Machine Learning for Health Informatics

Andreas Holzinger[1,2(✉)]

[1] Holzinger Group, HCI-KDD, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Graz, Austria
a.holzinger@hci-kdd.org
[2] Institute for Information Systems and Computer Media, Graz University of Technology, Graz, Austria

Abstract. Machine Learning (ML) studies algorithms which can learn from data to gain knowledge from experience and to make decisions and predictions. Health Informatics (HI) studies the effective use of probabilistic information for decision making. The combination of both has greatest potential to rise quality, efficacy and efficiency of treatment and care. Health systems worldwide are confronted with "big data" in high dimensions, where the inclusion of a human is impossible and automatic ML (aML) show impressive results. However, sometimes we are confronted with complex data, "little data", or rare events, where aML-approaches suffer of insufficient training samples. Here interactive ML (iML) may be of help, particularly with a *doctor-in-the-loop,* e.g. in subspace clustering, k-Anonymization, protein folding and protein design. However, successful application of ML for HI needs an *integrated* approach, fostering a concerted effort of four areas: (1) data science, (2) algorithms (with focus on networks and topology (structure), and entropy (time), (3) data visualization, and last but not least (4) privacy, data protection, safety & security.

Keywords: Machine learning · Health informatics

1 Introduction and Motivation

Since the early days of Machine Learning (ML) in the 1950ies [1] the goal was to learn from data, to gain knowledge from experience and to make predictions. The field accelerated by the introduction of *statistical learning theory* in the late 1960ies; although it was at that time a purely theoretical analysis of the problem of *function estimation* from a given collection of data [2]. With the introduction of new statistical learning algorithms (e.g. support vector machine [3]) statistical learning theory became more and more interesting as a tool for developing algorithms of practical use for the estimation of multidimensional functions [4].

Today, ML is the most growing subfield in computer science and Health Informatics (HI) is the greatest application challenge [5,6]. This is not surprising, because in the health domain we are confronted with probabilistic, uncertain,

A. Holzinger (Ed.): ML for Health Informatics, LNAI 9605, pp. 1–24, 2016.
DOI: 10.1007/978-3-319-50478-0_1

unknown, incomplete, heterogenous, noisy, dirty, unwanted and missing data sets which endangers the modelling of artifacts. Moreover, in the biomedical world we are confronted with a further problem: time. Whilst most computational approaches assume homogeneity in time, people and processes in the health domain are not homogenous in time and cannot be forecasted, sometimes it can happen the completely unexpected. That makes automatic solutions in this domain difficult, yet sometimes impossible.

A grand challenge in HI is to discover relevant *structural* patterns and/or *temporal* patterns ("knowledge") in such data, which are often hidden and not accessible to the human expert but would be urgently needed for better decision support. Another problem is that most of the data sets in HI are weakly-structured and non-standardized [7], and most data is in dimensions much higher than 3, and despite human experts are excellent at pattern recognition in dimensions of $\leq$3, high dimensional data sets make manual analysis difficult, yet often impossible.

The adoption of data-intensive methods can be found throughout various branches of health, leading e.g. to more evidence-based decision-making and to help to go towards personalized medicine [8]: A grand goal of future biomedicine is to tailor decisions, practices and therapies to the individual patient. Whilst personalized medicine is the ultimate goal, stratified medicine has been the current approach, which aims to select the best therapy for groups of patients who share common biological characteristics. Here, ML approaches are indispensable, for example *causal inference trees (CIT)* and aggregated grouping, seeking strategies for deploying such stratified approaches. Deeper insight of personalized treatment can be gained by studying the personal treatment effects with *ensemble CITs* [9]. Here the increasing amount of heterogenous data sets, in particular "-omics" data, for example from genomics, proteomics, metabolomics, etc. [10] make traditional data analysis problematic and optimization of knowledge discovery tools imperative [11,12]. On the other hand, many large data sets are indeed large collections of small data sets. This is particularly the case in personalized medicine where there might be a large amount of data, but there is still a relatively small amount of data for each patient available [13]. Consequently, in order to customize predictions for each individual it is necessary to build a model for each patient along with the inherent uncertainties, and to couple these models together in a hierarchy so that information can be "borrowed" from other similar patients. This is called *model personalization*, and is naturally implemented by using hierarchical Bayesian approaches including e.g. hierarchical Dirichlet processes [14] or Bayesian multi-task learning [15].

This variety of problems in the application of ML for HI requires a synergistic combination of various methodological approaches which are combined in the HCI-KDD approach, which is described in Sect. 3. In Sect. 4 an example curriculum is briefly discussed and Sect. 5 provides an outlook to three future challenges.

2 Glossary and Key Terms

automatic Machine Learning (aML) in bringing the human-out-of-the-loop is the grand goal of ML and works well in many cases with "big data" [16].

Big Data is a buzz word to indicate the flood of data today; however, large data sets are necessary for aML approaches to learn effectively, the problem is rather in "dirty data" and sometimes we have large collections of "little data".

Cognitive Science mainly deals with questions of human intelligence, problem solving and decision making and is manifested to a large extent in the field of Human–Computer Interaction (HCI) [17].

Computer Science today has a large focus on machine learning algorithms and these are manifested to a large part in the field of Knowledge Discovery/Data Mining (KDD). *Deep Learning* allows models consisting of multiple layers to learn representations of data with multiple levels of abstraction, e.g. in speech recognition, visual object recognition, object detection, genomics etc. [6].

Dimensionality of data is high, when the number of features p is larger than the number of observations n by magnitudes. A good example for high dimensional data is gene expression study data [18].

Entropy quantifies the expected value of information contained in data and can be used as a measure of uncertainty, hence it is of tremendous importance for HI with many applications to discover e.g. anomalies in data [19].

Health has been defined by the World Health Organization (WHO) in 1946 as *"a state of complete physical, mental, and social well-being"* and is undeniably one of the most important aspects concerning every human [20].

