses are searched among intersections (similarity as meet operations) of positive example descriptions (object intents), likewise for negative hypotheses [21]. The JSM method also identifies data patterns by means of induction and it can also be used as a logical rule-based classification method formulated in terms of formal concepts [200]. Machine learning has been studied from many perspectives in the context of poset theory and lattice theory. These include: machine learning performance comparison[39, 204], multilabel classification [109], ensemble classification[204], sequential clasification [212, 273, 272, 278]; natural language processing [116, 210, 150], deep unsupervised learning [31], semi-supervised learning [132], learning to rank[96], time series modelling [228, 212], ensemble clustering [109], model selection[270], topological deep learning [155], attention-based neural networks [154].

3.2 Performance Comparison and Model Selection

Performance comparison of different learning algorithms is a very common task in the field of machine learning. The evaluation of several machine learning algorithms based on the concept of depth functions was recently introduced in [39]. Depth functions [312] were originally formulated to extend the univariate notions of median, quantiles, ranks, signs, and order statistics to the setting of multivariate data. Blocher & Schollmeyer [39] uses an adaptation of the well-known simplicial depth, and then applied it to the set of all partial orders referred to as the union-free generic (ufg) depth. The purpose is to develop a framework to show how samples of poset-valued random variables can be descriptively analyzed by utilizing a generalized concept of data depth based on the concept of depth functions. A key property of the ufg depth is that it considers partial orders as a whole, and not limited to pairwise comparisons, since as the procedure for computing the depth of a partial order cannot be broken down using simple sum statistics, but requires the partial order as a holistic entity. The ufg depth for a comparison of machine learning algorithms is based on multidimensional performance measures. It works by comparing the various algorithms not with respect to one unidimensional criterion such as balanced accuracy, but instead it considers a whole set of performance measures. In Taeb et al. [270], a methodology is presented for organizing classes of models as partially ordered sets in order to address the challenges of model structures which are not characterized by an underlying Boolean logical structure, and which is a prerequisite to formalize and control the false-positive error.

3.3 Natural Language Processing

In natural language processing, compositionality [148] describes the ability to easily recombine familiar units like words into new phrases and sentences. In recent years, this has been the focus of intense interest in artificial intelligence. However, existing neural network based models have proven to be extremely deficient in such a capability [105]. In order to improve on the compositional generalization ability of neural encoder-decoder architectures for natural language understanding, Guo et al. [150] introduce a hierarchical poset decoding paradigm with the goal of taking into consideration the partial permutation invariance and the poset structure of semantics. Compositional generalization is important in tasks that can involve complex natural language utterances, which can be associated to many equivalent meaning representations. One of such task is dependency parsing. Semantic dependency parsing [199] is the task of mapping natural language sentences into representations of their meaning in the form of directed graphs on words. Dependency parsing is generally used to identify semantic relations between words in a sentence by analyzing their grammatical structure. Dyer [116] proposes a method for converting dependency trees to surface orders using syntactic word embeddings and edge-weighted posets. Graph neural network were employed to learn the edge weights of the poset representing the surface order, and the training was carried out on the Universal Dependencies (UD) corpora.

3.4 Classification

Most major studies in artificial intelligence has significantly concentrated on developing methods for classification. Classification in machine learning involves prediction tasks on discrete target variable. It is common to find learning tasks that involve ranking of objects, particularly in domains that generate relational data where the goal of inferring is conditioned on a particular target object. In classification, a single label $y \in Y$ can be associated with a covariate x. A generalization of this problem is conditional ranking, the task of assigning to x a full or partial ranking of the items in Y. Lebanon and Lafferty[204] propose a unifying algebraic framework for ensemble methods of classification and conditional ranking, focusing on the case of discrete and ranked inputs. Their approach is based on the ranking poset having a fixed size of items, and which consists of the collection of all

full and partial rankings endowed with a partial order given by refinement of rankings. The structure of the poset of partial ranking over Y gives rise to natural invariant distance functions which generalizes Kendall's Tau[188] and the Hamming distance. Using the invariance properties of the distances, a generative interpretation of the probabilistic model is derived, and this is suggested to be useful in model selection and validation [204]. Zhang[254] presents a framework for ranking with classification rules. Classification rules can be generated using inductive learning algorithms[6] on data. Sahami[254] introduces an inductive machine learning algorithm called Ruleamer, which generates a set of symbolic classification rules as output using lattices. The Ruleamer algorithm takes as input (i) a lattice, L, (ii) a set of instance classification labelings, C, which corresponds to the instance nodes in L, and (iii) a noise parameter, N, indicating a percentage by which each induced rule can misclassify some portion of the training instances to which it applies. Learning classification rules has also been extended to sequence data [102, 101]. The problem of sequential classification is considered when classification states form lattices and response distributions are experiment-specific. Tatsuoka[278] proposes a Bayesian framework for sequential classification on finite lattice models in which response distributions are allowed to vary according to experiment. When the models are latent and complex, such as in cognitive applications, it becomes imperative to have available a variety of data analytic tools for fitting the models, and for the validation of assumptions that are made regarding the specification of class conditional response distributions for experiments. A framework for model fitting and analysis of experiments has been developed by Tatsuoka [272]. This led to the formulation of the theory on asymptotically optimal sequential selection of experiments in [273, 274], which can be applied to Bayesian classification problems when the parameter space is finitely partially ordered. In general, lattices can be used as classification models when it is natural to consider classification states as partially ordered. Examples include cognitive modeling, such as in neuropsychological (NP) assessment, educational testing, and in group testing data.

An important task involving supervised machine learning techniques is classifying content employed in many contexts. Multi-label classification[36] deals with the problem in which each instance can be associated with a set of class labels. But a drawback on the application of multilabel classification in many real-world applications arises when classes overlap in the feature space or the number of class labels becomes very large resulting to an exponentially growing number of label combinations. As such it becomes computationally inefficient to use most multi-label classification algorithms in such scenarios. Multilabel classification has been suggested as a powerful tool for building AI systems that can handle complex uncertain scenarios. Uncertainty is a pervasive aspect of many AI systems, as they often deal with incomplete or conflicting information caused by the variability in data and models. The mathematical theory of belief functions as applied in the Dempster-Shafer theory (DST) introduced in [260, 108], provides a generalized framework which blends the concept of probability associated with uncertainty and the notion of sets related to logic. However, the computational complexity of DST increases significantly when confronted with a substantial number of events or sources of evidence, resulting in potential performance challenges caused by the exponential complexity of it's operations. In addition, problems which involve learning instances belonging to several classes (frame of discernment) simultaneously which may pose several difficulties in obtaining the right label combination during classification. Denoeux[109] suggests that when the frame of discernment has a lattice structure, it is possible to restrict the set of events to intervals in that lattice. This approach involving lattices leads to the application of Dempster-Shafer theory to computationally demanding machine Learning tasks such as multi-label classification and ensemble clustering.

3.5 Deep Learning

In this section, the use of posets in deep unsupervised learning, attention-based neural networks, and topological deep learning are reviewed. Deep unsupervised learning[186] is an emerging discipline in the field of artificial intelligence and machine learning. It can serve as an alternative to supervised approaches in representation learning. In particular, unlabelled data is much cheaper to obtain which is a key advantage when adopting deep unsupervised learning for complex processing tasks involving large datasets. One of such task is visual similarities learning[95] in the field of computer vision. Various approaches have been adopted for visual similarity learning such as the use of convolution neural networks(CNN) [294]. But this approach has a major limitation in that it requires millions of samples of supervised training in order to achieve a good performance. Bautista et al.[31], propose an unsupervised approach to visual similarity learning based on CNNs by framing it as a combination of surrogatee (i.e. artificially created) classification tasks and poset ordering. This approach which combines classification with a partial ordering of samples resolved some of the major problems of visual similarity learning

such as:

- a large number of training samples are not assigned to any of the compact surrogate classes since their mutual similarities or dissimilarities cannot be easily established;
- inability to perform joint optimization of the different classification tasks resulting to mutually conflicting relationships since transitivity cannot be established;
- expensive cost or difficulty in obtaining a large number of samples for training.

Neural attention models[222] focus on specific regions of images instead of the entire picture, originated in an attempt to reduce the computational complexity of image processing while improving performance. In general, the application of attention neural models in the field of computer vision was motivated from the idea of mimicking human attention. Many different neural attention models [256, 41] are now available and have been a very active research area over the past decade. Attention is an important mechanism that can be employed for a variety of deep learning models across many different domains and tasks. In natural language processing NLP, transformers and attention have been utilized successfully in a plethora of tasks including reading comprehension, abstractive summarization, word completion, and so forth. Hajij et al. [154] introduces higher-order attention networks (HOANs), as a special class of attention-based neural networks defined on a generalized higher-order domain called a combinatorial complex (CC). CCs define a structure that bridges the gap between simplicial/cell complexes and hypergraphs. In addition, CC is a poset; i.e., a partially ordered set with partial order relation given by set inclusion. Specifically, Hajij et al.[154] demonstrate the reducibility of any CC to a Hasse graph, allowing the characterization of certain computational and conceptual aspects of HOANs in terms of graph-based models. Furthermore, the Hasse graph is used to provide a definition of equivarence for HOANs. The developments of HOANs subsequently influenced the study of topological deep learning with respect to combinatorial complexes. Topological Deep Learning (TDL)[311] is a rapidly evolving field which provides a comprehensive framework that combines the principles of topological data analysis (TDA) with deep learning techniques. In particular, it uses topological features to understand and design deep learning models. TDL originated from the pioneering works on Topological Signal Processing (TSP)[26] that showed the benefits of considering higher-order (multiway) relationships among data points. The growing interest driving the current developments in topological deep learning is attributed to the fact that many systems are characterized by higher-order interactions that cannot be captured by the intrinsically pairwise structure of graphs. Recently, Hajij et al. [155] developed a general TDL framework along with the introduction of an abstract class of neural networks called combinatorial complex neural networks (CCNNs). CCNNs can also be viewed as a framework that generalizes many popular architectures, such as convolutional and attention-based neural networks. A CC-based neural network is reducible to graph-based models and is capable of exploiting all neighborhood matrices or a subset of them. The CC structure class is determined by the incidence matrices, adjacency matrices, coadjacency matrices (Proposition 8.1.[155]). Combinatorial complexes (CCs)[30] are able to handle both hierarchical and set-type higher-order interactions. Hasse graph of the CC, describes the poset structure between cells. The structure of combinatorial complexes permits graph-based pooling, facilitate message passing of topological features in deep learning, flexible modeling of relations among relations. Based on this structural benefits, Hajij et al. [155], present the Hasse graph interpretation of CCNNs as follows. (i) It is shown that any CCNN-based computational model can be realized as a message-passing scheme over a subgraph of the augmented Hasse graph of the underlying CC; (ii) A tensor diagram represents a CCNN via a directed graph. Any computation on tensor diagrams CCNN is realizable on augmented Hasse graphs.

3.6 Semi-Supervised Learning

Semi-supervised learning[286] is a branch of machine learning that bridges supervised and unsupervised learning by combining both labelled and unlabelled data to train artificial intelligence (AI) models for classification or regression. Semi-supervised classification methods are particularly relevant in scenarios where labelled data is scarce, expensive or difficult to obtain. Graph-based semi-supervised learning methods[264] have demonstrated their benefits in various domains due to their uniqueness of structure, the universality of applications, and their scalability to large scale data. Graph-based learning methods explicitly considers the relations between two entities (i.e. vertices). In an ordinary graph an edge connects exactly two vertices. On the other hand, a hypergraph represents a generalization of a graph in which an edge can join any number of vertices. But in most settings, hypergraph structure represents the hyperedge as an unordered set of vertices without taking into consideration

the possibility of ordering relationships among vertices which is inherent in most real-world data. This motivated Feng et al. [132] to suggest two related solutions which are:

- the construction of a new data structure named Partial-Order Hypergraph, which specifically injects the partial ordering relations among vertices into a hyperedge;
- the development of regularization based learning theories for partial-order hypergraphs which generalizes conventional hypergraph learning by incorporating logical rules that encode the partial-order relations.

Furthermore, graph convolution network(GCN)[191] was applied in a semi-supervised setting for learning over the constructed partial-order hypergraphs.

3.7 Time Series Modeling

Many machine learning and statistical models have been suggested for time series modeling of sequential data [302]. One way of generating sequence data is from a series of discrete events that occur over time. Event sequence data arise in many applications, such as web browsing, e-commerce, process monitoring etc. An important problem in mining sets of sequences of events is to get an overview of the ordering relationships in the data. Mannila & Meek[212] introduce a method for discovering partial orders from a sequence of events. The main idea of the method is to view a partial order as a generative model for sequences and by describing a set of sequences by means of mixture models of partial orders. In addition, it is also shown that the likelihood of a given partial order is inversely proportional to the number of total orders compatible with the partial order. The computation of the number of linear extensions is restricted to series parallel posets which are computationally efficient. A learning algorithm is constructed based on greedy search over partial orders defined on the set of possible events. This learning algorithm determines if a sequence is compatible with a partial order in linear time. Nicholls et al [228] explore the Bayesian inference for partial orders from random linear extensions using time-series data of social hierarchy covering the time period between the 11th and 12th centuries. This approach introduces a new class of models for rank-order time series data in which actors are listed in order of precedence which is presented as a collection of lists. The lists are modeled as a realisation of a queue in which queue-position is constrained by an underlying social hierarchy corresponding to a partial order, which is used in modelling the evolution of the social hierarchy over the given time period..

3.8 Learning to rank

Learning to rank (LTR)[70, 308] describes a class of algorithmic techniques that apply supervised machine learning to solve ranking problems from a wide range of domains such as information retrieval, recommendation systems, search engine optimization etc. It arises from the need to obtain prediction results based on the best order of their relevance to a machine learning classification problem. Several methods have been proposed for LTR and they can be categorized or grouped into three main approaches: pointwise, pairwise, listwise. Pointwise approach involves scoring items independently and then ranking them based on their scores. In contrast the pairwise or listwise ranking methods, consider the relative positions of items in pairs or lists respectively. The goal for the ranker is to minimize the number of inversions in ranking i.e. cases in which the pair of results are in the wrong order relative to the ground truth. In particular, the listwise learning approach addresses the ranking problem, by analyzing ranked lists of objects as input instances and then trains a ranking function through the minimization of a listwise loss function defined on the predicted list and the ground truth list. In [96], it is noted that not all rankings can follow a strict total order in the context of learning to rank. As such a proposed method of relaxing the conventional setting such that predictions are given in terms of partial instead of total orders is outlined. The key idea is that if a model is uncertain about the relative order of two alternatives, in which case it is unable to clearly determine whether the former should precede the latter and vice-versa, it may decline or pospone the making of a conclusive decision and instead declare such pair of alternatives as incomparable. In general, learning to rank problems as presented in [97] can be formulated as follows: Given a set of training instances $\{x_1,...,x_n\}\subseteq\mathcal{X}$ and a set of labels $\mathcal{Y} = \{y_1, ..., y_k\}$ endowed with an order $y_1 < y_2 < ..., < y_k$ such that for each training instance x_l it can be associated a label y_l . The problem is to determine a ranking function that orders a new set of instances $\{x_i'\}_{i=1}^{t}$ according to their (unknown) preference degrees. The performance measures for this are: AUC(k=2, for the bipartite ranking) and C-index in the polytomous case (k > 2), for k-partite ranking). Since the ranker has the ability to reject predictions, there is a trade-off between correctness and completeness. Correctness is measured

by gamma rank correlation [249] is a measure of rank correlation, i.e., the similarity of the orderings of the data when ranked by each of the quantities. Completeness measure penalizes the abstention from comparisons that should actually be made. Furthermore, a preference relation $P: A \times A \to [0,1]$ provides a measure of support for the pairwise preference a > b with P(a,b) = 1 - P(b,a) for all $a,b \in A$, which can be considered an application of a generic approach [97] to transform every ranker into a partial ranker via ensembling such that:

- 1. for any ranker L, train k ranking models $M_1
 ldots M_k$ by resampling from the original data set, i.e., by k bootstrap samples. By querying these models, k rankings $>_1 \dots >_k$ are generated;
- 2. for each pair of alternatives a and b, the degree of preference is defined as: $P(a,b) = \frac{1}{k} |\{i \mid a \succ_i b\}|$.