Health Informatics is concerned with the use of computational intelligence for the management of processes relevant for human health and well-being, ranging from the collective to the individual [21].

interactive Machine Learning (iML) in bringing the human-in-the-loop is necessary if we have small amounts of data ("little data"), rare events or deal with complex problems [22,23].

Knowledge Discovery (KDD) includes exploratory analysis and modeling of data and the organized process to identify valid, novel, useful and understandable patterns from these data sets [24].

Topological Data Mining uses algebraic geometry to recover parameters of mixtures of high-dimensional Gaussian distributions [25].

Visualization can be defined as transforming the symbolic into the geometric and the graphical presentation of information, with the goal of providing the viewer with a qualitative understanding of the information contents [12,26].

3 The HCI-KDD Approach

The original idea of the HCI-KDD approach [8,27,28] is in combining aspects of the best of two worlds: Human–Computer Interaction (HCI), with emphasis on cognitive science, particularly dealing with *human intelligence*, and Knowledge Discovery/Data Mining (KDD), with emphasis on machine learning, particularly dealing with *computational intelligence* [29].

Cognitive science (CS) studies the principles of human learning from data to understand intelligence. The Motto of Demis Hassabis from Google Deepmind is *"Solve intelligence - then solve everything else"* [30]. Our natural surrounding is in $\mathbb{R}^3$ and humans are excellent in perceiving patterns out of data sets with dimensions of $\leq$3. In fact, it is amazing how humans extract so much knowledge from so little data [31] which is a perfect motivator for the concept of iML.

The problem in HI is that we are challenged with data of arbitrarily high dimensions [7,18,32]. Within such data, relevant *structural* patterns and/or *temporal* patterns ("knowledge") are hidden, difficult to extract, hence not accessible to a human. A grand challenge is to bring the results from high dimensions into the lower dimension, where the health experts are working on 2D surfaces on different devices (from tablet to large wall-displays), which can represent data only in $\mathbb{R}^2$.

Machine Learning (ML) studies the principles of computational learning from data to understand intelligence [5]. Computational learning has been of general interest for a very long time, but we are far away from solving intelligence: facts are not knowledge and descriptions are not insight. A good example is the famous book by Nobel prize winner Eric Kandel *"Principles of Neural Science"* [33] which doubled in volume every decade - effectively, our goal should be to make this book shorter!

HCI and KDD did not harmonize in the past. HCI had its focus on specific experimental paradigms, embedded deeply in Cognitive *Science*; and aimed to be cognitively/neutrally plausible. KDD had its focus on computational learning problems and tried to optimize in the range of 1% because it was embedded in Computer *Engineering*, and aimed to have working systems to solve practical problems - whether mimicking the human brain or not.

Consequently, a concerted effort of both worlds and a comprehensive understanding of the data ecosystem along with a multi-disciplinary skill-set, encompassing seven specializations: (1) data science, (2) algorithms, (3) network science, (4) graphs/topology, (5) time/entropy, (6) data visualization and visual analytics, and (7) privacy, data protection, safety and security can be highly beneficial for solving the aforementioned problems (Fig.1).

3.1 Research Track 1 DAT: Data Preprocessing, Integration, Fusion

Understanding the data ecosystem is of eminent importance in HI. Considering the context in which the data is produced, we can determine between four large data pools: (1) Biomedical research data (e.g. clinical trial data, -omics data [10]), e.g. from genomic sequencing technologies (Next Generation Sequencing,

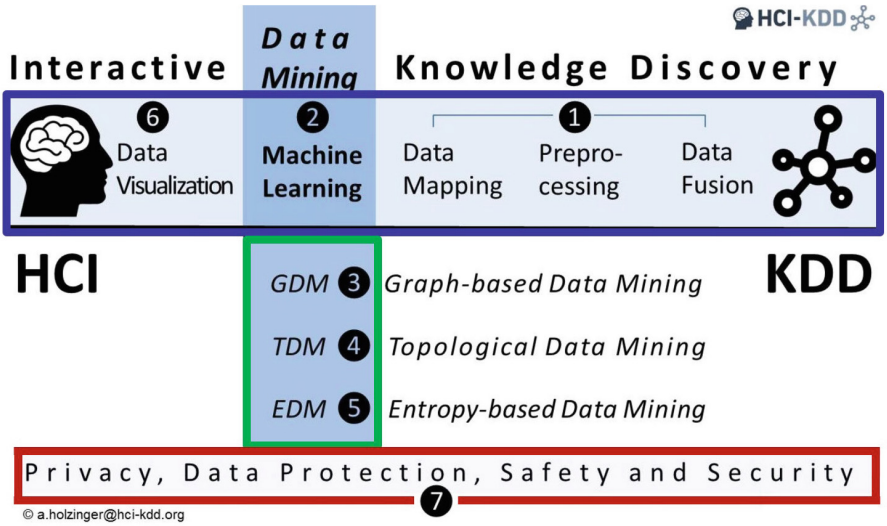

Fig. 1. The big picture of the HCI-KDD approach: The horizontal process chain (blue box) encompasses the whole machine learning pipeline from physical aspects of raw data, to human aspects of data visualization; while the vertical topics (green box) include important aspects of structure (graphs/networks), space (computational topology) and time (entropy); privacy, data protection, safety and security are mandatory topics within the health domain and provide kind of a base compartment (Color figure online) (Image taken from hci-kdd.org)

NGS etc.), microarrays, transcriptomic technologies, proteomic and metabolomic technologies, etc., which all plays important roles for biomarker discovery and drug design [34,35]. (2) Clinical data (e.g. patient records, clinicians documentations, medical terminologies (e.g. ICD, SNOMED-CT), medical surveys, laboratory tests, clinical and physiological parameters, ECG, EEG etc.), (3) Health business data (e.g. costs, utilization, management data, logistics, accounting, billing, resource planning, prediction etc.), and (4) private patient data, produced by various customers and stakeholders outside the clinical context (e.g., wellness data, Ambient Assisted Living data, sport data, insurance data, etc.) [36]. The US Department of Health and Human Services (HHS) created a taxonomy of health data with the following dimensions [37]: (1) Demographics and socio-economic Data: age, race, sex, education, etc. (2) Health Status Data: Health status of the patient, e.g., morbidities, problems, complaints, disabilities, diagnoses, symptoms, etc. (3) Health Resources Data: Characteristics and capacity of the health system, etc. (4) Healthcare Utilization Data: Characteristics(e.g., time, duration, tests, procedures, treatment) about medical care visits like discharge, stay, use of healthcare services, etc. (5) Healthcare Financing and Expenditure Data: Costs, charges, insurance status, etc. (6) Healthcare Outcomes of current and past prevention, treatments, etc. (7) Other data: -omics data, environmental exposures, etc.

Technically, there exist various levels of data structures [38] from physical level as basic indissoluble unit (bit, Shannon) to the logical level (Booleans, integers, floating-point numbers, strings, etc.) and conceptual (abstract) Level (arrays, lists, trees, graphs, etc.). Finally the technical level is the application data (text, graphics, images, audio, video, multimedia) an the Hospital Level includes narrative ("free text") patient record data (structured/unstructured and standardized/non-standardized), -omics data (genomics, proteomics, metabolomics, lipidomics, transcriptomics, microbiomics, fluxomics, phenomics, cytomics, connectomics, environomics, exposomics, exonomics, foodomics, toponomics, etc.), numerical measurements (physiological data, lab results, vital signs, etc.), recorded signals (ECG, EEG, EOG, etc.), Images (standard X-ray, MR, CT, PET, SPECT, microscopy, confocal laserscans, ultrasound imaging, molecular imaging, etc.)

Data preprocessing is often a required first step for machine learning because ML algorithms learn from data and the learning outcome for problem solving heavily depends on the proper data needed to solve a particular problem. Data preprocessing, however, inflicts a heavy danger, e.g. during the preprocessing data can be inadvertently modified, e.g. "interesting" data may be removed. Consequently, for discovery purposes it would be wise to have a look at the original raw data first.

Data integration is a hot topic generally and in health informatics specifically and solutions can bridge the gap between clinical and biomedical research [39]. This is becoming even more important due to the increasing amounts of heterogeneous, complex patient related data sets, resulting from various sources including picture archiving and communication systems (PACS) and radiological information systems (RIS), hospital information systems (HIS), laboratory information systems (LIS), physiological and clinical data repositories, and all sorts of -omics data from laboratories, using samples from Biobanks. The latter include large collections of DNA sequence data, proteomic and metabolic data; resulting from sophisticated high-throughput analytical technologies. Along with classical patient records, containing large amounts of unstructured and semi-structured information, integration efforts incorporate enormous problems, but at the same time offers new possibilities for translational research. However, before starting any data integration or machine learning task, it is necessary to get a deep understanding of the underlying physics of the available data. In this paper we provide an overview about the modern data landscape in a clinical and biomedical research domain, with a focus on typical clinical/biomedical research, imaging and -omics data-sources, and the structure, quality and size of the produced patient related health information.

Whilst data integration is on combining data from different sources and providing users with a unified view on these data (e.g. combining research results from different bioinformatics repositories), *data fusion* is matching various data sets which represent one and the same object into a single, consistent, and clean representation [40]; in health informatics these unified views are particularly important in high-dimensions, e.g. for integrating heterogeneous descriptions of

the same set of genes [41]. The main expectation is that fused data is more informative than the original inputs.

Capturing all information describing a biological system is the implicit objective of all -omics methods, however, genomics, transcriptomics, proteomics, metabolomics, etc. need to be combined to approach this goal: valuable information can be obtained using various analytical techniques such as nuclear magnetic resonance, liquid chromatography, or gas chromatography coupled to mass spectrometry. Each method has inherent advantages and disadvantages, but are complementary in terms of biological information, consequently combining multiple data sets, provided by different analytical platforms is of utmost importance. For each platform, the relevant information is extracted in the first step. The obtained latent variables are then fused and further analyzed. The influence of the original variables is then calculated back and interpreted. There is plenty of open future research to include all possible sources of information [42].

3.2 Research Track 2 ML: Machine Learning Algorithms

There are uncountable future challenges in the design, development, experimentation and evaluation of ML algorithms generally and in the application to health informatics specifically. The ultimate goal ever since is to develop algorithms which can *automatically* learn from data, hence can improve with experience over time *without any human-in-the-loop.* Most colleagues from the ML community are concentrating on automatic Machine Learning (aML), with the grand goal of excluding humans, hence to make it fully automatic and best practice real-world examples can be found in speech processing [43], recommender systems [44], or autonomous vehicles [45], just to mention a few.

However, the application of such aML approaches in the complex health domain seems elusive in the near future and a good example are Gaussian processes, where aML approaches (e.g. standard kernel machines) struggle on function extrapolation problems which are trivial for human learners. Consequently, *interactive ML-approaches,* by integrating a *human-into-the-loop* (e.g. a human kernel [46]), thereby making use of human cognitive abilities, is a promising approach for solving problems in the complex health domain. iML can be defined as algorithms that can interact with *both computational agents and human agents* and can optimize their learning behaviour through these interactions [22]. In Active Learning such agents are referred to as oracles [47].

iML-approaches can be of particular interest to solve problems, where we are lacking big data sets, deal with complex data and/or rare events, where traditional learning algorithms suffer due to insufficient training samples. Here the doctor-in-the-loop can help, where human expertise and long-term experience can assist in solving problems which otherwise would remain NP-hard; examples include subspace clustering [48], protein folding [49], or privacy preserving ML, which is an important issue, fostered by anonymization, in which a record is released only if it is indistinguishable from k other entities in the data, but where k-anonymity is highly dependent on spatial locality in order to effectively implement the technique in a statistically robust way. In high dimensionalities data

becomes sparse, hence the concept of spatial locality is not easy to define. Consequently, it becomes difficult to anonymize the data without an unacceptably high amount of information loss [50] - here iML could be of help.