3.9 Machine learning using formal concept analysis

Formal concept analysis(FCA) [11, 141] is a well known method which originates from partial order and lattice theory for the purpose of conceptual knowledge representation and data analysis in information science. Formal concept analysis was introduced in the seminal work of Wille(1982) as a result of an attempt to restructure mathematical order and lattice theory. The basic data format in FCA[248] is a cross-table given by the triple (O,A,I) called formal context, where O is a set of elements (referred to as formal objects), A a set of properties (called formal attributes) and I the relation has on $O \times A$. A pair (X,Y) where X is a maximal set of elements (called extent) and Y is a maximal set of shared properties (called intent) represents a formal concept. The set of all concepts of the context is partially ordered by extent inclusion which represents a complete lattice. The fundamental theory underlying FCA is the representability of complete lattices by ordered sets of their meetand join-irreducibles [27]. Since ordered sets of irreducibles are naturally represented by binary matrices, this makes it possible to apply some properties of lattice theory to the analysis of data given by object-attribute matrices[201]. Furthermore, the rows of the binary matrix are usually interpreted as an object and a column is interpreted as a binary attribute. If an object possesses an attribute the corresponding entry is 1 and if it does not possess it then the corresponding entry is 0 [200]. It has also been shown that through a process of discretising and booleanising data, it is possible to convert a variety of datasets into formal contexts or concept lattices [9, 8, 10]. FCA can also be viewed as an unsupervised machine learning technique. In this direction, it takes an input binary relation which is represented by a binary matrix. Its goal is usually to determine the natural concepts described in the data, and then organizes the concepts in the form of a Hasse diagram corresponding to a partial ordering [91, 143, 164]. Due to its generality, simplicity, and powerful mathematical foundation, there is currently a growing adoption of FCA in a wide variety of data science tasks [202]. Its advantage can be seen in the several FCA based machine learning methods that gave competitive results in various machine learning problems compared to classical methods [5]. In a recent study on the application of formal concept analysis in image classification problems, Khatri [196] establishes several advantages of formal concept analysis over convolutional neural networks for image classification task. These advantages are :

- FCA provides clarity which can be useful for interpretability of machine learning classification tasks. Furthermore, FCA facilitates examination of classification hierarchy by providing a visual representation of data in the form of a lattice.
- During the classification process, data can be added and removed from the lattice without the need for retraining. Moreso, the FCA classifier outperforms the random forest classifier, which is a human interpretable model.

An important feature of FCA-based classification methods is that they do not make any assumptions regarding statistical models of a dataset [237]. A typical refinement of the idea of learning, is learning from positive and negative examples, when a learning system obtains descriptions of positive and negative examples and constructs a generalization of positive examples that does not "cover" negative examples [201]. There are several FCA-based models for classification based on learning from positive and negative examples [202], [229], [201]. In Jabin [170] a genetic algorithm based machine learning method is presented for automatically learning object-oriented hierarchies which are very similar to lattice based concept hierarchies from large datasets. Ikeda and Yamamoto [165] propose a method for classifying data that generates a concept lattice and selects appropriate formal concepts in the lattice. The method is useful for robust feature selection during classification by securing sufficient storage space in order to maintain both the selected concepts and redundant concepts. This is beneficial in terms of reduced time duration in solving practical problems related to a variety of large datasets. Formal

concept analysis has been applied in the context of ensemble learning: Dagging [218, 5], Boosting[217], Bagging [184], Recommendatory-based Multiple Classifier System [187]. In Xie[301], concept lattice(CL) model, the core structure in formal concept analysis integrates the simple base classifer (Naive Bayes(NB) and Nearest Neighbour(NN)) into each node of the concept lattice to form a new composite classifer namely-Concept Lattice Naive Bayes(CLNB) and Concept Lattice Nearest Neighbour(CLNN) classification systems. Experimental results on 26 datasets indicate that two hybrid classification systems (CLNN and CLNB) perform better than their corresponding base classifiers and CLNB even outperforms state-of-the-art classifiers. Many machine learning based AI systems have been designed as black boxes. which makes it challenging in achieving interpretability in AI models due to the complexity of AI algorithms. Sangroya et al. [255] propose a general concept lattice theory based framework for explainability of AI models. The main idea is that given an outcome of a deep learning model and a domain ontology, the objective of their proposed solution is to identify an explanation that can point the user to the prominent feature set for a certain outcome. FCA based approaches have also been used for the automated discovery of domain-specific ontologies from textual descriptions of domain entities [25, 283, 310]. Ontology learning [225] refers to the process of automatically extracting and constructing knowledge structures or models from unstructured or semi-structured data sources such as text, speech, images, or sensor measurements etc. These knowledge structures typically take the form of annotated taxonomies, concept hierarchies, or domainspecific ontologies that capture various aspects of the underlying domain or subject matter. Ontologies are a key element of the semantic web. They aim to capture basic knowledge by providing appropriate terms and formal relationships between them, so that they can be used in a machine-processable manner. Relational Concept Analysis (RCA)[163] is an extension of the FCA framework to take into account multi-relational datasets, RCA generates a family of concept lattices, precisely one for each category of objects Using formal concept analysis as well as relational concept analysis different approaches [159, 179, 153, 230, 100] have been implemented for ontology learning and ontology extraction.

4 Clustering Partially Ordered Data

Cluster analysis is a multivariate data analysis technique for identifying structural patterns and relationships within complex datasets with the goal of grouping similar data points into clusters. In the process, it attempts to discover the internal structure of a data set by considering the similarity between objects based on some type of attributes that the objects may possess in varying degrees. The ordinal model for clustering with posets was introduced in [174], where it is shown that the characterization of flat cluster methods [173] leads to a universal mapping problem in the theory of partially ordered sets. A persistent underlying theme involves generalized notions of adjoints of order preserving mappings between posets [175]. In addition, Janowitz [175] attempts to develop a clustering scheme based on dissimilarities measured in posets. Many of the cluster methods that are used in the construction of classificatory systems operate on data in the form of a dissimilarity coefficient on a set of objects [35]. There are two methods in cluster analysis, namely hierarchical and non-hierarchical. Mathematically, hierarchical clustering involves the creation of a nested sequence of partitions of a set, whereas non hierarchical clustering deals with a single partition. The hierarchical clustering method consists of the agglomerative and divisive methods. Divisive clustering uses a top-down approach to combine all the data points as a single cluster and then divides them as the distance between them increases. On the other hand, agglomerative clustering is a bottom-up approach which divides the data points into different clusters and then aggregates them as the distance decreases. Agglomerative clustering can be categorized into single linkage and complete linkage. Single linkage is a method that focuses on minimum distances or nearest neighbor between clusters while complete linkage focuses on maximum distance or furthest neighbor between clusters [240]. Sabara et al. [251] use the single linkage and complete linkage agglomerative hierarchical clustering method on the Hasse graph of posets in a multidimensional data analysis setting. Similarly, Wu et al. [300] and Kardaetz et al. [185] also develop methods that combine poset methodology with hierarchical clustering from a multidimensional data perspective. In particular, Wu et al. [300] develop a method of hierarchical stratification which is interpretable using the partial order Hasse graph, and the evaluation process is such that it takes into consideration the characteristics of multi-dimensional indicators, division into several layers whereby different layers have different properties and characteristics. Janowitz [176] presents a comprehensive review of clustering including through the use of lattices that generalize trees. Ontologies represent data relationships as hierarchies of possibly overlapping classes. Ontologies are closely related

to clustering hierarchies. Dissimilarity is a common intermediary used by clustering methods to classify data. In

Liu et al.[209], it is shown that modeling ontologies as posets over the subset relation enables classical clustering algorithms which takes dissimilarity matrices as inputs and as such it incorporates all available information and therefore there is no loss of information. This served as the basis for the development of a clustering algorithm that generates a partially ordered set of clusters from a dissimilarity matrix. Dissimilarity matrices simplify some of the problems associated with clustering high-dimensional datasets, since their size is only a function of the number of objects $\mathcal{O}(N^2)$, and independent of the object's dimensions.

Formal concept analysis(FCA) is a powerful tool for investigating and identifying hidden structures in large data sets. Janowitz[176] introduces a mathematical treatment of the subject ordinal and relational clustering using FCA. The FCA can be seen as a conceptual clustering method. Michalski[220] introduced conceptual clustering as a new branch of machine learning. Since its introduction there has been significant interest in the potential machine learning tasks that can utilize conceptual clustering methods. Work in artificial intelligence has concentrated significantly on developing methods for classification, and the conceptual representations has been pivotal in supporting this effort. Conceptual clustering[90] is concerned with the problem of grouping unlabelled objects into classes. Furthermore, Carpineto[90] presents the three key features of conceptual clustering methods as follows:

- 1. Every output class, in addition to being characterized by its extensional description (i.e., the set of objects covered by the class), is also characterized by an intensional (conceptual);
- 2. Output classes (also referred to as concepts) are arranged into a hierarchy based on their generality or specificity;
- 3. The process of class formation is incremental in the sense that the processing of the nth object does not require extensive reprocessing of the previously processed (n-1)th objects.

The theory of concept or Galois lattices provides a simple and formal approach to conceptual clustering. The Galois (or concept) lattice which is generated from a binary relation has been proven to be useful for many applications. The construction of Galois lattice [146] can be considered a conceptual clustering method because it results in a concept hierarchy. Carpineto [89] developed the clustering algorithm GALOIS which is able to determine the concept lattice corresponding to a given set of objects. It was demonstrated to be useful for class discovery and class prediction, with the time complexity of each update ranging from O(n) to $O(n^2)$ where n is the number of concepts in the lattice. Fisher[133] presented conceptual clustering as an extension of numerical taxonomy. In Restrepo et al. [242] the theoretical background of the application of Hierarchical Cluster Analysis(HCA) to improve on the interpretability of Hasse Diagrams derived from the ranking process is presented. The use of HCA is based on the idea of reducing the number of elements in a poset by the selection of representatives from the original set used for ranking. As such the clusters derived from HCA can be interpreted as similarity classes in which one of its members (the nearest to the centre of the cluster) can be selected as representative of its class. The process of determining the concepts in formal concept context is time-consuming, as it is considered an NP problem. Zhang et al. [309] present a hierarchical conceptual clustering analysis based on distance function for the purpose of determining the concepts in formal context. In this way, the discovery of all concepts is facilitated by subsetting of the feature set used in the clusters. On the other hand Markov[214] uses hierarchical clustering based on a generalization operator to induce a lattice structure to the clusters. As such it maximizes the overall clustering quality since the local evaluation goes through all levels of the hierarchy in a bottom-up fashion. Yoneda et al. [305] present a method that can learn a graph structured representation from multivariate data where each node represents a cluster of data points and each edge represents the subset-superset relationship between clusters. The main idea of the approach uses formal concept analysis to extract hierarchical relationships between clusters based on the algebraic closedness property which leads to a directed graph representation. Similarly, other graph-based hierarchical conceptual clustering methods which can be applicable to partially ordered datasets have been studied in [180, 181]. In clustering, the frame of discernment is the set of all partitions of a finite set E, denoted $\mathcal{P}(E)$ This set can be partially ordered using the following relation. A partition p is said to be finer than a partition p'(or, equivalently p' is coarser than p) if the clusters of p can be obtained by splitting those of p'; therefore $p \leq p'$. In this case the poset $(\mathcal{P}(E), \leq)$ can be obtained. Ensemble clustering aims at combining the outputs of several clustering algorithms (clusterers) to form a single clustering structure. Using evidential reasoning based on the Dempster-Shafer theory [109] by assuming that there exists a "true" partition p^* and that each clusterer provides evidence about p^* . The evidence from multiple clusterers can then be combined to draw plausible conclusions about p^* in situations which are not immediately evident.

5 Multidimensional Data Analysis

Multidimensional data analysis was primarily established as a consequence of the development of relational databases and OLAP(On-Line Analysis Processing) by E.. Codd in 1993. The main idea of OLAP is based on the notion of dimensions which provides a flexible approach to interpreting data from several angles. That is, multidimensional data analysis [236, 161, 139, 206] allows the observation of data from multiple viewpoints. In statistics, econometrics and related fields, multidimensional analysis (MDA) is a data analysis process that groups data into two categories, namely data dimensions and measurements. Dimensions are a set of axis on which analysis is carried out. Measurements indicate values which are associated with different dimensional entities. As such the analyses and categorization of data based on multiple dimensions and measures constitutes multidimensional data analysis. In MDA, data analysis is based on hierarchy, sequential property and dependency relationships in the dimensions [161]. This form of data can most often be organized into meaningful hierarchies [42, 236]. Multi-indicator models [269] are very useful in the domain of multidimensional data analysis. Multiple indicators consist of two or more "alternative" measures of the same concept. That is, a causal model contains alternative measures of the same thing. For example, two different ways of asking how satisfied an individual is with a service. Most of its familiar applications can be found in fields that involve the evaluation of subjective data. Many public opinion surveys employ this kind of data, usually measured among many thousands of individuals [71, 168]. Many environmental systems and infrastructure systems [227] are monitored using a set of indicator values that assess several features of site condition. Peculiarly, the values of the different indicators often convey different comparative messages for the study sites. It is also useful in the mathematical modelling of business processes, where large and complex data is involved across multiple geographic regions, channels and products. Due to the complexity of multi-indicator systems[57, 65], quantifying these indicators can be difficult and time expensive. Specifically, when dealing with nominal or ordinal multivariate data, a preliminary step is often scaling them with quantitative data. However, this can lead to inconsistent results[118]. The most popular approach for quantifying multi-indicator data values is based on weighted-average or other aggregation techniques (additive [169]. hierarchical[232] etc.) which is then used in a ranking procedure[59]. Ranking data represents a peculiar form of multivariate ordinal data taking values in the set of permutations [224]. The mathematical aggregation of a set of individual indicators that measures multi-dimensional concepts leads to a composite indicator [226]. In the field of social sciences, composite indicators based on scoring or index functions are a commonly used method for ranking or measuring objects based on different properties [22]. A composite indicator may include several dimensions, where the dimensions represent different domains or aspects of the phenomenon being measured. For example, a composite indicator of well-being may cover dimensions such as income, employment, health and education. The problem of ranking objects based on some ordinal scale is a fundamental issue in economics and social statistics. In addition, the validity and robustness of the results obtained from composite indices or other multidimensional data sources through aggregation has been challenged particularly due to the unavoidable aspect of subjectivity in their calculations. In addition, aggregative methods tend to oversimplify the complexity of a phenomenon and they also produce the same results for different situations [3, 2]. Moreover, such data often lacks clear ordering criteria and many ambiguities arise when trying to compare individuals consistently [118]. In order to address these issues, many studies ([22, 50, 233, 120, 57, 71, 129, 12, 18, 69, 124, 82] etc.) advocate the use of partially ordered set theory, which requires subjectivity only in the choice of the properties to be considered. The key motivation for advocating a poset oriented approach derives from the observation that classical aggregative methods of analysis may not clearly express the true complex nature of certain phenomena. Partial orders help to reveal why an object of interest holds a certain ranking position and how much it is subject to change if a composite indicator is upgraded [57]. This methodology is convenient in providing comparable indices or ranks of objects while reducing subjectivity to a minimum. The main advantages of managing multivariate data for ordering purposes based on the concept of posets and its properties are summarized as follows.

- Partial order theory provides a technique to derive rankings in which ties can be considered, which avoids the need of a weighting of the indicators[83]. The horizontal arrangement of objects within a Hasse diagram and depicted by level structure gives a first approximation to a weak order (i.e. tied ranks are not excluded).
- With partial orders there is no restriction to mapping indicators on a single scale.
- Evaluation of phenomenon such as subjective well-being can be addressed in a consistent and effective
 way, using tools from partial order theory and overcoming the limitations of composite and counting
 paradigms[123].

- Partial order theory provides tools for dealing with multidimensional systems of ordinal data which avoids
 the procedure of performing variable aggregation into composite indicators. As such, there is no necessity
 for converting ordinal scores into numerical values which can lead to inconsistency in the evaluation of the
 real phenomena under consideration [120, 12].
- There is no requirement of "weighting" evaluation dimensions, to account for their different relevance when partial order theory is applied [120].
- The evaluative process for ranking using partial order theory is advantageous in that the comparison of the objects of interest is done simultaneously for all indicators without the need of any prior aggregation [84].
- Partial order ranking offers a non-parametric method that neither includes any assumptions about linearity
 nor any assumptions about distribution properties [74]. That is, it is parameter free and as such there are
 no assumptions needed about the statistical distributions of the attributes.
- The theory of partially ordered sets and its graphical representation in the form of Hasse graphs does not need any additional information to sort objects. Beyond sorting, many conclusions can be drawn from Hasse diagrams because they represent a well-defined mathematical structure [50, 45].
- Posets are used to define the structure of comparabilities underlying multi-indicator systems [2]. The poset structure permits the analyses of the phenomenon under study using mathematical tools. A key goal of this analysis is the evaluation of scores to reduce the dimensionality of the phenomenon. In this case, quantitative variables are used for the analysis since the resulting structure of comparabilities defines the poset. Thus, there is no loss of information since it does not involve the possible evaluations of any difference or ratio of their values to obtain distances or proportions.
- Partial order recognizes that the values of different indicators often convey a different comparative message and most often there is no unique way of solving many ranking problems [227], while taking all of the indicator information into account. For example, item A may be considered better than item B with respect to the first indicator, but item B is considered superior to item A with respect to the second indicator. The different comparative messages of an indicator set complicates the process of prioritization of the items based on the indicator scores. In particular, the properties of partially-ordered sets and rank matrices can provide additional useful information to aid decision making beyond an overall ordering of the items.
- The presentation by a Hasse diagram of posets avoids the arbitrariness in constructing a ranking index [50].
- Hasse diagrams of posets allows a holistic view on all the objects which are to be ranked, without introducing
 artificial ranking indices [46].

We summarize how to obtain poset structures from subjective data set as follows. Let k be the number of subjective indicators for the study of some observed phenomenon and let n denote the number of individuals in the study. Furthermore, let $A = \{q_1, q_2, ..., q_k\}$ be such that each element of the set A represents a different variable corresponding to a subjective indicator. Score the response information of each subjective indicator on an m degree ordinal scale $(y_1, y_2, ..., y_m)$ which are ordered in such a way that $y_1 < y_2 < ... < y_m$. For each statistical unit associate the set of degrees to the variables. Each of the n individual profiles is determined by the m ordinal scores given in a set B on each of the k indicators in the set A. This is followed by pairwise comparison of all the n individuals using their individual profiles corresponding to the ordinal scale in B. This evaluation procedure if not totally ordered leads to a poset structure which can be depicted by a Hasse graph of size n. For instance, using this set up for a four degree scale, the profile (y_2, y_3, y_2, y_3) is better than (y_4, y_1, y_3, y_1) , but (y_3, y_3, y_4, y_2) and (y_4, y_3, y_3, y_2) are incomparable. Different types of methodology have been put forward which use poset theory for the analysis of multivariate datasets that have attributes which require a multidimensional approach. Some of these are based on 1) fuzzy set approach [119, 118], partial order linear extensions [120] 3) average rank [71] 4) poset theory with Adjusted Mazziotta-Pareto Index [1] 5) interpersonal comparability [258] 6) fuzzy-first order dominance [129] and so forth. These approaches are summarized as follows.

• The fuzzy set approach was originally developed in the seminal work of Zadeh (1965)[307] as a class of sets that allows its members to have different degrees of membership. It can address uncertainty associated with vague concepts by allowing partial membership to a set. Fuzzy set approach has been used for the statistical evaluation of ordinal data associated to socio-economic phenomena in combination with posets,

in order to overcome the problems associated with the classical aggregative approach based on composite indicators[119, 118, 16]. The main feature of this approach is that the measurement level of the data is fully respected, avoiding any form of improper scaling.

- In Fattore et al. [120] a ranking method is developed based on the linear extensions from partial orders. Particularly, it aims to solve the problem of ranking a finite collection of objects which can be represented as a cloud of points having a suite of indicator values corresponding to each member of the collection. It solves the problem of arbitrary assignment and aggregation of composite numerical score for each object by viewing the relative positions as defining a partial ordering and any given pair of objects may not necessarily be comparable. The interval having the possible ranks assigned to each object can then be determined from the Hasse diagrams of the partial orders corresponding to the collection of all rankings that are compatible with the linear extensions from partial orders. In cases involving very large datasets which require enumerating a large number of linear extensions which can result to computational intractability, it is suggested in [233] to use the method of Discrete Markov Chain Monte Carlo(MCMC) for such problems. The MCMC method involves the sampling of linear extensions [68], followed by the computation of cumulative frequency ranking based on the sampled linear extensions. In Lerche[140], the method of calculating ranking probabilities based on random linear extensions is highlighted. The approach based on random linear extensions can be very useful for the prediction of ranking probabilities and computation of the average rank for larger partially ordered sets, which are practically impossible to handle using a very large set of generated linear extensions.
- Caperna & Boccuzzo[71] suggest a methodology for dealing with big complex data when some of the attributes require the measurement of complex concepts on ordinal or dichotomous scales. The main idea consist of a procedure for sampling units from a big population using a simple criterion to summarize the resulting values appropriately. This is followed by an application of the central limit theorem which makes it possible to compare the results obtained from different groups using statistical tests on the means. The Height of Groups by Sampling method is then used to compare the average rank among groups that are defined by one or more socio-demographic variables influencing the level of the measurable complex concept.
- In Patil et al. [234], a procedures for extracting insights on ordering properties that are embodied in multivariate datasets and applicable in configuring sets of indicators are reviewed.
- The applied methodology described in Alaimo et al[1] for synthesizing the evolution of multi-indicator systems over time is based on partial order theory and the aggregative method of the Adjusted Mazziotta—Pareto Index, to one of the 15 sustainable development goals. In particular, the role of posets was to define the structure of comparabilities underlying multi-indicators systems. This is then followed by a mathematical evaluation of scores to reduce the dimensionality of the phenomenon. Temporal posets are then obtained by merging the different posets. Furthermore, an embedded five level scale divided into(minimum, maximum, first, second and third quartiles of the indicator values are also used to improve the quality of measurement.
- Sen [258] proposes a framework for interpersonal comparability in the context of aggregating the welfare
 measures of individuals. An aggregation relation between any pair of social states which is a quasi-ordering
 and identical with respect to the Pareto quasi-ordering under noncomparability, and is a complete ordering
 under unit comparability, and under full comparability are all considered.
- Fattore & Maggino [123] outline a comprehensive procedure to address evaluation problems in a multidimensional ordinal setting. The procedure incorporates partial order theory for evaluation study that consist of identification of evaluation dimensions, addition of attribute relevance, threshold selection, computation of evaluation scores at the statistical unit and population level.
- Fattore[125] proposes a procedure for the assessment of multidimensional deprivation with ordinal data, based on partial order theory. The procedure focuses on achievement profiles and their comparabilities and incomparabilities The evaluation problem is treated in terms of multidimensional comparisons among profiles. The usefulness of the procedure is noted in its consideration for both vagueness and intensity of multidimensional deprivation and produces a complete set of synthetic indicators for the description of deprivation on statistical populations. In addition, it also allows for the introduction of information on attribute importance and it facilitates the handling of missing data.

- In Fattore[129] the procedure "fuzzy-first order dominance" (F-FOD), employs concepts and tools from partially ordered set theory and from fuzzy relational calculus with the goal of overcoming the main limitations of previously developed algorithms for FOD analysis. F-FOD is useful in that it produces full pairwise comparison matrices which allows for partial orderings and rankings of the statistical units to be derived from the input data.
- Bruggemann et al.[51] also proposes the average rank method which is based on the structure of the Hasse duagram associated to the order relations encoded in the data matrix.

6 Exploratory Data Analysis

In the field of data analysis, exploratory data analysis (EDA) is the preliminary investigation of data to discover patterns, spot anomalies, test hypothesis, check assumptions, relationships and insights. A major challenge in the exploratory data analysis phase is to identify noisy data which can significantly influence its outcome. Noisy data is often characterized by a high degree of variability which cannot be explained in the dataset. In partial orderings, there is the problem of data uncertainty [265] as a result of data noise. Noise may cause changes in the overall orderings of objects. In order to address the issue of data noise, Bruggemann & Carlsen [64] construct a probability scheme which specifies a noise-model from which explicit expressions outlining the distribution for noisy values are derived. This facilitates a priori estimations which determines noise level effects with respect to the order relation between any pair of objects which are affected. As such, the expected level of noise perturbations can be quantified. Furthermore, Carlsen & Bruggemann[79] restudy the effects of data noise or uncertainties on the partial ordering of a series of objects by adopting a fuzzy approach such that partially ordered sets are obtained as a function of noise. Based on this, it becomes possible to identify the stability range in terms of noise where the original partial order is retained. Trend analysis can be a suitable method of exploratory data analysis that can be applied in the context of posets. As shown in Alaimo et al.[1], the evolution of Hasse diagrams can be analyzed over time for the synthesis of phenomenon from which a preliminary depiction of the temporal trend of the phenomenon can be derived. Missing data has been identified to be a problem in non linear trend analysis [259]. A poset methodology for handling missing values is presented in Alaimo [125] using the case study of deprivation data. The pair-wise comparison of the objects of a poset are shown to be a useful way of dealing with missing data. The idea is that if there is no sufficient information to connect a given object to another then there is still the possibility to have information to make comparisons with other objects in the dataset [22]. In the context of formal concept analysis, it is particularly suitable for exploratory data analysis [32, 164] because of its human-centeredness.

7 Descriptive Data Analysis

Descriptive analysis is one of the most crucial phases of statistical data analysis and data exploration that involves summarizing and describing the primary properties of a dataset. It also involves the process of using current and historical data to identify trends and relationships between variables. In addition, it is particularly useful for communicating the evolution of data and it uses trends as the basis for further analysis to drive decision-making. Poset theory has been used in a variety of contexts for data summarization and visualization. Hasse diagrams serve the purpose of being the partial order graphical representation which conveys a considerable amount of information on the partial order structure of the data. For this, Hasse acyclic directed graphs of posets have been used for data visualization tasks that correspond to: causality[282], socioeconomic analysis [121], linear models and analysis of variance [24], experimental design [183], learning analytics [189], image analytics [104], medical data analysis [111], molecular structure prediction [195, 219], spatial analysis [182, 48], multidimensional analysis [103, 86, 57], biomonitoring[285], multivariate analysis[67], sustainability analysis[88], multi-criteria decision analysis[87], environmental data analysis[284], decision support[298, 227] etc. The principal aim of a visualization tool for multidimensional and partially ordered datasets is to provide a direct representation of the data structure, reducing its complexity but retaining the essential patterns in it [121]. Several formal data analysis methods have been developed that are based on the Hasse diagram. One of such methods is the Hasse Diagram Technique (HDT) introduced by Brüggemann & Voigt[54], which is essentially an application of partial order theory based on a data matrix to analyze the structure of multivariate datasets whenever a number of options can be characterized

by multiple attributes (indicators). The sorting approach using the Hasse Diagram Technique (HDT)[51], can be summarized as follows. Consider a dataset of objects to be sorted with respect to some criteria or property. Let m(i) be an appropriate attribute by which the objects are ordered. Furthermore, let x and y represent any two objects with m(i,x) the value of the i^{th} attribute of the object x(i=1,..,n). Then $x \ge y($ which can mean that x is evaluated better than or equal to y) if and only if $m(i,x) \ge m(i,y)$ for every i=1,...,n. The HDT or use of posets can be useful in ranking problems that involves multi-criteria assessment[45] particularly when the objects from a dataset can be viewed relative to all others. There are several key motivations for the use of HDT in ranking problems which are highlighted as follows:

- HDT makes the ranking process transparent.
- Ranking using the Hasse Diagram Technique has the advantage that it can be performed without any normative constraints.
- HDT enables the extraction of structural information and the dimension of Hasse diagrams with respect to visualization of the ranking results [50].
- HDT enables the visualization of parameters in sensitivity analysis [44].

In Bruggemann et al.[51] the Hasse Diagram Technique is proposed for sorting of chemicals with respect to their potential environmental hazards. In particular, in the Hasse Diagram Technique, the dominance of one chemical over another can be established whenever all attributes can simultaneously support the dominance. A disadvantage of the HDT is the absence of aggregating functions with weights as parameters which would permit the removal of conflictive descriptor values. In order to resolve this shortcoming, Simon et al [262] developed METEOR(METhod of Evaluation by ORder) as an extension of the HDT procedure using partial order theory. In contrast to the HDT method, METEOR[263] combines transparent decision support and convenient tools for data analysis with the ability to include stakeholder's preferences in the decision process. Furthermore, METEOR solves the problem of obtaining a single high-ranked object. Another significant advantage of using METEOR is that it avoids the time consuming and resource inefficient process of trial and error selection of different descriptor priorities, since it enables the possibility to calculate the probability of having a particular linear order by descriptor prioritization [243]. METEOR has been used in the computational evaluation of chemicals and environmental hazards as presented in [291, 55]. Halfon et al. [156] suggest an innovative ranking procedure using a vectorial approach for partial ordering applicable to a variety of problems in environmental toxicology. The motivation to use the poset is that it can depict not only situations of comparison amongst chemicals but also when chemicals are incomparable with respect to each other based on environmental hazard. The ranking is then visualized with the aid of computer programs displaying the Hasse diagrams which can be applicable to both small and large datasets. The set of chains and the antichains of a Hasse diagram represent important structural features that can be identified from their vertical and horizontal components respectively. Bruggemann[60] found that tripartite graphs are useful in the interpretation of complex datasets by clarifying the role of indicators causing incomparabilities which are associated with the horizontal components of Hasse diagrams. In the study of the evaluation of analytic performance using partial ordering as described in Carlsen et al. [78], it is demonstrated that the presence of incomparabilities which are considered an inherent feature in partial order ranking may disclose important characteristics. It is also shown that summary statistics of the absolute z-score, absolute skewnesss and the standard deviation from calculations using data on analytical performnce can be visualised on a Hasse diagram to provide a holistic view of the results. In [241, 69] the graphical technique called POSAC is developed for multivariate data analysis in a two-dimensional space such that observations and variables can be studied simultaneously. It works by mapping the rows (e.g., states) of a matrix in a way that maximizes the preservation of their partial order with similar states located in close proximity on the map. POSAC has been demonstrated to be useful in detecting the evolution of crime analytics in large geographical areas over time.

The well-established use of lattice theory is in formal concept analysis (FCA) which provides a conceptual framework for structuring, analyzing and visualizing data, in order to make them more understandable [138]. In particular, the concept lattice with its Hasse graph allows the visualization and summarization of data in a more concise representation. The main characteristic of FCA which is of interest here is its "symmetric" view on objects and properties, whereas a Hasse diagram only shows the order relations which connect the objects in a graphical structure [60]. Large datasets may be difficult to deal with computationally in FCA. More especially, it is the number of formal concepts derived from a dataset that is the key factor in determining if a concept lattice will be useful for visualisation purposes. Andrews & Orphanides[9] suggests that interpretable results can be derived

from existing data sources in which it would not usually have been possible. This can be explained by the fact that formal contexts which would normally be intractable for visualisation have been processed by firstly, focusing on information of interest and secondly, reducing 'noise' in the context, thus revealing readable lattices that precisely represent conceptual meaning in large datasets. A limitation of a Hasse diagram as a representation of a poset, primarily results from its complexity when there exist a large number of edges and vertices. In this case, they are not easily interpretable as the number of vertices increases and they do not also provide any metric information when this is available in the original data. Cluster analysis, on the other hand, reduces the complexity of the data, but it is not designed to preserve information on comparabilities and incomparabilities. Bruggemann & Carlsen [61] combine Hasse diagrams and cluster analysis in a complexity reduction process, producing a visual output that allows the end users to jointly grasp the partial order and the metric structure of the data. This visualization process involves the following steps as outlined in Fattore et al [121]:

- 1. Reducing dataset complexity through a clustering process based on a SOM(Self-Organizing Map);
- Building a classical Hasse diagram on the population of clusters which are associated to the weight vectors of SOM;
- 3. Visually adding information on statistical units and clusters. In particular, information pertaining to the value of the covariates.

Alternatively, the use of cluster analysis and principal component analysis [50] with HDT may be helpful in obtaining statistically relevant data representation and in avoiding insignificant numerical differences of the attributes. As a consequence this would lead to insignificant comparabilities and incomparabilities and thus to very complex Hasse diagrams.