Despite these apparent findings, so far there is little quantitative evidence on effectiveness and efficiency of iML-algorithms. Moreover, there is practically no evidence, how such interaction may really optimize such algorithms. Even though such "natural" intelligent agents are present in large numbers on our world and are studied by cognitive scientists for quite a while [51]. One possible explanation for the dominance of aML-approaches could be, that these are much better to evaluate and therefore are more rapidly publishable. In iML approaches methodically correct evaluations are not only much more difficult and time-consuming, but also very difficult or even impossible to replicate, due to the fact that human agents are subjective, individual and therefore can not be copied - in contrast to data, algorithms and computational agents. Robustness of iML is an open question.

3.3 Research Track 3 GDM Graph-Based Data Mining

Graph-Theory [52] provides powerful tools to map data structures and to find novel connections between single data objects [53,54]. The inferred graphs can be further analyzed by using graph-theoretical, statistical and machine learning techniques [55]. A mapping of already existing and in medical practice approved *knowledge spaces* as a conceptual graph (as e.g. demonstrated in [56]) and a subsequent visual and graph-theoretical analysis can bring novel insights on hidden patterns in the data, which exactly is the goal of knowledge discovery. Another benefit of a graph-based data structure is in the applicability of methods from network topology and network analysis and data mining, e.g. small-world phenomenon [57,58], and cluster analysis [59,60]. However, the first question is "How to get a graph?", or simpler "How to get point sets?", because point cloud data sets (PCD) are used as primitives for such approaches. The answer to this question is not trivial (see [61]), apart from "naturally available" point clouds, e.g. from laser scanners, protein structures [62], or text mapped into a set of points (vectors) in $\mathbb{R}^n$. Sticking on the last example, graphs are intuitively more informative as example words/phrase representations [63], and graphs are the best studied data structures in computer science, with a strong relation to logical languages [64]. The beginning of graph-based data mining approaches was two decades ago, some pioneering work include [65–67]. According to [64] there are five theoretical bases of graph-based data mining approaches such as (1) subgraph categories, (2) subgraph isomorphism, (3) graph invariants, (4) mining measures and (5) solution methods. Furthermore, there are five groups of different graph-theoretical approaches for data mining such as (1) greedy search based approach, (2) inductive logic programming based approach, (3) inductive database based approach, (4) mathematical graph theory based approach and (5) kernel function based approach [68]. However, the main disadvantage of graph-theoretical text mining is the computational complexity of the graph representation, consequently the goal of future research in the field of graph-theoretical

approaches for text mining is to develop efficient graph mining algorithms which implement effective search strategies and data structures [63].

In [69] a graph-theoretical approach for text mining is used to extract relation information between terms in "free-text" electronic health care records that are semantically or syntactically related. Another field of application is the text analysis of web and social media for detecting influenza-like illnesses [70].

Moreover there can be content-rich relationship networks among biological concepts, genes, proteins and drugs developed with topological text data mining like shown in [71]. According to [72] network medicine describes the clinical application field of topological text mining due to addressing the complexity of human diseases with molecular and phenotypic network maps.

3.4 Research Track 4 TDM Topological Data Mining

Closely related to graph-based methods are topological data mining methods; for both we need point cloud data sets - or at least distances - as input. A set of such primitives forms a space, and if we have finite sets equipped with proximity or similarity measure functions $sim_q \colon S^{q+1} \to [0,1]$, which measure how "close" or "similar" $(q+1)$-tuples of elements of S are, we speak about a *topological space*. A value of 0 means totally different objects, while 1 corresponds to equivalent items. Interesting are manifolds, which can be seen as a topological space, which is locally homeomorphic (that means it has a continuous function with an inverse function) to a real n-dimensional space. In other words: X is a d-manifold if every point of X has a neighborhood homeomorphic to $\mathbb{B}^d$; with boundary if every point has a neighborhood homeomorphic to $\mathbb{B}$ or $\mathbb{B}^d_+$ [73].

A topological space may be viewed as an abstraction of a metric space, and similarly, manifolds generalize the connectivity of d-dimensional Euclidean spaces $\mathbb{B}^d$ by being locally similar, but globally different. A d-dimensional chart at $p \in X$ is a homeomorphism $\phi : U \to \mathbb{R}^d$ onto an open subset of $\mathbb{R}^d$, where U is a neighborhood of p and open is defined using the metric. A d-dimensional manifold (d-manifold) is a topological space X with a d-dimensional chart at every point $x \in X$ [74].

For us also interesting are simplicial complexes ("simplicials") which are spaces described in a very particular way, the basis is in Homology. The reason is that it is not possible to represent surfaces precisely in a computer system due to limited computational storage; thus, surfaces are sampled and represented with triangulations. Such a triangulation is called a simplicial complex, and is a combinatorial space that can represent a space. With such simplicial complexes, the topology of a space from its geometry can be separated. Zomorodian [74] compares it with the separation of syntax and semantics in logic.