8 Applications

In this section, we summarize the datasets, software and algorithmic solutions for the analysis of partially ordered data. Since as most of the applications of posets in machine learning problems has already being outlined in Section 3, here we focus mostly on practical data analysis applications of posets in various domains. Partially ordered data and structures are ubiquitous, and therefore they are associated with a wide variety of applications. Although the applications are very broad, a significant number of its applications pertains to data-driven decisions in problems of sustainable development [162, 2, 160, 1]. In this case, the main examples are applicable to the analysis of complex and multidimensional systems of ordinal data and to problems of multi-criteria decision making which are relevant in socio-economic and environmental sciences [127]. The application of partial order methodology can sometimes be used as an interim process[84] before other tools are used. Posets and other methods arising from lattice theory are generally useful in comparative evaluation processes [45]. It is more appropriate to use poset methodology for performing comparative evaluation than using statistical tools since as decisions on binary choices based on comparisons by applying some criteria are naturally based on order theory. Furthermore, ordering based on posets can be visualised and additional results can be derived from the mathematical theory of relations between objects. Order relations also facilitate the identification and evaluation of most relevant objects in a study, as well as deriving the sequence of the importance of a criteria to perform a ranking process for ordinal data. The Hasse diagram technique presents a powerful tool to perform comparative evaluations [48]. A procedure has been outlined in Newlin and Patil [227] for identifying the minimal and maximal elements of a poset using their Hasse graph structures in complex scenarios, which are of fundamental interest whenever priority setting procedures are to be performed. Partially ordered sets often comprise a big amount of information on the degree of dominance among their elements and proper tools can be employed to extract and turn it into rankings for various data analysis purposes.

8.1 Environmental Chemistry and Toxicology

The concept of partially ordered sets and their visualisation by Hasse diagrams turns out to be very useful in many applications of environmental pollution data studies where evaluative considerations and assessments are important [156, 128, 73, 76, 291, 77, 266, 287, 288, 289, 49, 43]. Environmental data have been extensively studied by Hasse Diagram Technique in [47, 48, 50, 52] with the goal to identify significant relationships among sediment samples (objects) and degradation indices (attributes). Further investigations using formal concept analysis[11]

with the aim to highlight interaction among hygienic compounds and a synergism between toxicity tests applied to the sediment samples from surface water sources. Hasse diagram technique [48] supports the visualization of more than two-dimensional problems to identify pollution patterns by characterizing geographical regions using their chemical pollution levels, and then suggesting priority regions for further examination. The pollution pattern which determines the location of the regions within a Hasse diagram is useful in the search for remediation strategies. In an alternative approach outlined by Restrepo & Bruggemann [242], the pollution pattern from regional studies was examined using a methodology that combined hierarchical clustering analysis (HCA) and Hasse Diagram Technique. The HCA was used for classification of the objects in the set in order to find similarity classes resulting to reduced set of object repreentatives, thereby making the Hasse diagrams for analysis depicting the network structure more easily understandable. Posets can also be very useful when ranking environmental chemicals by multicriteria analysis [55, 106, 13]. Posets in this case provides a solid formal framework for the ranking of objects without assigning a common scale or weights to the criteria. In the prediction of toxicity levels of chlorobenzene in environmental chemical polychlorinated hydrocarbons, a scheme is developed by Ivanciuc et al. [167] based on poset theory and embedded within an overall reaction network.

8.2 Socio-Economic Analysis

Partially ordered data are prevalent in many branches of the social and behavioral sciences. This has been driven by the fact that partial order theory employed in the applied sciences overcomes the intrinsic disadvantage hidden in aggregation if a multiple attribute system is available [12]. A typical example is in response categories: "Agree", "Neutral", "Disagree", and "Don't Know", of which the first three can be ordered and the last forms a category of its own. This type of data can be derived from a wide range of applications covering different topics such as multidimensional poverty, economic development, inequality measurement etc. Partially ordered set theory has been shown to be particularly suitable to address multi-criteria decision problems since it shows where multidimensional indicator data values are expressing a conflict which can then be identified and resolved by appropriate corrective action [86]. The use of posets with quantitative data reduces the set of operations and choices to be made in order to synthesize indicators (normalization, aggregation), even if they are the natural representation of multidimensional ordinal data [125, 126]. Using this approach, it is possible not only to investigate the nature of the phenomenon in more detail, but also to help policy-makers in their assessments. Infact, it can be considered an effective policy tool aimed at promoting the identification of disadvantaged and under-developed profiles and contexts [2]. The application of the poset-based method provides an understanding of the complexity of the evolution of a social phenomena in terms of temporal trends and comparisons between regional structures in a countrywide basis using a multidimensional data source for the process [1, 71]. Incomparability captures the existence of intrinsically different forms of it, thereby providing a more realistic picture of the phenomenon under investigation. In this respect, the impracticability to compare the profiles of different cities, due to the existence of dimensions where they perform in conflicting ways, reveals the irreducible complexity of sustainability. In census data on deprivation within regions of a country, poset theory has been applied because they can account for incomparabilities which are at the basis of deprivation complexity[16, 118, 119]. Partial order theory has been extensively applied to the construction of synthetic indicators in various socio-economic contexts [1, 2, 14, 3, 23, 110, 59, 80, 81, 107, 124, 126, 130, 244, 245]. In the study of crime analytics [69], POSAC(Partially Ordered Scalogram Analysis with Coordinates) is a graphical representation, providing dimensional reduction procedures that preserve partial order. It has been used to study the structure of criminal networks in terms of their longevity, biographical data the, network size and so forth. Levy [207] performs data analysis by employing the technique of partial order analysis of crime indicators characterizing cities. The results were visualized with a scalogram based on the method of Partial Order Structuple Analysis which was developed for non-metric data analysis. Furthermore, it is noted that the approach described in Levy[207], can be useful in a broad range of problems on the stratification of cities, individuals, as well as several varieties of social indicators for classification.

8.3 Neurocognition Modelling

Finite partially ordered sets are natural models for cognition as it is reasonable to assume that some cognitive states have higher levels of functionality than others [272, 273, 271, 275, 276]. Additionally, finite partially ordered classification models are useful for many statistical applications including cognitive modelling. In particular, a data analytic framework for implementing latent finite partially ordered classification models introduced by

Tatsuoka[272], provides useful methods for evaluating cognitive applications that are latent and complex. Poset models are flexible and can become quite rich and complex, enabling them to be effective models for describing response phenomena from educational test data or neuropsychological assessment data[272, 274, 279]. An objective of neuropsychological assessment is to determine differences in cognitive functioning in clinical settings. Carr et al. [92] used poset classification models of neuropsychological test data to classify samples into detailed cognitive profiles using ADNI2(Alzheimer's Disease Neuroimaging Initiative) and AIBL(Australian Imaging, Biomarker & Lifestyle) datasets. In the risk of disease progression of Alzheimer disease for individuals with mild cognitive impairment (MCI), Tatsuoka et al. [277] suggest that poset-based modeling methods may be useful in providing more precise classification of cognitive subgroups among MCI for imaging and genetics studies, and for developing more efficient and focused cognitive test batteries. An order structure arises naturally with skill profiles. The flexibility to not necessarily assume that one state is greater than another is an appealing feature of posets[275]. In addition, posets are comprised of states, into which cases are classified, that are associated with distinct patterns of attribute strengths and weaknesses [172]. Posets have several advantages over conventional statistical methods for handling large numbers of polyfactorial neuropsychological test variables such as the ability to mimic the expert judgment of a clinical neuropsychologist for each case in a large sample. Posets are efficient in these tasks since as valid conclusions can be drawn based on relatively few measures, and classification of large samples can be accomplished rapidly [171]. On the other hand, Jaeger et al. [171] noted that the primary limitation in poset modeliing of large conventional neuropsychological test datasets is based on its restriction to a selected set of attributes, while additional important distinctions remain to be tested.

8.4 Other

The application of posets and lattice theory are not limited to the aforementioned domains and tasks for data analysis. Posets have also been used in: 1) hypothesis management in large scale research projects [147]; 2) longitudinal data analysis[113]; 3) social macroeconomics analysis[93]; 4) adaptive testing for cognitive assessment [278]; 5) psychometrics[166]; 6) medical statistics [261]; 7) machine learning of reviewer's paper preferences [96]; 8) visual similarity learning using pose estimations [31]; 9) gender analysis [85]; 10) Pattern Mining [268] etc.

8.5 Some Selected Datasets

1. Computer Vision Datasets

- The Olympic Sports dataset [317], The Leeds Sports Pose(LSP)[316], MPII Pose [315]. They are employed in the study of visual similarity learning from pose estimation using posets.[31].
- Micro-video dataset [99]. It is used for the learning problem on hypergraph partial order [132].

2. Natural Language Datasets:

- The Compositional Freebase Questions (CFQ)(Keysers et al., 2020)[318] is a dataset that is specifically designed to measure compositional generalization. It is used for the study of hierarchical poset decoding for compositional generalization in language [150].
- Universal Dependencies (UD) corpora [319]. It is used in the learning algorithm for generating a surface word order for a sentence given its dependency tree [116].

3. Sustainanble Development Data:

The sustainable datasets are categorized into socio-economic and environmetal datasets.

Socio-economic data

- Equitable and sustainable well-being data [320]. Used in [1, 3] for multidimensional data analysis in the context of sustainable development.
- Service Performance Dataset [321]. Used for multidimensional data analysis in [12].
- Data on crime rates and their ranking for sixteen American cities is described in Levy [207].

4. Environmental data

- A battery of biochemical, microbiological and bioassay tests were used to identify degraded or degrading sediments in waters [47, 50, 11]. The dataset is available in [115].
- Collection of toxicity data [167].
- Multiple indicator data for stream channel stability at bridge crossing is described in [227].

5. Other datasets:

- (i)Data covering a time period up to the end of 2022 on the historical head-to-head matches of six professional tennis players.(ii) Data on educational testing from 15 OECD countries in reading comprehension based on test performance in 2015 from the Programme for International Student Assessment (PISA). Both of these are described in [270] for the purpose of partial ranking of tennis players and total ranking of educational systems respectively.
- A mass spectroscopy dataset consisting of 11 phosphoproteins and phospholipids containing approximately 854 measurements of abundance levels in an observational setting described in the supplementary material in [252]. It is used in Taeb et al. [270] for the purpose of learning causal relations and structures in proteins.
- UCI machine learning repository [313], used in [217].
- Poset/Hypergraph Currvature Datasets[324], used in [303] to perform an empirical study involving computation and analysis of the Forman–Ricci curvature of hyperedges in 12 real-world hypergraphs.

8.6 Selected Algorithms and Software Packages

1. Machine learning and Deep learning

- Python implementation of deep visual similarity learning using posets [322]. Used in [31].
- Given as input an edge-weighted poset, the algorithm in [116] constructs a total order such that nodes with smallest weights are adjacent. The algorithm works by attempting to order a set of words as closely as possible to their original surface realization in the Universal Dependencies(UD) corpus. Due to the fact that words may repeat in the sentence, each order is instead represented by a list of integers, and it is these lists of integers which are compared [116]. For example, assuming a target reference order of [1,2,3] for the red horse, the generated order of red the horse would be [2,1,3].
- Causal Structure Discovery Algorithm-Greedy Sparsest Poset (GSPo), python implementation in [323]. Used for causal structure learning in the presence of latent variables [34].
- A learning algorithm for discovering partial orders from sequences of events is described in [212].
- The learning algorithm of pertinent concept described in [217] uses the Adaboost algorithm with formal concept analysis on a learning dataset to discover lattice concepts used for classification rules.
- An algorithm framework for integrating base classifier into concept node of concept lattice is described in [301].
- Greedy sequential algorithm for model selection described in [270]. It is used in partial ranking of tennis players, total ranking in educational systems, and causal structure learning on proteins dataset.

2. Multidimensional data analyis:

The Hasse Diagram Technique (HDT) described in section 7 provides a huge collection of methods that are simple but tend to be complicated if the number of elements in the poset increases. To address this chalenge, special software packages have been developed to support the HDT. Some of these are described as follows.

• The PARSEC[15, 122] is an R package for poset-based evaluation of socio-economic data. Its main goal is to provide socio-economic scholars with an integrated set of elementary functions for multidimensional poverty evaluation based on ordinal information. The package is organized in four main parts

- i) Data management.. ii) Basic poset analysis. iii) Poset-based evaluation. iv) OPHI counting approach.
- An algorithm based on posets for multidimensional data analysis is described in [17].
- Self-organizing map algorithm(SOM)[314] is a popular tool for non-linear dimensionality reduction and pattern recognition. It is used for multidimensional data analysis described in [16].
- A non-aggregative partial order algorithm for the construction of sustainability synthetic indicators on multi-indicators systems is described by Arcagni et al. [17].
- An application of partial order theory for object rank correlation analysis of multiple variables is implemented by the software package PO Correlation [267]. The design is made transparent for rank correlation analysis by a detailed mapping of the rank relations between all objects.
- The software WHASSE is applied for chemical monitoring data analysis [51, 53, 157], medical data analysis [111].
- PLMIX[224] an R package for modeling and clustering partially ranked data.
- PyHasse software [62, 66, 198] is used for the purpose of ordinal analysis in data matrices, identification
 and analysis of partial order relations as well as in computing ranks. Furthermore, it has been used
 in [293, 55, 290] for the analysis and evaluation of environmental data and in [87] for multi-criteria
 decision analyses.
- ProRank[238] software for partial order ranking. It is used in the evaluation of environmental databases [292].

9 Future Directions

Posets have proven their usefulness in a broad range of applications. But there still remains many more data centric applications where existing and new methods of learning and analyzing partially ordered data can be relevant in generating novel results. In this section we suggest four areas.

Learning Algorithms A major challenge in data science is the identification of geometric structure in highdimensional data. The structural understanding of data is very relevant for designing efficient algorithms for optimization and machine learning. Classically, the structure of data has been studied under Euclidean assumptions. The fundamental representation of the features for any machine learning model is the vector and its multidimensional generalization which is the tensor. As such many machine learning algorithms and data analyis pipelines have been developed with the aim of computing vectors or matrices of real numbers. There are currently a wide range of well-studied tools and algorithms that assume such structure. However, many scientific fields study data with an underlying structure that can only be represented in non-Euclidean space. Such data can be found in graphs and many kinds of manifolds. There is currently growing research on developing machine learning models that work with data representations for features and embeddings, and or model parameters in the non Euclidean domains. Posets as a class of directed acyclic graphs have also begun to recieve attention with regards to developing appropriate learning algorithms. Recently, Wendler [296] introduced new methods for the fourier-sparse learning on data indexed by lattices and posets. The graph neural network class of algorithms also works on posets as specified in section 3. But there is not yet widespread use of graph neural networks relative to the number of i applications based on posets and lattices. The success of machine learning algorithms generally depends on data representation since as different representations encode different explanatory factors of variation behind the data [33, 112]. There is enormous scope to develop new robust and efficient algorithms for learning the key structural characteristics of data represented as Hasse graphs of lattices and posets from a graph based perspective. There is also the need to develop specialized measures of model evaluation metrics for the use in any potential poset learning algorithms.

Posets and Topological Data Analysis Topological data analysis (TDA)[94] is a recent and fast-growing field in applied mathematics which provides a set of new topological and geometric tools to infer relevant features from complex data. In topological data analysis[178], the shape of a dataset is often encoded into a system of vector spaces and linear maps over a partially ordered set. There is a growing body of literature which connects posets to topological data analysis [177, 72, 40, 281]. The current survey did not take into consideration the posets that appear in topological data analysis. More recently, there has also been the introduction of posets in the

construction of topological deep learning models[155]. There are many new applications which can arise from the topological analysis of partially ordered data and the development of their associated deep learning framework. There is the need to study situations of multiway interactions within data points endowed with a poset structure, by leveraging concepts such as Latent Topology Inference (LTI) [30].

Machine Learning and Formal Concept Analysis Several key aspects of machine learning in the context of formal concept analysis was introduced in section 3.9. However, the scope of applications is not exhaustive especially for ontology learning using formal concept analysis techniques. A more indepth review of existing and new models for machine learning using formal concept approaches can be useful in highlighting latest results.

Learning to rank with partially ordered data Learning to rank refers to machine learning techniques for training the model in a ranking task. The prediction of structured outputs in general and rankings in particular has attracted considerable attention in machine learning in recent years, and different types of ranking problems have already been studied. Several ranking problems have been introduced in the context of data analysis of partially ordered data in sections 5 and 7. There is a enormous potential of introducing existing and new techniques of learning to rank, for ranking problems with datasets from which can be derived a partial order structure such as big data for sustainable development etc. In particular, it can be useful to implement solutions from the various graph based[142, 117] approaches of learning to rank for this purpose or developing new learning to rank algorithms for partially ordered data.