The two most popular techniques are *homology* and *persistence*. The connectivity of a space is determined by its cycles of different dimensions. These cycles are organized into groups, called homology groups. Given a reasonably explicit description of a space, the homology groups can be computed with linear algebra. Homology groups have a relatively strong discriminative power and a clear meaning, while having low computational cost. In the study of persistent

homology the invariants are in the form of persistence diagrams or barcodes [75]. For us it is important to extract significant features, and thus these methods are useful, since they provide robust and general feature definitions with emphasis on global information, e.g. Alpha Shapes [76]. A recent example for topological data mining is given by [77]: Topological text mining, which builds on the well-known vector space model, which is a standard approach in text mining [78]: a collection of text documents (corpus) is mapped into points (=vectors) in $\mathbb{R}^n$. Moreover, each word can be mapped into so-called term vectors, resulting in a very high dimensional vector space. If there are n words extracted from all the documents then each document is mapped to a point (*term vector*) in $\mathbb{R}^n$ with coordinates corresponding to the weights. This way the whole corpus can be transformed into a point cloud data set. Instead of the Euclidean metric the use of a similarity (proximity) measure is sometimes more convenient; the *cosine similarity measure* is a typical example: the cosine of the angle between two vectors (points in the cloud) reflects how "similar" the underlying weighted combinations of keywords are. Amongst the many different text mining methods (for a recent overview refer to [79]); topological approaches are promising, but need a lot of further research. One of the main tasks of applied topology is to find and analyse higher dimensional topological structures in lower dimensional spaces (e.g. point cloud from vector space model as discussed in [80]). A common way to describe topological spaces is to first create simplicial complexes, because a simplicial complex structure on a topological space is an expression of the space as a union of simplices such as points, intervals, triangles, and higher dimensional analogues. Simplicial complexes provide an easy combinatorial way to define certain topological spaces [81]. A simplical complex K is defined as a finite collection of simplices such that $\sigma \in K$ and τ, which is a face of σ, implies $\tau \in K$, and $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ can either be a face of both σ and σ' or empty [82]. One way to create a simplical complex is to examine all subsets of points, and if any subsets of points are close enough, a p-simplex (e.g. line) is added to the complex with those points as vertices. For instance, a Vietoris-Rips complex of diameter ϵ is defined as $VR(\epsilon) = \sigma | diam(\sigma) \leq \epsilon$, where $diam(\epsilon)$ is defined as the largest distance between two points in σ [82]. A common way a analyse the topological structure is to use persistent homology, which identifies cluster, holes and voids therein. It is assumed that more robust topological structures are the one which persist with increasing ϵ. For detailed information about persistent homology, see [82–84].

3.5 Research Track 5 EDM Entropy-Based Data Mining

Information Entropy can be used as a measure of *uncertainty in data*. To date, there have emerged many different types of entropy methods with a large number of different purposes and applications; here we mention only a few: *Graph Entropy* was described by [85] to measure structural information content of graphs, and a different definition, more focused on problems in information and coding theory, was introduced by Körner in [86]. Graph entropy is often used for

the characterization of the structure of graph-based systems, e.g. in mathematical biochemistry, but also for any complex network [87]. In these applications the entropy of a graph is interpreted as its structural information content and serves as a complexity measure, and such a measure is associated with an equivalence relation defined on a finite graph; by application of Shannons Eq. 2.4 in [88] with the probability distribution we get a numerical value that serves as an index of the structural feature captured by the equivalence relation.

Topological Entropy (TopEn), was introduced by [89] with the purpose to introduce the notion of entropy as an invariant for continuous mappings: Let (X, T) be a topological dynamical system, i.e., let X be a nonempty compact Hausdorff space and $T : X \to X$ a continuous map; the TopEn is a nonnegative number which measures the complexity of the system [90].

Hornero et al. [91] performed a complexity analysis of intracranial pressure dynamics during periods of severe intracranial hypertension. For that purpose they analyzed eleven episodes of intracranial hypertension from seven patients. They measured the changes in the intracranial pressure complexity by applying ApEn, as patients progressed from a state of normal intracranial pressure to intracranial hypertension, and found that a decreased complexity of intracranial pressure coincides with periods of intracranial hypertension in brain injury. Their approach is of particular interest to us, because they proposed classification based on ApEn tendencies instead of absolute values.

Pincus et al. took in [92] heart rate recordings of 45 healthy infants with recordings of an infant one week after an aborted sudden infant death syndrom (SIDS) episode. They then calculated the ApEn of these recordings and found a significant smaller value for the aborted SIDS infant compared to the healthy ones.

3.6 Research Track 6 DAV Data Visualization

Visualization is a very important method of transforming the symbolic into the geometric, offers opportunities for discovering knowledge in data and fosters insight into data [26]. There are endless examples for the importance of visualization in health, e.g. Otasek et al. [12] present a work on Visual Data Mining (VDM), which is supported by interactive and scalable network visualization and analysis. Otasek et al. emphasize that knowledge discovery within complex data sets involves many workflows, including accurately representing many formats of source data, merging heterogeneous and distributed data sources, complex database searching, integrating results from multiple computational and mathematical analyses, and effectively visualizing properties and results. Mueller et al. [93] demonstrate the successful application of data Glyphs in a disease analyser for the analysis of big medical data sets with automatic validation of the data mapping, selection of subgroups within histograms and a visual comparison of the value distributions. A good example for the catenation of visualization with ML is clustering: Clustering is a descriptive task to identify homogeneous groups of data objects based on the dimensions (i.e. values of the attributes). Clustering

methods are often subject to other systems, for example to reduce the possibilities of recommender systems (e.g. Tag-recommender on Youtube videos [94]); for example clustering of large high-dimensional gene expression data sets has widespread application in -omics [95]. Unfortunately, the underlying structure of these natural data sets is often fuzzy, and the computational identification of data clusters generally requires (human) expert knowledge about cluster number and geometry. The high-dimensionality of data is a huge problem in health informatics general and in ML in particular, and the curse of dimensionality is a critical factor for clustering: With increasing dimensionality the volume of the space increases so fast that the available data becomes sparse, hence it becomes impossible to find reliable clusters; also the concept of distance becomes less precise as the number of dimensions grows, since the distance between any two points in a given data set converges; moreover, different clusters might be found in different sub spaces, so a global filtering of attributes is also not sufficient. Given that large number of attributes, it is likely that some attributes are correlated, therefore clusters might exist in arbitrarily oriented affinity sub spaces. Moreover, high-dimensional data likely includes *irrelevant* features, which may obscure to find the relevant ones, thus increases the danger of modeling artifacts. The problem is that we are confronted with subjective similarity functions; the most simplest example is the grouping of cars in a showroom: a technician will most likely group the cars differently than a mother of three kids (cylinder capacity versus storage capacity). This subspace clustering problem is hard, because for the grouping very different characteristics can be used: highly subjective and context specific. What is recognized as comfort for end-users of individual systems, can be applied in scientific research for the interactive exploration of high-dimensional data sets [96]. Consequently, iML-approaches can be beneficial to support finding solutions in hard biomedical problems [48]. Actually, humans are quite good in comparison for the determination of similarities and dissimilarities - described by nonlinear multidimensional scaling (MDS) models [97]. MDS models represent similarity relations between entities as a geometric model that consists of a set of points within a metric space. The output of an MDS routine is a geometric model of the data, with each object of the data set represented as a point in n-dimensional space.

3.7 Research Track 7 DAP Privacy

Privacy aware machine learning and privacy preserving machine learning is an important issue [98,99], fostered by anonymization concepts, in which a record is released only if it is indistinguishable from k other entities in the data. k-anonymity is highly dependent on spatial locality in order to effectively implement the technique in a statistically robust way and in high dimensions data becomes sparse, hence the concept of spatial locality is not easy to define. Consequently, it becomes difficult to anonymize the data without an unacceptably high amount of information loss [50]. Consequently, the problem of k-Anonymization is on the one hand NP-hard, on the other hand the quality of the result obtained

can be measured at the given factors (k-Anonymity, l-diversity, t-closeness, delta-presence), but not with regard to the actual security of the data, i.e. the re-identification through an attacker. For this purpose certain assumptions about the background knowledge of the hypothetical enemy must be made. With regard to the particular demographic and cultural clinical environment this is best done by a human agent. Thus, the problem of (k-)Anonymization represents a natural application domain for iML.

4 Example Curriculum

Most universities offer excellent courses on machine learning, neural networks, data mining, and visualization, so a course on ML for HI should be complementary and follow a research-based teaching (RBT) style, showing the students state-of-the-art science and engineering example from biomedicine and the life sciences for discussing the underlying concepts, theories, paradigms, models, methods and tools on practical cases and examples (Fig. 2). For practical reasons the exercises can be done with Python [100], which is to date still the

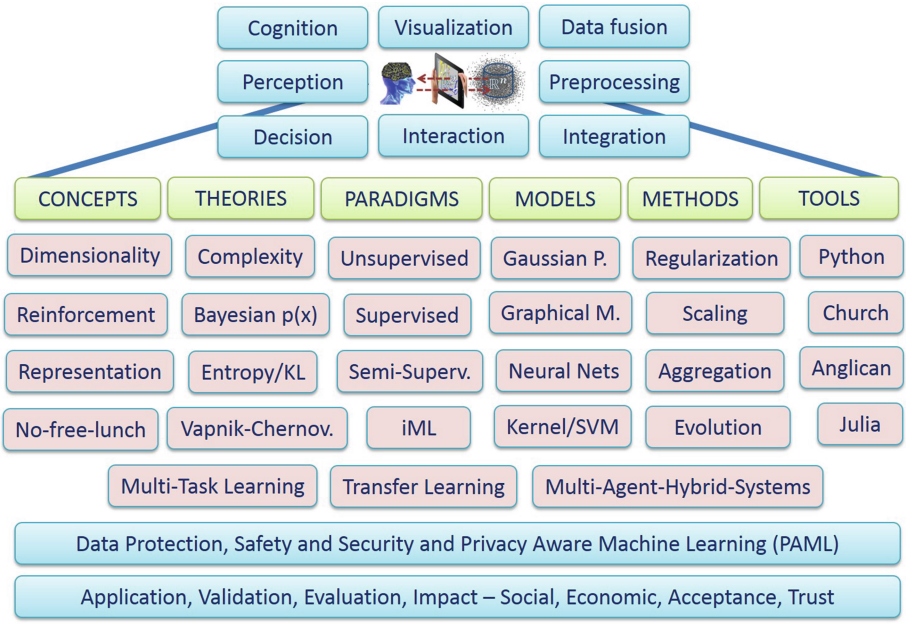

Fig. 2. The top level view of the contents of the Machine Learning for Health Informatics course at Vienna University of Technology, developed by A. Holzinger. Besides from focusing on practical examples from biology, biomedicine, clinical medicine and healthcare, issues including privacy, safety, security, data protection, validation, evaluation, social and economic impact, acceptance and trust are important parts of this course

most used ML-tool worldwide, and probabilistic programming [101] should be fostered (with at least a short touch on, e.g., Anglican, Church, or PyMC). The course 183.A83 at Vienna University of Technology (http://hci-kdd.org/machine-learning-for-health-informatics-course/) is consisting of twelve lectures plus practicals for a one-semester course on Master level with the following contents:

Lecture 01: Introduction and Overview of ML and HI explains the HCI-KDD approach, shows the complexity of the application area health informatics, demonstrates what aML can do and shows the limitations of aML, and the usefulness iML with a human-in-the-loop on practical examples and outlines some future challenges.

Lecture 02: Fundamentals of Data and Information discusses the underlying physics of data and biomedical data sources, taxonomy of data, data structures, data integration, data fusion, and a clinical view on data, information and knowledge; focuses then on probabilistic information, information theory, cross-entropy, mutual information and Kullback-Leibler Divergence.

Lecture 03: Dimensionality Reduction and Subspace Clustering provides an introduction into classification vs. clustering, feature spaces, feature engineering, discusses the curse of dimensionality and methods of dimensionality reduction, and demonstrates the usefulness of subspace clustering with the expert-in-the-loop; finally discusses the hard question "what is interesting?" by showing projection pursuit.