10 Conclusion

This article can be considered the first exhaustive review summarizing the applications of posets in relation to data analysis and machine learning. The vast use of poset theory in many domains of machine learning and data analysis demonstrates their relevance for a long time to come. We have examined the growing adoption of poset methodology in several areas of machine learning and deep learning. Poset theory was also systematically categorized into multidimensional, descriptive and explorative data analysis from many perspectives. The practical applications of posets was drawn from a broad range of domains. The main contributions of this work is to highlight the deep connections poset theory has with data analysis and machine learning. It is expected that this article will provide a basic and in-depth understanding of the applications of posets in data science and the discovery of new directions for future research on the subject.

References

- [1] Alaimo, L.S., Arcagni, A., Fattore, M. et al. Synthesis of Multi-indicator System Over Time: A Poset-based Approach. Soc Indic Res 157, 77–99 (2021).
- [2] Alaimo, L.S., Ciacci, A. & Ivaldi, E. Measuring Sustainable Development by Non-aggregative Approach. Soc Indic Res 157, 101–122 (2021).
- [3] L. S. Alaimo & F. Maggino, 2020. "Sustainable Development Goals Indicators at Territorial Level: Conceptual and Methodological Issues—The Italian Perspective," Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, Springer, vol. 147(2), pages 383-419.
- [4] A. Albuquerque, T. Amador, R. Ferreira, A. Veloso and N. Ziviani, "Learning to Rank with Deep Autoencoder Features," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489646.
- [5] M. A. Ali, A. Jaoua and S. A. Al-Maadeed, "A novel Conceptual Machine Learning Method using Random Conceptual Decomposition," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 18-22.
- [6] A An., Learning Classification Rules from Data, (2003) http://www.cs.yorku.ca/~aan/research/paper/cam03.pdf.
- [7] Anđelić, M., da Fonseca, C.M. Cover matrices of posets and their spectra. Czech Math J 59, 1077–1085 (2009). https://doi.org/10.1007/s10587-009-0075-6
- [8] Andrews, S.: Data Conversion and Interoperability for FCA. In: CS-TIW 2009, pp. 42-49, http://www.kde.cs.uni-kassel.de/ws/cs-tiw2009/proceedings_final_15July.pdf

- [9] S. Andrews and C. Orphanides, "Analysis of large data sets using formal concept lattices," In: M. Kryszkiewicz and S. Obiedkov, (eds.) Proceedings of the 7th International Conference on Concept Lattices and Their Applications. Seville, University of Seville, 2010, 104-115.
- [10] Andrews, S., Polovina, S., Visualising computational intelligence through converting data into formal concepts(2011), https://shura.shu.ac.uk/2720/.
- [11] P. Annoni, R. Bruggemann, The dualistic approach of FCA: A further insight into Ontario Lake sediments, Chemosphere 70 (2008) 2025–2031.
- [12] Annoni, P., Brüggemann, R. Exploring Partial Order of European Countries. Soc Indic Res 92, 471–487 (2009).
- [13] P. Annoni, R. Brüggemann, A. Saltelli, Partial order investigation of multiple indicator systems using variance-based sensitivity analysis, *Environmental Modelling & Software*, Vol. 26; 7, 2011, pp. 950-958, https://doi.org/10.1016/j.envsoft.2011.01.008
- [14] Annoni, P., Fattore, M. & Bruggemann R. (2011). A Multi-Criteria Fuzzy Approach for Analyzing Poverty structure. Statistica & Applicazioni, Special Issue, 7-30.
- [15] Arcagni, A., & Fattore, M. (2014). PARSEC: An R package for poset-based evaluation of multidimen-sional poverty. In R. Bruggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order*. Berlin: Springer.
- [16] Arcagni, A., Barbiano di Belgiojoso, E., Fattore, M. et al. Multidimensional Analysis of Deprivation and Fragility Patterns of Migrants in Lombardy, Using Partially Ordered Sets and Self-Organizing Maps. Soc Indic Res 141, 551–579 (2019).
- [17] Arcagni A., Cavalli L., and Fattore, M., Partial Order Algorithms for the Assessment of Italian Cities Sustainability (February 3, 2021). FEEM Working Paper No. 1.2021, Available at https://ssrn.com/abstract=3778559orhttp://dx.doi.org/10.2139/ssrn.3778559
- [18] Arcagni, A., A. Avellone, and M. Fattore (2022). Complexity reduction and approximation of multidomain systems of partially ordered data. Computational Statistics & Data Analysis 173, 107520.
- [19] Avogadri, R., Valentini, G. (2008). Ensemble Clustering with a Fuzzy Approach. In: Okun, O., Valentini, G. (eds) Supervised and Unsupervised Ensemble Methods and their Applications. Studies in Computational Intelligence, vol 126. Springer, Berlin, Heidelberg.https://doi.org/10.1007/978-3-540-78981-9_3
- [20] N. A. Asif et al., "Graph Neural Network: A Comprehensive Review on Non-Euclidean Space," in IEEE Access, vol. 9, pp. 60588-60606, 2021, doi: 10.1109/ACCESS.2021.3071274.
- [21] M. A. Babin and S. O. Kuznetsov, Enumerating Minimal Hypotheses and Dualizing Monotone Boolean Functions on Lattices
- [22] J. Bachtrögler, H. Badinger, A. F. de Clairfontaine, and W. H.Reuter, "Summarizing data using partially ordered set theory: anapplication to fiscal frameworks in 97 countries," *Stat. J. IAOS*, vol.32, no. 3, pp. 383–402
- [23] Badinger H., Reuter W. H. (2015). Measurement of Fiscal Rules: Introducing the Application of Partially Ordered Set (POSET) Theory. Journal of Macroeconomics, 43, 108–23.
- [24] R. A. Bailey (2021) Hasse diagrams as a visual aid for linear models and analysis of variance, Communications in Statistics - Theory and Methods, 50:21, 5034-5067, DOI: 10.1080/03610926.2019.1676443
- [25] Bain, M. (2003). Inductive Construction of Ontologies from Formal Concept Analysis. In: Gedeon, T.(.D., Fung, L.C.C. (eds) AI 2003: Advances in Artificial Intelligence. AI 2003. Lecture Notes in Computer Science(), vol 2903. Springer, Berlin, Heidelberg.https://doi.org/10.1007/978-3-540-24581-0_8.
- [26] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *IEEE Transactions on Signal Processing*, 2020.
- [27] Barbut, M. and Monjardet, B., Ordre et classification, II, Paris: Hachette, 1970.
- [28] Barthélemy, J.P., Flament, C., Monjardet, B. (1982). Ordered Sets and Social Sciences. In: Rival, I. (eds) Ordered Sets. NATO Advanced Study Institutes Series, vol 83. Springer, Dordrecht.

- [29] S. M. Basha, D. S. Rajput, Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap, https://doi.org/10.1016/B978-0-12-816718-2.00016-6.
- [30] Battiloro C. et al, From Latent Graph to Latent Topology Inference: Differentiable Cell Complex Module, arXiv:2305.16174v2.
- [31] M.A. Bautista, A. Sanakoyeu, B. Ommer, Deep Unsupervised Similarity Learning using Partially Ordered Sets, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu HI USA 2017,pp. 1923–1932, doi: 10.1109/CVPR.2017.208.
- [32] Belohlavek, R., Sklenar, V., & Zacpal, J. (2004). Formal concept analysis with hierarchically ordered attributes. International Journal of General Systems, 33(4), 383–394. https://doi.org/10.1080/03081070410001679715
- [33] Y. Bengio, A. Courville, & P. Vincent[†], (2014) Representation Learning: A Review and New Perspectives: arXiv:1206.5538v3.
- [34] Bernstein, D.I.; Saeed, B.; Squires, C.; Uhler, C.; Ordering-Based Causal Structure Learning in the Presence of Latent Variables, Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.
- [35] Bertrand, M.F. Janowitz, Pyramids and weak hierarchies in the ordinal model for clustering, Discrete Applied Mathematics, Vol. 122; 1–3, 2002, pp. 55-81, https://doi.org/10.1016/S0166-218X(01)00354-7.
- [36] Bi, W. &; Kwok, J. (2013). Efficient Multi-label Classification with Many Labels. Proceedings of the 30th International Conference on Machine Learning, in Proceedings of Machine Learning Research 28(3):405-413 Available from https://proceedings.mlr.press/v28/bi13.html.
- [37] G. Birkhoff, Lattice Theory, American Mathematical Society, Vol. 25; 3 1967.
- [38] Blocher, H., Schollmeyer, G., Jansen, C. (2022). Statistical Models for Partial Orders Based on Data Depth and Formal Concept Analysis. In: Ciucci, D., et al. Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2022. Communications in Computer and Information Science, vol 1602. Springer, Cham. https://doi.org/10.1007/978-3-031-08974-9_2.
- [39] H. Blocher, G. Schollmeyer, C. Jasen, M. Nalenz, Depth Functions for Partial Orders with a Descriptive Analysis of Machine Learning Algorithms, Proceedings of Machine learning Research 215: 59–71, 2023.
- [40] MB Botnan, Topological Data Analysis, Lecture Notes, https://www.few.vu.nl/~botnan/lecture_notes.pdf
- [41] G. Brauwers and F. Frasincar, "A General Survey on Attention Mechanisms in Deep Learning" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 04, pp. 3279-3298, 2023. doi: 10.1109/TKDE.2021.3126456.
- [42] Brisaboa, N.R., Cerdeira-Pena, A., López-López, N., Navarro, G., Penabad, M.R., Silva-Coira, F. (2016). Efficient Representation of Multidimensional Data over Hierarchical Domains. In: Inenaga, S., Sadakane, K., Sakai, T. (eds) String Processing and Information Retrieval. SPIRE 2016. Lecture Notes in Computer Science, vol 9954. Springer, Cham. https://doi.org/10.1007/978-3-319-46049-9_19.
- [43] Brüggemann, R., Münzer, B., & Halfon, E. (1994). An algebraic/graphical tool to compare ecosystems with respect to their pollution. The German River Elbe as an example. I: Hasse-diagrams. *Chemosphere*, 28, 863-872.
- [44] R. Bruggemann, J. Schwaiger, R. D. Negele, Applying Hasse diagram technique for the evaluation of toxicological fish tests, *Chemosphere* **30** (1995) 1767–1780.
- [45] Bruggemann, R. and K. Voigt (1995). "An Evaluation of Online Databases by Methods of Lattice Theory." Chemosphere 31: 3585-3594.
- [46] Bruggemann, R.; Oberemm, A.; Steinberg, C. Ranking of Aquatic Effect Tests Using Hasse Diagrams. *Toxicol. En Viron. Chem.* 1997, 63, 125-139.
- [47] Bruggemann, R.; Halfon, E. Comparative Analysis of Nearshore Contaminated Sites in Lake Ontario: Ranking for Environmental Hazard. J. EnViron. Sci. Health 1997, A32(1), 277-292.
- [48] R. Bruggemann, S. Pudenz, K. Voigt, A. Kaune, K. Kreimes, An algebraic/graphical tool to compare ecosystems with respect to their pollution. IV: Comparative regional analysis by Boolean arithmetics, *Chemosphere* 38(1999) 2263–2279.

- [49] Brüggemann, R. and H.-G. Bartel (1999) A Theoretical Concept to Rank Environmentally Significant Chemicals. J.Chem.Inf.Comp.Sc. 39, 211-217.
- [50] R. Bruggemann, E. Halfon, G. Welzl, K. Voigt, C. Steinberg, Applying the concept of partially ordered sets on the ranking of near-shore sediments by a battery of tests, J. Chem. Inf. Comp. Sci. 41 (2001) 918–925.
- [51] R. Bruggemann, U. Simon, S. Mey, Estimation of averaged ranks by extended local partial order models, MATCH Commun. Math. Comput. Chem. 54 (2005) 489–518.
- [52] Brüggemann, R.; Carlsen, L. Partial Order in Environmental Sciences and Chemistry; Springer, 2006.
- [53] R. Bruggemann, G. Restrepo, K. Voigt, Structure–fate relationships of organic chemicals derived from the software packages E4CHEM and WHASSE, J. Chem. Inf. Model. 46 (2006) 894–902.
- [54] R. Bruggemann, K. Voigt, Basic principles of Hasse diagram technique in chemistry, Comb. Chem. & High Throughput Screen. 11 (2008) 756–769.
- [55] R. Bruggemann, K. Voigt, G. Restrepo, U. Simon, The concept of stability fields and hot spots in ranking of environmental chemicals, J. Environ. Model. & Soft. 23 (2008) 1000–1012.
- [56] R. Bruggemann, K. Voigt, Analysis of partial orders in environmental systems apply-ing the new software Py-Hasse, in: J. Wittmann, M. Flechsig (Eds), Simulation in Umwelt- und Geowissenschaften- Workshop Potsdam 2009, Shaker-Verlag, Aachen, 2009, pp. 43–55.
- [57] Bruggemann, R., Patil, G.P. Multicriteria prioritization and partial order in environmental sciences. Environ Ecol Stat 17, 383-410 (2010).https://doi.org/10.1007/s10651-010-0167-3
- [58] R. Bruggemann, L. Carlsen, An improved estimation of averaged ranks of partial orders, MATCH Commun. Math. Comput. Chem. 65 (2011) 383–414.
- [59] Bruggemann R, Patil G.P. Ranking and prioritization for multi-indicator systems—Introduction to partial order applications. Springer; 2011.
- [60] Brüggemann, R. and K. Voigt (2011). "A New Tool to Analyze Partially Ordered Sets. Application: Ranking of Polychlorinated Biphenyls and Alkanes/Alkenes in River Main, Germany." MATCH: Communications in Mathematical and in Computer Chemistry 66: 231–251.
- [61] R. Bruggemann, L. Carlsen, Incomparable: What now II? Absorption of incomparabilities by a cluster method, Quality & Quantity 49(4), (2014).
- [62] Brüggemann, R.; Carlsen, L.; Voigt, K.; Wieland, R. PyHasse Software for Partial Order Analysis: Scientific Background and Description of Selected Modules. In *Multi-Indicator Systems and Modelling in Partial Order*; Springer: New York, NY, USA, 2014; pp. 389–423.
- [63] Bruggemann, R.; Carlsen, L. Incomparable—What now? MATCH Commun. Math. Comput. Chem. 2014, 71, 694–716.
- [64] Bruggemann, R. and Carlsen, L. An attempt to Understand Noisy Posets. MATCH Commun.Math.Comput.Chem., 2016, 75, 485-510.
- [65] Bruggemann, R. et al.(Eds.)(2021). Measuring and Understanding Complex Phenomena: Indicators and Their Analysis in Different Scientific Fields. London, UK, Springer Nature.
- [66] Bruggemann, R., Kerber, A., Koppatz, P., Pratz, V. (2021). PyHasse, a Software Package for Applicational Studies of Partial Orderings. In: Bruggemann, R., Carlsen, L., Beycan, T., Suter, C., Maggino, F. (eds) Measuring and Understanding Complex Phenomena. Springer, Cham.https://doi.org/10.1007/978-3-030-59683-5_18.
- [67] Brunsdon, C. Rank Inadequacy: A Partially Ordered Set Approach For Multivariate Data Analysis. https://huckg.is/gisruk2017/GISRUK_2017_paper_31.pdf.
- [68] Bubley R, Dyer M (1999) Faster random generation of linear extensions. Discrete Math 201(1-3):81-88.
- [69] Canter, D. A Partial Order Scalogram Analysis of Criminal Network Structures. Behaviormetrika 31, 131–152 (2004). https://doi.org/10.2333/bhmk.31.131
- [70] Cao et al., Learning to Rank: From Pairwise Approach to Listwise Approach, ICML '07:

 Proceedings of the 24th international conference on Machine learningJune 2007Pages 129–136

 https://doi.org/10.1145/1273496.1273513.

- [71] Caperna, G., & Boccuzzo, G. (2018). Use of poset theory with big datasets: A new proposal applied to the analysis of life satisfaction in Italy. *Social Indicators Research*, 136, 1071–1088.
- [72] Caputi, L., Collari, C. & Di Trani, S. Combinatorial and topological aspects of path posets, and multipath cohomology. J Algebr Comb 57, 617–658 (2023). https://doi.org/10.1007/s10801-022-01180-9.
- [73] L. Carlsen, Partial order ranking of organophosphates with special emphasis on nerve agents, MATCH Commun. Math. Comput. Chem. 54 (2005) 519–534.
- [74] Carlsen L. Assessment of chemicals applying partial order ranking techniques. Comb Chem High Throughput Screen. 2008 Dec;11(10):794-805. doi: 10.2174/138620708786734280. PMID: 19075601
- [75] Carlsen, R. Bruggemann, Partial order ranking as a tool in environmental impact assessment. PAH and PCB pollution of the river Main as an illustrative example, in:G. T. Halley, Y. T. Fridian (Eds.), Environmental Impact Assessment, Nova Science Publishers, 2009, pp. 335–354.
- [76] Carlsen, B. N. Kenessov, S. B. Batyrbekova, A QSAR/QSTR study on the human health impact of the rocket fuel 1,1-dimethylhydrazine and its transformation products. Multicriteria hazard ranking based on partial order methodologies, *Environ. Tox. Pharm.* 27 (2009) 415–423.
- [77] L. Carlsen, The interplay between QSAR/QSPR studies and partial order ranking and formal concept analyses, Int. J. Mol. Sci. 10 (2009) 1628–1657.
- [78] Carlsen L.; Bruggemann R.; Kenessova O.; Erzhigitov E. Evaluation of analytical performance based on partial order methodology. Talanta 2015, 132, 285-293.
- [79] Carlsen, L.; Bruggemann, R. On the influence of data noise and uncertainty on ordering of objects, described by a multi-indicator ystem. A set of pesticides as an exemplary case. J. Chemom. 2016, 30, 22–29.
- [80] Carlsen L., Brueggemann R. (2017). Fragile State Index: Trends and Developments. A Partial Order Data Analysis. Social Indicators Research 133, 1-14).
- [81] Carlsen L. (2017). An Alternative View on Distribution Keys for the Possible Relocation of Refugees in the European Union. Social Indicators Research., 130, 1147-1163.
- [82] Carlsen, L. (2018). Happiness as a sustainability factor. The world happiness index: A posetic-based dataanalysis. Sustainability Science, 13(2), 549–571
- [83] Carlsen, L., & Bruggemann, R. (2019). An analysis of the 'Failed States Index' by partial order methodology. Journal of Social Structure, 14(1), 1–31.
- [84] Carlsen, L.; Bruggermann, R. Inequalities in the European Union—A Partial Order Analysis of the Main Indicators, Sustainability (2021), 13, 62–78.
- [85] Carlsen, L., & Bruggemann, R. (2021). Gender equality in Europe: The development of the sustainabledevelopment goal No. 5 illustrated by exemplary cases. Social Indicators Research, 158(3), 1127-1151.
- [86] Carlsen, L.; Bruggemann, R. Partial Order as Decision Support Between Statistics and Multi-criteria Decision Analyses. Standards 2022, 2, 22.
- [87] Carlsen, L.; Bruggemann, R. Combining different stakeholders' opinions in multi-criteria decision analyses applying partial ordermethodology. Standards 2022, 2, 35.
- [88] Carlsen, L. Food Waste: The Good, the Bad, and (Maybe) the Ugly, Standards 3(1):43-56, 2023.
- [89] Carpineto C. and Giovanni R., "GALOIS: An Order-Theoretic Approach to Conceptual Clustering." International Conference on Machine Learning (1993).
- [90] Carpineto, C., & Romano, G. (1996). A Lattice Conceptual Clustering System and Its Application to Browsing Retrieval. *Machine Learning*, 24, 95-122.
- [91] C. Carpineto, G. Romano, Concept Data Analysis, Wiley, Chichester, 2004.
- [92] Carr et al., Associating Cognition With Amyloid Status Using Partially Ordered Set Analysis. Front Neurol. 2019;10:976.
- [93] Cavalletti, B., Corsi, M. "Beyond GDP" Effects on National Subjective Well-Being of OECD Countries. Soc Indic Res 136, 931–966 (2018). https://doi.org/10.1007/s11205-016-1477-0

- [94] F. Chanza & B. Michel, I An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists Front. Artif. Intell., Vol. 4 (2021)https://doi.org/10.3389/frai.2021.667963
- [95] Y. Chen, M. Mancini, X. Zhu and Z. Akata, "Semi-Supervised and Unsupervised Deep Visual Learning: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 3, pp. 1327-1347, March 2024, doi: 10.1109/TPAMI.2022.3201576.
- [96] Cheng, W., Rademaker, M., De Baets, B., Hüllermeier, E. (2010). Predicting Partial Orders: Ranking with Abstention. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2010. Lecture Notes in Computer Science(), vol 6321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15880-3_20.
- [97] Cheng et al., Partial Orders: Ranking with Abstention, https://www.weiweicheng.com/research/slidesposters/cheng-ecml
- [98] G.-S Cheon, B. Curtis, G. Kwon and A.M. Mwafise, Riordan posets and associated incidence matrices, *Linear Algebra Appl.*, 632: 308–331 (2022).
- [99] J. Chen, X. Song, Liqiang Nie, X. Wang, H. Zhang, and Tat-Seng Chua. 2016. Micro tells macro: predicting the popularity of micro-videos via a transductive model. In MM. 898–907.
- [100] Cimiano, P., Hotho, A., & Staab, S. (2005). Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. *Journal of Artificial Intelligence Research* 2005 DOI:10.1613/jair.1648
- [101] M Collery(2022), Learning binary classification rules for sequential data, https://strl2022.github.io/files/short1.pdf.
- [102] M. Collery et al. (2023), Neural-based classification rule learning for sequential data, arXiv:2302.11286.
- [103] Comim, F. A Poset-Generalizability Method for Human Development Indicators. Soc. Indic. Res. 2021, 158, 1179–1198.
- [104] M. Crampes, J. Oliveira-Kumar, S. Ranwez, J. Villerd. Visualizing Social Photos on a Hasse Diagram for Eliciting Relations and Indexing New Photos. IEEE Computer Graphics and Applications, 2009, 15 (6), pp.985-992.
- [105] V. Dankers, E. Bruni, and D. Hupkes. 2022. The Paradox of the Compositionality of Natural Language: A Neural Machine Translation Case Study. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4154–4175, Dublin, Ireland. Association for Computational Linguistics.
- [106] De Loof K, De Baets B, De Meyer H, Brüggemann R. A hitchhiker's guide to poset ranking. Comb Chem High Throughput Screen. 2008;11(9):734-44. doi: 10.2174/138620708786306032.PMID:18991576.
- [107] della Queva, S. (2017). Analysis of Social Participation: A Multidimensional Approach Based on the Theory of Partial Ordering. In M. Fattore, R. Bruggemann Partial Order Concepts in Applied Sciences, Springer.
- [108] Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*. 38 (2): 325–339. doi:10.1214/aoms/1177698950.
- [109] T. Denoeux, M-H Masson (2010), Dempster-Shafer Reasoning in large Partially Ordered Sets: Applications in Machine Learning, Advances in Intelligent and Soft Computing, book series (AINSC) volume 68. Springer, Berlin, Heidelberg.
- [110] Di Brisco, A., & Farina, P. (2018). Measuring gender gap from a poset perspective. Social Indicators Research, 136, 1109–1124.
- [111] Diaz et al., Predicting Proteome-Early Drug Induced Cardiac Toxicity Relationships (Pro-EDICToRs) with Node Overlapping Parameters (NOPs) of a new class of Blood Mass-Spectra graphs ,The 11th International Electronic Conference on Synthetic Organic Chemistry session Computational Chemistry https://sciforum.net/paper/view/1371
- [112] Diego Colombo, Marloes H Maathuis, Markus Kalisch, and Thomas S Richardson. Learning highdimensional directed acyclic graphs with latent and selection variables. The Annals of Statistics, pages 294–321, 2012.
- [113] di Bella, E., Corsi, M., Leporatti, L. (2017). POSET Analysis of Panel Data with POSAC. In: Fattore, M., Bruggemann, R. (eds) Partial Order Concepts in Applied Sciences. Springer, Cham.https://doi.org/10.1007/978-3-319-45421-4_11.

- [114] S. Dong, P. Wang, and K. Abbas. 2021. A survey on deep learning and its applications. *Comput. Sci. Rev.* 40, C (May 2021).
- [115] B.J. Dutka, 1, K. Jones, 1, K.K. Kwan, 1, H. Bailey, 2, R. McInnis 1. Use of microbial and toxicant screening tests for priority site selection of degraded areas in water bodies, Water Research, Vol. 22; 4, 1988, 503-510. https://doi.org/10.1016/0043-1354(88)90047-4
- [116] W. Dyer, Weighted Posets: Learning surface order from dependency trees. Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019) https://api.semanticscholar.org/CorpusID: 204885369
- [117] U. Ergashev, E. C. Dragut, W. Meng, Learning To Rank Resources with GNN, WWW '23: Proceedings of the ACM Web Conference 2023, pp. 3247–3256, https://doi.org/10.1145/3543507.3583360
- [118] Fattore, M. (2008). Hasse diagrams, poset theory and fuzzy poverty measures. Rivista Internazionale Di Scienze Sociali, Anno, 116(1), 63–75.
- [119] Fattore, M., Brueggemann, R., OWSIŃSKI, J., (2011). Using poset theory to compare fuzzy multidimensional material deprivation across regions. In: Ingrassia S., Rocci R. and Vichi, M. New Perspectives in Statistical Modeling and Data Analysis. Berlin: Springer-Verlag.
- [120] Fattore, M., Maggino, F. and Colombo, E. (2012) "From Composite Indicators to Partial Orders: evaluating socio-economic phenomena through ordinal data". In: Maggino F, Nuvolati G (eds.) Quality of Life in Italy: research and reflections. Social Indicators Research Series, 48, 41–68.
- [121] Fattore, M., Arcagni, A., Barberis, S., (2014). Visualizing Partially Ordered Sets for Socioeconomic analysis. Revista Colombiana de Estadistica, 34(2), pp. 437–450.
- [122] Fattore, M., Arcagni, A., (2014). PARSEC: an R package for poset-based evaluation of multidimensional poverty. In: Bruggemann R., Carlsen L. and Wittmann J. Multi-Indicator Systems and Modelling in Partial Order. Berlin: Springer.
- [123] Fattore M.; Maggino, F.; Arcagni, A.(2015) Exploiting Ordinal Data For Subjective Well-Being Evaluation The Measurement of Subjective Well-Being in Survey Research Vol. 16, No. 3, pp. 409–428.
- [124] Fattore M., Maggino F., (2015). A new method for measuring and analyzing suffering Comparing suffering patterns in Italian society. In Anderson R. E. (Eds.). World Suffering and the Quality of Life. New York: Springer.
- [125] Fattore, M. (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. Social Indicators Research, 128(2), 835–858.
- [126] Fattore M., Maggino F., Arcagni A. (2016) "Non-aggregative assessment of subjective well-being", In G. Alleva, A. Giommi (eds.), Topics in Theoretical and Applied Statistics, Springer International Publishing Switzerland,
- [127] M. Fattore, R. Bruggemann. Partial Order Concepts in Applied Sciences, Springer, Cham, 2017, pp. 71–86. D0I10.1007/978-3-319-27274-0_20.
- [128] Fattore M. (2018) "Non-aggregated indicators of environmental sustainability", Silesian Statistical Review, 16(22), 7-22.
- [129] Fattore, M., & Arcagni, A. (2018). F-FOD: Fuzzy first order dominance analysis and populations ranking over ordinal multi-indicators system. *Social Indicators Research*, 144(1), 1–29.
- [130] Fattore M., Zenga Ma. (2019) "New posetic tools for the evaluation of financialliteracy", In Bianco A., Gnaldi M., Conigliaro. P. (eds.), Italian Studies on Quality of Life, Springer, 978-3-030-06021-3.
- [131] Fattore, M., Arcagni, A., & Maggino, F. (2019). Optimal scoring of partially ordered data, with an application to the ranking of smart cities. https://research.uniromal.it/pubblicazioni/49725.
- [132] F. Feng, X. He, Y. Liu, L. Nie, and T-S Chua. 2018. Learning on Partial-Order Hypergraphs. In WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France. ACM, New York, NY, USA 10 Pages.
- [133] Fisher, D.H., & Langley, P. (1985). Approaches to Conceptual Clustering. International Joint Conference on Artificial Intelligence.

- [134] H. Fisher. Knowledge Acquisition Via Incremental Conceptual Clustering. Machine Learning 2:139–172, 1987.
- [135] Finn, V.K., and others (1982). Many valued logics as fragments formalize semantics. Acta philosophica Fennica, vol.35.
- [136] V. K. Finn, M. I. Zabezhailo and O. M. Anshakov On a Computer-Oriented Formalization of Plausible Reasoning In F. Bacon-J. S. Mill'S Style (Main Principles and Computer Experiments) IFAC Artificial Intelligence. Leningrad. USSR 1983.
- [137] Finn, V.K.: Plausible Reasoning in Systems of JSM Type. Itogi Nauki i Tekhniki, Seriya Informatika 15, 54–101 (1991).
- [138] A. Formica, Ontology-based concept similarity in Formal Concept Analysis, Information Sciences, Volume 176, Issue 18, 2006, Pages 2624-2641, https://doi.org/10.1016/j.ins.2005.11.014.
- [139] Francés, O.; Abreu-Salas, J.; Fernández, J.; Gutiérrez, Y.; Palomar, M. Multidimensional Data Analysis for Enhancing In-Depth Knowledge on the Characteristics of Science and Technology Parks. Appl. Sci. 2023, 13, 12595. https://doi.org/10.3390/app132312595
- [140] Lerche, D. and P. Sorensen (2003) Evaluation of the ranking probabilities for partial orders based on random linear extensions. Chemosphere 53, 981-992.
- [141] Ganter, B., Wille, R.: Formal Concept Analysis, Mathematical Foundations. Springer-Verlag (1999).
- [142] H. Gao et al. "Graph-augmented Learning to Rank for Querying Large-scale Knowledge Graph." ArXiv abs/2111.10541 (2021):
- [143] Garriga, G.C. (2011). Formal Concept Analysis. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA.
- [144] Gennari, J., Langley, P., & Fisher, D. (1989). Models of incremental concept formation. Artificial Intelligence, 40:12-61.
- [145] P. Gepner, Machine Learning and High-Performance Computing Hybrid Systems, a New Way of Performance Acceleration in Engineering and Scientific Applications, Proceedings of the 16th Conference on Computer Science and Intelligence Systems, ACSIS, Vol. 25, pages 27–36 (2021).
- [146] R. Godin, R. Missaoui, H. Alaoui. Incremental concept formation algorithms based on Galois(concept) Lattice, Computational Intelligence Volume 11: 2, pp. 246–267.
- [147] Goncalves, B. and Porto, F. Research lattices: towards a scientific hypothesis data model. In Conference on Scientific and Statistical Database Management, SSDBM '13, Baltimore, MD, USA, July 29 31, 2013.
- [148] E. Goodwin et al., Compositional Generalization in Dependency Parsing, arXiv:2110.06843 (cs).
- [149] G. Grätzer, General Lattice Theory: Academic Press, Inc., New York. 1978.
- [150] Y. Guo, Z. Lin, J-G Lou, D. Zhang, Hierarchical Poset Decoding for Compositional Generalization in Language, 34^{th} conference on Neural Processing Systems(NeurlPS 2020) Vancouver Canada.
- [151] Guo et al. RankDNN: Learning to Rank for Few-Shot Learning, (2023) The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)
- [152] Hadzikadic, M., & Yun, D. (1989). Concept formation by incremental conceptual clustering. Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (pp. 831–836). Detroit, MI: Morgan Kaufmann.
- [153] M. R. Hacene, A. Napoli, P. Valtchev, Y. Toussaint and R. Bendaoud, "Ontology Learning from Text Using Relational Concept Analysis," 2008 International MCETECH Conference on e-Technologies (mcetech 2008), Montreal, QC, Canada, 2008, pp. 154-163, doi: 10.1109/MCETECH.2008.29.
- [154] M. Hajij et al., Higher-Order Attention Networks (2022), https://arxiv.org/abs/2206.00606v1.
- [155] M. Hajij et al., Topological Deep Learning: Going Beyond Graph Data (2023), arXiv: arXiv:2206.00606v3.
- [156] E. Halfon, M.G. Reggiani, On ranking chemicals for environmental hazard Environ. Sci. Technol. 1986, 20, 11, 1173–1179.