Lecture 04: Human Learning vs. Machine Learning: Decision Making starts with reinforcement learning and discusses the differences of humans and machines on the example of decision making under uncertainty, shows then multi-armed bandits and applications in health and finally gives an outlook on the importance of transfer learning.

Lecture 05: Probabilistic Graphical Models I starts with reasoning under uncertainty and expected utility theory, highlights the importance of graphs and knowledge representation in network medicine, shows some basic metrics and measures and discusses practical examples of graphical model learning and how to get graphs.

Lecture 06: Probabilistic Graphical Models II continues with graphical models and decision making, shows factor graphs, graph isomorphism and applications, Bayes nets, ML on graphs, similarity and correspondence, and probabilistic topic models for natural language to get insight into unknown document collections, concluded by Graph bandits.

Lecture 07: Evolutionary Computing for HI I poses medical decision making as search problem and shows evolutionary principles (Lamarck, Darwin, Baldwin, Mendel) and applications of evolutionary computing with the special case of genetic algorithms and k-armed bandits and genetic algorithms (global optimization problem).

Lecture 08: Evolutionary Computing for HI II continues with examples from medical applications for EA, discusses natural computing concepts and their usefulness in principle, focuses then on Ant Colony Optimization and the traveling salesman problem with motivation on protein folding, simulated annealing, and the human-in-the-loop, and finalizes with multi-agents and neuro evolution.

Lecture 09: Towards Open Data Sets: Privacy Aware Machine Learning motivates privacy, data protection safety and security and discusses anonymization methods (k-Anonymization, l-diversity, t-closeness, delta-presence, pertubative approaches, differentially private kernel learning, etc.), and how iML can help anonymization.

Lecture 10: Active Learning, Multi-Task Learning and Transfer Learning discusses the principles of active learning, preference learning, active preference learning with an excursion on PAC-learning, and programming by feedback, highlights some problems of the human-in-the-loop and continues with MTL and TL, where humans are still better than machines.

Lecture 11: Machine Learning from Text focuses on natural language understanding and the problems involved, and highlights word vectors for sentiment analysis (continous bag-of-words model, skip-gram model, global vectors for word embedding) with giving an outline on neural probabilistic language models and alternative models.

Lecture 12: Discrete Multi-Agent Systems on the topic of stochastic simulation of tumor kinetics and key problems for cancer research, tumor growth modeling, cellular potts model, tumor growth visualization and towards using open tumor growth data for machine learning in the international context [102].

5 Future Challenges

Much future research has to be done, particularly in the fields of Multi-Task Learning and Transfer Learning to go towards Multi-Agent-Hybrid Systems as applications of the iML-approach.

5.1 Future Challenge 1: Multi-task Learning

Multi-task learning (MTL) aims to improve the prediction performance by learning a problem together with multiple, different but related other problems through shared parameters or a shared representation. The underlying principle is *bias learning* based on probable approximately correct learning (PAC learning) [103]. To find such a bias is still the hardest problem in any ML task and essential for the initial choice of an appropriate hypothesis space, which must be large enough to contain a solution, and small enough to ensure a good generalization from a small number of data sets. Existing methods of bias generally require the input of a human-expert-in-the-loop in the form of heuristics and domain knowledge to ensure the selection of an appropriate set of features, as such features

are key to learning and understanding. However, such methods are limited by the accuracy and reliability of the expert s knowledge (robustness of the human) and also by the extent to which that knowledge can be transferred to new tasks (see next subsection). Baxter (2000) [104] introduced a model of bias learning which builds on the PAC learning model which concludes that learning multiple related tasks reduces the sampling burden required for good generalization and bias that is learnt on sufficiently many training tasks is likely to be good for learning novel tasks drawn from the same environment (the problem of transfer learning to new environments is discussed in the next subsection). A practical example is *regularized MTL* [105], which is based on the minimization of regularization functionals similar to Support Vector Machines (SVMs), that have been successfully used in the past for singletask learning. The regularized MTL approach allows to model the relation between tasks in terms of a novel kernel function that uses a taskcoupling parameter and largely outperforms singletask learning using SVMs. However, multi-task SVMs are inherently restricted by the fact that SVMs require each class to be addressed explicitly with its own weight vector. In a multi-task setting this requires the different learning tasks to share the *same set of classes.* An alternative formulation for MTL is an extension of the large margin nearest neighbor algorithm (LMNN) [106]. Instead of relying on separating hyper-planes, its decision function is based on the nearest neighbor rule which inherently extends to many classes and becomes a natural fit for MTL. This approach outperforms state-of-the-art MTL classifiers, however, much open research challenges remain open in this area [107].

5.2 Future Challenge 2: Transfer Learning

A huge problem in ML is the phenomenon of *catastrophic forgetting*, i.e. when learned one task and transferred to another task the ML algorithm "forgets" how to perform the learned task. This is a well-known problem which affects ML-systems and was first described in the context of connectionist networks [108]; whereas natural cognitive systems rarely completely disrupt or erase previously learned information, i.e. natural cognitive systems do not forget "catastrophically" [109]. Consequently the challenge is to discover how to avoid the problem of catastrophic forgetting, which is a current hot topic [110].

According to Pan & Yang (2010) [111] a major assumption in many ML algorithms is, that both the training data and future (unknown) data must be in the same feature space and required to have the same distribution. In many real-world applications, particularly in the health domain, this is not the case: Sometimes we have a classification task in one domain of interest, but we only have sufficient training data in another domain of interest, where the latter data may be in a completely different feature space or follows a different data distribution. In such cases transfer learning would greatly improve the performance of learning by avoiding much expensive data-labeling efforts, however, much open questions remain for future research [112].