- [157] Halfon, E. (2006). Hasse Diagrams and Software Development. In: Brüggemann, R., Carlsen, L. (eds) Partial Order in Environmental Sciences and Chemistry. Springer, Berlin, Heidelberg.
- [158] Hanson, S., & Bauer, M. (1989). Conceptual clustering, categorization, and polymorphy. *Machine Learning*, 3:343–372.
- [159] B. A. Hassan, Ontology Learning Using Formal Concept Analysis and WordNet, arXiv:2311.14699, 2023.
- [160] Hilckmann, A., Bach, V., Bruggemann, R., Ackermann, R., & Finkbeiner, M. (2017). Partial order analysis of the government dependence of the sustainable development performance in Germany's federal states. In M. Fattore & R. Bruggemann (Eds.), (2017) Partial Order Concepts in Applied Sciences. Springer.
- [161] S. Hira, P.S. Deshpande, Data Analysis using Multidimensional Modeling, Statistical Analysis and Data Mining on Agriculture Parameters, Procedia Computer Science, Vol. 54, 2015, pp. 431-439, https://doi.org/10.1016/j.procs.2015.06.050.
- [162] T. Hirai a, F. Comim, Measuring the sustainable development goals: A poset analysis, Ecological Indicators, Vol. 145, 2022.
- [163] Huchard, M., Rouane-Hacène, M., Roume, C., Valtchev, P.: Relational concept discovery in structured datasets. Ann. Math. Artif. Intell. 49(1-4) (2007) 39-76.
- [164] D. I. Ignatov, Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields (2017), arXiv:1703.02819
- [165] Ikeda, M., & Yamamoto, A. (2013). Classification by Selecting Plausible Formal Concepts in a Concept Lattice.
- [166] Ip, E., Chen, S.-H., & Quandt, S. (2016). Analysis of multiple partially ordered responses to belief items with don't know option. *Psychometrika*, 81(2), 483–505.
- [167] T. Ivanciuc, O. Ivanciuc, D. J. Klein, Posetic Quantitative Superstructure/Activity Relationships (QSSARs) for Chlorobenzenes, J. Chem. Inf. Model. 2005, 45, 4, 870–879. https://doi.org/10.1021/ci0501342
- [168] Ivaldi, E., Ciacci, A., & Soliani, R. (2020). Urban deprivation in Argentina: A poset analysis. Papers in Regional Science, 99(6), 1723-1747.
- [169] S. D. Langhans, P. Reichert, N. Schuwirth, The method matters: A guide for indicator aggregation in ecological assessments, Ecological Indicators, Volume 45, 2014, pp. 494-507, https://doi.org/10.1016/j.ecolind.2014.05.014.
- [170] S. Jabin, Machine Learning methods and applications using Formal Concept Analysis, International Journal of New Technologies in Science and Engineering Vol. 2: 3, 2015,
- [171] Jaeger J, Tatsuoka C, Berns S, Varadi F, Czobor P, Uzelac S: Associating functional recovery with neurocognitive profiles identified using partially ordered classification models. *Schizophr Res* 2006, 85:40-48.
- [172] Jaeger J, Tatsuoka C, Berns SM, Varadi F: Distinguishing neurocognitive functions in schizophrenia using partially ordered classification models. Schizophr Bull 2006, 32:679-691
- [173] M.F. Janowitz., Semi-Flat Cluster Methods, Discrete Mathematics 21 (1978), 47-60.
- [174] M.F. Janowitz., An order theoretic model for cluster analysis. SIAM Journal on Applied Mathematics,34:55-72, 1978.
- [175] Janowitz, M.F. (2007). Cluster Analysis Based on Posets. In: Brito, P., Cucumel, G., Bertrand, P., de Carvalho, F. (eds) Selected Contributions in Data Analysis and Classification. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg.
- [176] M. F. Janowitz: Ordinal and Relational Clustering. Interdisciplinary Mathematical Sciences 10, World Scientific 2010, pp. 1-200.
- [177] R. Jardine, Posets, Metric Spaces, Topological Data Analysis(2020), http://www.sci.brooklyn.cuny.edu/~noson/CTslides/S
- [178] W. Kim, F. Memoli, Persistence Over Posets, Notices of American Mathematical Society, Vol 20, No. 8, (2023).
- [179] Jia, H., Newman, J., Tianfield, H. (2009). A new Formal Concept Analysis based learning approach to Ontology building. In: Sicilia, MA., Lytras, M.D. (eds) Metadata and Semantics. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-77745-0_42.

- [180] I. Jonyer, L. B. Holder, and D. J. Cook. Graph-Based Hierarchical Conceptual Clustering. *Proceedings of the Florida Artificial Intelligence Research Symposium*, pp. 91-95, 2000.
- [181] I. Jonyer, D. J. Cook, L. B. Holder, Graph-Based Hierarchical Conceptual Clustering, Journal of Machine Learning Research 2 (2001) 19-43.
- [182] Kainz, W., Egenhofer, M. J., and Greasley, I. (1993). Modelling spatial relations and operations with partially ordered sets. *International Journal of Geographical Information Systems*, 7(3):215–229.
- [183] H-M. Kaltenbach, Teaching Design of Experiments using Hasse diagrams arXiv:1912.08567v1
- [184] Kang, X., Li, D., & Wang, S. (2011). A multi-instance ensemble learning model based on concept lattice. Knowl. Based Syst., 24, 1203-1213.
- [185] S. Kardaetz, T. Strube, R. Bruggemann, G. Nützmann, Ecological scenarios analyzed and evaluated by a shallow lake model, J. Environ. Manag. 88 (2008) 120–135.
- [186] J. Karhunen, T. Raiko, K. Cho, Unsupervised deep learning: A short review, Advances in Independent Component Analysis and Learning Machines, Academic Press, 2015, Pages 125-142,https://doi.org/10.1016/B978-0-12-802806-3.00007-5.
- [187] Y. Kashnitsky and D. I. Ignatov, "Can FCA-based Recommender System Suggest a Proper Classi er? 2 Multiple Classi er Systems," FCA do Artif. Intell. 2014), p. 17, 2014.
- [188] M. Kendall, A new measure of rank correlation, Biometrika, Volume 30, Issue 1-2, June 1938, Pages 81–93 (doi:10.1093/biomet/30.1-2.81)
- [189] Kickmeier-Rust, M.D., Albert, D. (2013). Using Hasse Diagrams for Competence-Oriented Learning Analytics. In: Holzinger, A., Pasi, G. (eds) Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data. HCI-KDD 2013. Lecture Notes in Computer Science, vol 7947. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39146-0_6
- [190] Kim, W., Memoli, F.: Persistence over posets. Notices of the American Mathematical Society 2761, 1214–1224 (September 2023).
- [191] T. N Kipf and M. Welling. 2017. Semi-supervised classification with graph convolutional networks. ICLR (2017).
- [192] Klein, D. J. Similarity and Dissimilarity in Posets. J. Math. Chem. 1995, 18, 321-348.
- [193] Klein, D. J.; Babic, D. Partial Orderings in Chemistry. J. Chem. Inf. Comput. Sc. 1997, 37, 656–671.
- [194] D. J. Klein, Prolegomenon on partial orderings in chemistry, MATCH Commun. Math. Comput. Chem. 42 (2000) 1–290.
- [195] Klein, D. J., & Ivanciuc, T. (2006). Directed reaction graphs as posets. In R. Bruggemann & L. Carlsen (Eds.), Partial order in environmental sciences and chemistry (pp. 35–57). Berlin: Springer.
- [196] Khatri, Minal, "Formal Concept Analysis for Image Classification and Machine Learning Models for Anti-CRISPR Protein Discovery in Bioinformatics" (2023). Dissertations and Doctoral Documents from University of Nebraska-Lincoln, 2024–. 47.
- [197] Kotsianti, S.B., Kanellopoulos, D. (2007). Combining Bagging, Boosting and Dagging for Classification Problems. In: Apolloni, B., Howlett, R.J., Jain, L. (eds) Knowledge-Based Intelligent Information and Engineering Systems. KES 2007. Lecture Notes in Computer Science(), vol 4693. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74827-4_62
- [198] V. Kristina Et Al., "Features of PyHasse software used for the evaluation of chemicals in human breast milk samples," Simulation in Umwelt- und Geowissenschaften, vol.141, Hamburg, Germany, pp.169-180, 2012.
- [199] R. Kurtz(2020), Contributions to Semantic Dependency Parsing. https://www.diva-portal.org/smash/get/diva2:1457198/
- [200] S. O. Kuznetsov, Mathematical aspects of concept analysis, Journal of Mathematical Sciences 80(2),(1996).
- [201] Kuznetsov, S.O. Machine Learning on the Basis of Formal Concept Analysis. Automation and Remote Control 62, 1543–1564 (2001). https://doi.org/10.1023/A:1012435612567
- [202] Kuznetsov, S.O, Machine Learning and Formal Concept Analysis Conference: Concept Lattices, Second International Conference on Formal Concept Analysis, ICFCA 2004, Sydney, Australia, February 23-26, 2004.

- [203] S. Kuznetsov, Ordered Sets for Data Analysis(2019), arXiv:1908.11341.
- [204] G. Lebanon, J. Lafferty, Conditional models on the ranking poset, NIPS'02: Proceedings of the 15th International Conference on Neural Information Processing Systems, January 2002, pp. 431–438.
- [205] Lebowitz, M. (1986). Concept learning in a rich input domain: generalization-based memory. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach* (Vol. 2). Morgan Kaufmann, San Mateo, CA.
- [206] C. Lei et al., The Application of Multidimensional Data Analysis in the EIA Database of Electric Industry, 2011 3rd International Conference on Environmental Science and Information Application Technology (ESIAT 2011), In: Procedia Environmental Sciences 10 (2011) 1210 – 1215.
- [207] Levy, S. (1985). Partial order analysis of crime indicators. Social Indicators Research, 16(2), 195–199.
- [208] Linear Extensions https://mathworld.wolfram.com/LinearExtension.html.
- [209] J. Liu, Q. Zhang, W. Wang, L. McMillan, J. Prins, Clustering pair-wise dissimilarity data into partially ordered sets. In Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2006) 637–642.
- [210] Liu, Q., An, S., Lou, J.-G., Chen, B., Lin, Z., Gao, Y., Zhou, B., Zheng, N., and Zhang, D. Compositional generalization by learning analytical expressions. In Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20, Red Hook, NY, USA, 2020.
- [211] Maarek, Y.S. (1990). An Incremental Conceptual Clustering Algorithm that Reduces Input-Ordering Bias. In: Golumbic, M.C. (eds) Advances in Artificial Intelligence. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-9052-7_7
- [212] H. Mannila, C. Meek, Global partial orders from sequential data, KDD'00: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining, August 2000, pages 161–168.
- [213] L. Markowsky and G. Markowsky, "Lattice Data Analytics and an Exploratory Analysis of the Carver2 Dataset," to appear in IDAACS 2019.
- [214] Markov, A lattice-based approach to hierarchical clustering. Proceedings of the Florida Artificial Intelligence Research Symposium, pp. 389–393, 2001.
- [215] Martin, J., & Billman, D. (1994). Acquiring and combining overlapping concepts. *Machine Learning*, 16(1–2):121–155.
- [216] McKusick, K., & Langley, P. (1991). Constraints on tree structure in concept formation. *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence* (pp. 810–816). Sydney, Australia: Morgan Kaufmann.
- [217] Meddouri, N., Maddouri, M. (2009). Boosting Formal Concepts to Discover Classification Rules. In: Chien, BC., Hong, TP., Chen, SM., Ali, M. (eds) Next-Generation Applied Intelligence. IEA/AIE 2009. Lecture Notes in Computer Science(), vol 5579. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02568-6_51.
- [218] N. Meddouri, H. Khoufi, M. Maddouri, Parallel Learning and Classification for Rules based on Formal Concepts Procedia Computer Science Volume 35, 2014, pp. 358-367.
- [219] J-P. Métivier et al., Discovering Structural Alerts for Mutagenicity Using Stable Emerging Molecular Patterns, J. Chem. Inf. Model. 2015, 55, 5, 925–940, 2015 https://doi.org/10.1021/ci500611v.
- [220] R.S. Michalski, Knowledge Acquisition Through Conceptual Clustering: A Theoretical Framework and an Algorithm for Partitioning Data into Conjunctive Concepts. *International Journal for Policy Analysis and Information Systems*. Vol. 4, No. 3, 1980.
- [221] Michalski and R. E. Stepp. Learning From Observation: Conceptual Clustering. In R.S.Michalski, J.G. Carbonell, and T.M. Mitchell (Eds.), Machine Learning: An Artificial Intelligence Approach, Volume 1, Tioga Publishing Company, pp. 331-363, 1983.
- [222] V. Mnih, N. Heess, A. Graves, and k. kavukcuoglu, "Recurrent models of visual attention," in 27th Annual Conference on Neural Information Processing Systems (NIPS 2014). Curran Associates, Inc., 2014, pp. 2204–2212.

- [223] A. M. Mwafise, G-S Cheon, H. J. Choi, S. Giraudo, Operad Structure of Poset Matrices (2024), arXiv:2401.06814.
- [224] Mollica, C., & Tardella, L. (2020). PLMIX: an R package for modelling and clustering partially ranked data. Journal of Statistical Computation and Simulation, 90(5), 925–959. https://doi.org/10.1080/00949655.2020.1711909.
- [225] M. Nabeel et al., A survey of ontology learning techniques and applications, Database, Volume 2018, 2018, bay101, https://doi.org/10.1093/database/bay101
- [226] Nardo, M., M. Saisana, A. Saltelli and S. Tarantola: 2005b, Tools for Composite IndicatorsBuilding. European Commission, report EUR 21682 EN (Joint Research Centre, Ispra,Italy).
- [227] Newlin, J.T., Patil, G.P. Application of partial order to stream channel assessment at bridge infrastructure for mitigation management. Environ Ecol Stat 17, 437–454 (2010). https://doi.org/10.1007/s10651-010-0162-8.
- [228] Nicholls GK, Lee JE, Karn N, Johnson D, Huang R, Muir-Watt A, Bayesian inference for partial orders from random linear extensions: power relations from 12th Century Royal Acta, , 2022. arXiv: 2212.05524.
- [229] Onishchenko, A.A., Gurov, S.I. Classification based on formal concept analysis and biclustering: possibilities of the approach. Comput Math Model 23, 329–336 (2012). https://doi.org/10.1007/s10598-012-9141-2.
- [230] Ouyang, C.; Liu, Y., Formal Concept Analysis Supporting Ontology Learning From Database, Advanced Science Letters, Volume 7 2012, pp. 473-477(5).
- [231] V. Padhye, K. Lakshmanan, A deep actor critic reinforcement learning framework for learning to rank, Neurocomputing, Volume 547, (2023.) https://doi.org/10.1016/j.neucom.
- [232] Pakkar, M.S. A Hierarchical Aggregation Approach for Indicators Based on Data Envelopment Analysis and Analytic Hierarchy Process. Systems 2016, 4, 6. https://doi.org/10.3390/systems4010006
- [233] G. P. Patil, C. Taillie, Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization, *Environ. Ecol. Stat.* 11 (2004) 199–228.
- [234] G. Patil, W. L. Myers, R. Bruggemann, Multivariate datasets for inference of order: some considerations and explorations, in: R. Bruggemann, L. Carlsen, J. Wittmann (Eds.), Multi-Indicator Systems and Modelling in Partial Order, Springer, New York, 2014,pp.13–45.
- [235] L. Pellegrina, C. Cousins, F. Vandin, M. Riondato. MCRapper: Monte-Carlo Rademacher Averages for Poset Families and Approximate Pattern Mining. Extended version. https://arxiv.org/abs/2006.09085.
- [236] Popescu-Spineni, S. (1998). Hierarchy Techniques of Multidimensional Data Analysis (MDA) in Social Medicine Research. In: Rizzi, A., Vichi, M., Bock, HH. (eds) Advances in Data Science and Classification. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-72253-0_87
- [237] O. Prokasheva, A. Onishchenko, S. Gurov, Classification Methods Based on Formal Concept Analysis, Conference: FCAIR 2013 – Formal Concept Analysis Meets Information Retrieval. Workshop co-located with the 35th European Conference on Information Retrieval (ECIR 2013)At: Moscow.
- [238] S. Pudenz, ProRank software for partial order ranking, MATCH Commun. Math. Comput. Chem. 54 (2005) 611–622.
- [239] Qiu, J., Wu, Q., Ding, G. et al. A survey of machine learning for big data processing. EURASIP J. Adv. Signal Process. 2016, 67 (2016). https://doi.org/10.1186/s13634-016-0355-x.
- [240] N.R Rashidah, A Comparison Between Single Linkage and Complete Linkage in Agglomerative Hierarchical Cluster Analysis for Identifying Tourists Segments, IIUM Engineering Journal, Vol. 12, No. 6, 2011: Special Issue in Science and Ethics.
- [241] A. Raveh and S. F. Landau, Partial Order Scalogram Analysis with Base Coordinates (POSAC): Its Application to Crime Patterns in All the States in the United States, *Journal of Quantitative Criminology* Vol. 9, No. 1., pp. 83-99 (1993).
- [242] G. Restrepo & R. Brüggemann, Ranking regions using cluster analysis, Hasse diagram technique and topology, 3^{rd} International Congress on Environmental Modelling and Software, Burlington, Vermont, USA 2006.

- [243] G. Restrepo, R. Bruggemann, M. Weckert, S. Gerstmann, H. Frank, Ranking patterns, an application to refrigerants, MATCH Commun. Math. Comput. Chem. 59 (2008) 555–584.
- [244] Rimoldi S.M.L., Arcagni A., Fattore M., Barbiano di Belgiojoso E. (2020) "Targeting policies for multidimensional poverty and social fragility relief among migrants in Italy, using F-FOD analysis", Social Indicators Research - doi10.1007/s11205-020-02485-7
- [245] Rimoldi S.M.L., Arcagni A., Fattore M., Terzera T. (2020) "Social and material vulnerability of the Italian municipalities: comparing alternative approaches", Social Indicators Researchdoi:10.1007/s11205-020-02330-x.
- [246] Sagi O, Rokach L. 2018. Ensemble learning: a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8(4):e1249
- [247] G.-C. Rota, On the foundations of combinatorial theory I. Theory of Mobius functions, Z. Wahrsch. Verw. Gebiete 2 (2964) 340-368.
- [248] Rouane-Hacene, M., Valtchev, P., Nkambou, R.: Supporting Ontology Design through Large-Scale FCA-Based Ontology Restructuring. In: ICCS 2011. (2011) 257–269.
- [249] M. D. Ruiz, E. Hüllermeier, A formal and empirical analysis of the fuzzy gamma rank correlation coefficient, Information Sciences, Vol. 206, 2012, pp. 1-17, https://doi.org/10.1016/j.ins.2012.04.006.
- [250] F. Ruskey, Generating linear extensions of posets by transpositions, Journal of Combinatorial Theory, Series B, Volume 54, Issue 1, 1992, Pages 77-101,https://doi.org/10.1016/0095-8956(92)90067-8.
- [251] I.M. Sabara, F. Rozi, M. N. Jauhari, Agglomerative Hierarchical Clustering Analysis Based on Partially-Ordered Hasse Graph of Poverty Indicators in East Java, Proceedings of the 12th International Conference on Green Technology (ICGT 2022).
- [252] K. Sachs, O. Perez, D. Lauffenburger, G. Nolan, Causal protein-signaling networks derived from multiparameter single-cell data. Science. https://www.science.org/doi/10.1126/science.1105809#supplementary-materials.
- [253] Salman, H.E. Leveraging a combination of machine learning and formal concept analysis to locate the implementation of features in software variants 2023, Information and Software Technology, vol. 164.
- [254] Sahami, M. (1995). Learning classification rules using lattices (Extended abstract). In: Lavrac, N., Wrobel, S. (eds) Machine Learning: ECML-95. ECML 1995. Lecture Notes in Computer Science, vol 912. Springer, Berlin, Heidelberg.https://doi.org/10.1007/3-540-59286-5_83.
- [255] Sangroya, A., Anantaram, C., Rawat, M., & Rastogi, M. (2019). Using Formal Concept Analysis to Explain Black Box Deep Learning Classification Models. FCA4AI@IJCAI.
- [256] A. Santana, E. Colombini, Neural Attention Models in Deep Learning: Survey and Taxonomy(2021), arXiv:2112.05909v1.
- [257] Semeraldi et al., Partial Order Rank Features in Colour Space, Appl. Sci. 2020, 10(2), 499; https://doi.org/10.3390/app10020499
- [258] Sen, A. (1970b). Interpersonal aggregation and partial comparability. Econometrica, 38(3), 393–409.
- [259] Seymour, R. B. (1960). Missing Data in Non-Linear Trend Analysis of Repeated Measurements on the Same Individuals. The Journal of Educational Research, 54(4), 141–144. http://www.jstor.org/stable/27530403.
- [260] Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976).
- [261] Silan M, Boccuzzo G, Arpino B. Matching on poset-based average rank for multiple treatments to compare many unbalanced groups. *Statistics in Medicine*. 2021;40(28):6443–6458. doi: 10.1002/sim.9192.
- [262] U. Simon, R. Bruggemann, S. Mey, S. Pudenz, METEOR- application of a decision support tool based on discrete mathematics, MATCH Commun. Math. Comput. Chem. 54 (2005) 623–642.
- [263] Simon, U.; Brüggemann, R.; Behrendt, H.; Shulenberger, E.; Pudenz, S. METEOR: a step-by-step procedure to explore effects of indicator aggregation in multi criteria decision aiding application to water management in Berlin, Germany. Acta hydrochim. Hydrobiol. 2006, 34, 126-136.
- [264] Song et al., Graph-based Semi-supervised Learning: A Comprehensive Review 2021.

- [265] Sorensen, P. B.; Mogensen, B. B.; Carlsen, L.; Thomsen, M. The Influence on Partial Order Ranking from Input Parameter Uncertainty Definition of a Robustness Parameter. *Chemosphere* **2000**, 595–600.
- [266] P. B. Sorensen, R. Bruggemann, L. Carlsen, B. B. Mogensen, J. Kreuger, S. Pudenz, Analysis of monitoring data of pesticide residues in surface waters using partial order ranking theory, *Envir. Tox. Chem.* 22 (2003) 661–670.
- [267] Sorensen, P. B.; Brüggemann; M. Thomsen; D. B. Lerche, Application of multidimensional rank-correlation. MATCH Communications in Mathematical and in Computer Chemistry, 54(3), 643-670.
- [268] M. Sugiyama, Machine Learning and Information Geometry II, https://mahito.info/files/Sugiyama_NII_IISS_2018_02.pd
- [269] J.L. Sullivan & S. Feldman Multiple Indicators An Introduction (1979), https://www.ojp.gov/ncjrs/virtual-library/abstracts/multiple-indicators-introduction.
- [270] Taeb, A.; Buhlmann, P.; Chandrasekaran, V. Model Selection over Partially Ordered Sets(2024), ArXiV: 2308.10375 .
- [271] Tatsuoka C: Sequential classification on partially ordered sets. PhD thesis Cornell University, Statistics Department; 1996.
- [272] Tatsuoka C. Data analytic methods for latent partially ordered classification models. Applied Statistics (Journal of the Royal Statistical Society Series C). 2002;51:337–350.
- [273] Tatsuoka C, Ferguson T: Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society*, Series B 2003, 65:143-157.
- [274] Tatsuoka C, Corrigendum:Data analytic methods for latent partially ordered classification models, Appl.Statist.(2005) **54**,Part2,pp.465–467
- [275] Tatsuoka C. Diagnostic models as partially ordered sets. Measurement. (2009) 7:49-53.
- [276] Tatsuoka C, Varadi F, Jaeger J. Latent partially ordered classification models and normal mixtures. J Edu Behav Stat. (2013) 38:267–94.
- [277] Tatsuoka C, Tseng H, Jaeger J, Varadi F, Smith MA, Yamada T, Smyth KA,Lerner AJ. com AsDNInn: Modeling the heterogeneity in risk of progres-sion to Alzheimer's disease acrosscognitive profiles in mild cognitive impairment. Alzheimer's Res Ther. 2013;5:1–19.
- [278] Tatsuoka C. Sequential classification on lattices with experiment-specific response distributions. Sequent Anal Design Methods Appl. (2014) 33:400–20.
- [279] Tatsuoka C, McGowan B, Yamada T, Espy KA, Minich N, Taylor HG. Effects of extreme prematurity on numerical skills and executive function in kindergarten children: an application of partially ordered classification modeling. Learn Individ Differ. (2016) 49:332–40.
- [280] K. M. Ting, I. Witten, Stacking Bagged and Dagged Models (1997), https://www.researchgate.net/publication/2516354_S-
- [281] F Tombari (2023), Tame representations in Topological Data Analysis, https://www.diva-portal.org/smash/get/diva2:1759244/FULLTEXT01.pdf.
- [282] F. J. Torres-Rojas & Castro-Mora, Partially Ordered Sets and Logical Clocks for Distributed Systems
- [283] Tovar, M., Pinto, D., Rendón, A.M., Serna, J.G., & Ayala, D.V. (2014). Identification of Ontological Relations Using Formal Concept Analysis. *LANMR*.
- [284] Tsakovski, V. Simeonov, Hasse diagrams as explanatory tool in environmental data mining: A case study, in: J. Owsinski, R. Bruggemann (Eds.) Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues, Sys. Res. Inst. Polish Acad. Sci., Warsaw, 2007, pp. 50–68.
- [285] Pirintsos, S., Bariotakis, M., Kalogrias, V., Katsogianni, S., Brüggemann, R. (2014). Hasse Diagram Technique Can Further Improve the Interpretation of Results in Multielemental Large-Scale Biomonitoring Studies of Atmospheric Metal Pollution. In: Brüggemann, R., Carlsen, L., Wittmann, J. (eds) Multi-indicator Systems and Modelling in Partial Order. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-8223-9_11
- [286] van Engelen, J.E., Hoos, H.H. A survey on semi-supervised learning. Mach Learn 109, 373-440 (2020). https://doi.org/10.1007/s10994-019-05855-6

- [287] K. Voigt, G. Welzl, R. Bruggemann, Data analysis of environmental air pollutant monitoring systems in Europe, *Environmetrics* **15**(2004) 577–596.
- [288] K. Voigt, R. Bruggemann, Ranking of pharmaceuticals detected in the environment: Aggregation and weighting procedures, Combin. Chem. High Through. Screen. 11 (2008) 770–782.
- [289] K. Voigt, R. Bruggemann, M. Kirchner, K.-W. Schramm, Influence of altitude concerning the contamination of humus soils in the German Alps: a data evaluation approach using PyHasse, Environ. Sci. Pollut. Res. 17 (2010) 429–440.
- [290] K. Voigt, R. Bruggemann, H. Scherb, H. Shen, K. H. Schramm, Evaluating therelationship between chemical exposure and cryptorchidism by discrete mathematical method using PyHasse software, J. Environ. Model. Soft. 25 (2010) 1801–1812.
- [291] Voigt, K.; Brüggemann, R. Water contamination with pharmaceuticals: data availability and evaluation approach with Hasse diagram technique and METEOR. MATCH Commun. Math. Comput. Chem. 2005, 54, 671-689.
- [292] Voigt, Kristina et al. "A multi-criteria evaluation of environmental databases using the Hasse Diagram Technique (ProRank) software." Environ. Model. Softw. 21 (2006): 1587-1597.
- [293] K. Voigt et al., Application of the PYHASSE Program Features: Sensitivity, Similarity, and Separability for Environmental Health Data, Statistica & Applicazioni Special Issue, 2011, pp. 155-168.
- [294] X. Wang and A. Gupta. Unsupervised learning of visual representations using videos. In ICCV, 2015.
- [295] X. Wang, P. Li, M. Zhang, Improving Graph Neural Networks on Multi-node Tasks with Labeling Tricks, (2023) arXiv:2304.10074. Available in: https://arxiv.org/abs/2304.10074
- [296] C. Wendler, Machine Learning on non-Euclidean domains: powersets, lattices, posets, (2023) ETH Zurich.
- [297] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in: I. Rival (Ed.), Ordered Sets, D. Reidel Publishing, Dordrecht, 1982, pp. 445–470.
- [298] J. Wittmann, J. Markert, S. Plura, R. Brüggemann, A Software Platform Towards a Comparison of Cars-A Case Study for Handling Ratio-Based Decisions, EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information, Shaker Verlag Aachen. Available in: http://enviroinfo.eu/sites/default/files/pdfs/vol6919/0467.pdf.
- [299] Wu et al., A Comprehensive Survey on Graph Neural Networks, (2019) arXiv:1901.00596.
- [300] Wu, C., Li, H., & Ren, J. (2021). Research on hierarchical clustering method based on partially-ordered Hasse graph. Future Generation Computer Systems. Doi: https://doi.org/10.1016/j.future.2021.07.025
- [301] Z. Xie, W. Hsu, Z. Liu, M. L. Lee: Concept Lattice based Composite Classifiers for high Predictability. Artificial Intelligence, vol. 139, pp. 253–267, Wollongong, Australia (2002).
- [302] Xiao, S., Yan, J., Farajtabar, M., Song, L., Yang, X., & Zha, H. (2017). Joint Modeling of Event Sequence and Time Series with Attentional Twin Recurrent Neural Networks. ArXiv, abs/1703.08524.
- [303] Y. Yadav, A. Samal, E. Saucan, A Poset-Based Approach to Curvature of Hypergraphs, Symmetry, 14(2):420, 2022.
- [304] M. Yang, R. Wang, Y. Shen, H. Qi and B. Yin, "Breaking the Expression Bottleneck of Graph Neural Networks," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 5652-5664, 1 June 2023, doi: 10.1109/TKDE.2022.3168070.
- [305] Yoneda, Y., Sugiyama, M., & Washio, T. (2018). Learning Graph Representation via Formal Concept Analysis. *ArXiv*, *abs/1812.03395*.
- [306] H. Yu, Y. Chen, P. Lingras, G. Wang, A three-way cluster ensemble approach for large-scale data, International Journal of Approximate Reasoning, Volume 115, 2019, Pages 32-49, https://doi.org/10.1016/j.ijar.2019.09.001.
- [307] L. A. Zadeh, Fuzzy Sets, Information and Control, 8, 338–353 (1965).
- [308] J. Zhang , Learning to Rank Cases with Classification Rules(2008), http://www.ecmlpkdd2008.org/files/pdf/workshops/pl/8.pdf.

- [309] L. Zhang, D. Wang, X. Liu, Computing Concept Lattices with Clustering Approaches.
- [310] Zhao, M., Zhang, S., Li, W. et al. Matching biomedical ontologies based on formal concept analysis. J Biomed Semant 9, 11 (2018). https://doi.org/10.1186/s13326-018-0178-9.
- [311] Zia, A., Khamis, A., Nichols, J. et al. Topological deep learning: a review of an emerging paradigm. Artif Intell Rev 57, 77 (2024).https://doi.org/10.1007/s10462-024-10710-9.
- [312] Y. Zuo and R. Serfling, General Notions of Statistical Depth Function, *The Annals of Statistics*, Vol. 28, No. 2 (Apr., 2000), pp. 461-482.
- [313] UCI machine learning repository, http://www.ics.uci.edu/~mlearn/MLRepository.html.
- [314] Kohonen, Teuvo (January 2013). "Essentials of the self-organizing map". Neural Networks. 37: 52–65. doi:10.1016/j.neunet.2012.09.018.
- [315] MPII Human Pose Dataset, http://human-pose.mpi-inf.mpg.de/.
- [316] Leeds Sports Poset Dataset https://datasets.activeloop.ai/docs/ml/datasets/lsp-dataset/.
- [317] Olympic Sports Dataset http://vision.stanford.edu/Datasets/OlympicSports/.
- $[318] \ \ Compositional \ Freebase \ Questions \ (CFQ) \ Dataset, \\ \textbf{https://github.com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-research/tree/master/com/google-research/google-google-research/google-research/google-research/google-google-research/google-research/googl$
- [319] Universal Dependencies Corpora https://www.tensorflow.org/datasets/catalog/universal_dependencies
- [320] Istat. (2018a). Indicators for the measurement of equitable and sustainable well-being. https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being/indicators.
- [321] Eurobarometer dataset for service performance https://www.gesis.org/en/eurobarometer-data-service/
- [322] Python implementaion of deep supervised visual similarity learning(2017) In :https://github.com/asanakoy/deep_unsupervised_posets.
- [323] C. Squires, causaldag (2018), https://github.com/uhlerlab/causaldag.
- [324] Poset-Hypergraph-Curvature Datasets, https://github.com/asamallab/Poset-Hypergraph-Curvature.