5.3 Future Challenge 3: Multi-agent-Hybrid Systems

Multi-Agent-Systems (MAS) are collections of many agents interacting with each other. They can either share a common goal (for example an ant colony, bird flock, or fish swarm etc.), or they can pursue their own interests (for example as in an open-market economy). MAS can be traditionally characterized by the facts that (a) each agent has incomplete information and/or capabilities for solving a problem, (b) agents are autonomous, so there is no global system control; (c) data is decentralized; and (d) computation is asynchronous [113]. For the health domain of particular interest is the *consensus problem*, which formed the foundation for distributed computing [114]. The roots are in the study of (human) experts in group consensus problems: Consider a group of humans who must act together as a team and each individual has a subjective probability distribution for the unknown value of some parameter; a model which describes how the group reaches agreement by pooling their individual opinions was described by DeGroot (1974) [115] and was used decades later for the aggregation of information with uncertainty obtained from multiple sensors [116] and medical experts [117]. On this basis Olfati-Saber et al. (2007) [118] presented a theoretical framework for analysis of consensus algorithms for networked multi-agent systems with fixed or dynamic topology and directed information flow. In complex real-world problems, e.g., for the epidemiological and ecological analysis of infectious diseases, standard models based on differential equations very rapidly become unmanageable due to too many parameters, and here MAS can also be very helpful [119]. Moreover, collaborative multi-agent reinforcement learning has a lot of research potential for machine learning [120].

6 Conclusion

There are uncountable future challenges in ML generally and in the application of ML to health informatics specifically. The ultimate goal is to design and develop algorithms which can *automatically* learn from data, hence can improve with experience over time *without any human-in-the-loop.* However, the application of such aML approaches in the complex health domain seems elusive in the near future and a good example are Gaussian processes, where aML approaches (e.g. standard kernel machines) struggle on function extrapolation problems which are trivial for human learners. Consequently, iML-approaches, by integrating a human-into-the-loop (e.g. a human kernel [46]), thereby making use of human cognitive abilities, seems to be a promising approach. iML-approaches can be of particular interest to solve problems in HI, where we are lacking big data sets, deal with complex data and/or rare events, where traditional learning algorithms suffer due to insufficient training samples. Here the doctor-in-the-loopcan help, where human expertise and long-term experience can assist in solving problems which otherwise would remain NP-hard. A cross-domain integration and appraisal of different fields provides an atmosphere to foster different perspectives and opinions and is an ideal think-tank and incubator to foster novel ideas and a fresh look on different methodologies to put these ideas into Business.

Acknowledgments. I am very grateful for fruitful discussions with members of the HCI-KDD network and I thank my Institutes both at Graz University of Technology and the Medical University of Graz, my colleagues and my students for the enjoyable academic freedom, the inspiring intellectual environment, and the opportunity to follow my personal motto: Science is to test crazy ideas - Engineering is to put these ideas into Business. Last but not least, I thank all students of my course LV 185.A83 (http://hci-kdd.org/machine-learning-for-health-informatics-course), at Vienna University of Technology for their kind interest and motivating feedback.

References

1. Samuel, A.L.: Some studies in machine learning using the game of checkers. IBM J. Res. Dev. **3**, 210–229 (1959)
2. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**, 988–999 (1999)
3. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory COLT, pp. 144–152. ACM (1992)
4. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer, New York (2009)
5. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science **349**, 255–260 (2015)
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
7. Holzinger, A., Dehmer, M., Jurisica, I.: Knowledge discovery and interactive data mining in bioinformatics - state-of-the-art, future challenges and research directions. BMC Bioinform. **15**, I1 (2014)
8. Holzinger, A.: Trends in interactive knowledge discovery for personalized medicine: cognitive science meets machine learning. IEEE Intell. Inform. Bull. **15**, 6–14 (2014)
9. Su, X., Kang, J., Fan, J., Levine, R.A., Yan, X.: Facilitating score and causal inference trees for large observational studies. J. Mach. Learn. Res. **13**, 2955–2994 (2012)
10. Huppertz, B., Holzinger, A.: Biobanks a source of large biological data sets: open problems and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 317–330. Springer, Heidelberg (2014)
11. Mattmann, C.A.: Computing: a vision for data science. Nature **493**, 473–475 (2013)
12. Otasek, D., Pastrello, C., Holzinger, A., Jurisica, I.: Visual data mining: effective exploration of the biological universe. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 19–33. Springer, Heidelberg (2014)
13. Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. Nature **521**, 452–459 (2015)
14. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical Dirichlet processes. J. Am. Stat. Assoc. **101**, 1566–1581 (2006)
15. Houlsby, N., Huszar, F., Ghahramani, Z., Hernndez-lobato, J.M.: Collaborative gaussian processes for preference learning. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems (NIPS 2012), pp. 2096–2104 (2012)
16. Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N.: Taking the human out of the loop: a review of bayesian optimization. Proc. IEEE **104**, 148–175 (2016)

17. Clark, A.: Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. **36**, 181–204 (2013)
18. Lee, S., Holzinger, A.: Knowledge discovery from complex high dimensional data. In: Michaelis, S., Piatkowski, N., Stolpe, M. (eds.) Solving Large Scale Learning Tasks. Challenges and Algorithms. LNCS (LNAI), vol. 9580, pp. 148–167. Springer, Heidelberg (2016). doi:10.1007/978-3-319-41706-6_7
19. Mayer, C., Bachler, M., Holzinger, A., Stein, P., Wassertheurer, S.: The effect of threshold values and weighting factors on the association between entropy measures and mortality after myocardial infarction in the cardiac arrhythmia suppression trial. Entropy **18**, 1–15 (2016)
20. Jadad, A.R., OGrady, L.: How should health be defined? Br. Med. J. **337**, a2900 (2008)
21. Parry, D.: Health informatics. In: Kasabov, N. (ed.) Springer Handbook of Bio-/Neuro-informatics, pp. 555–564. Springer, Heidelberg (2014)
22. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. (BRIN) **3**, 119–131 (2016)
23. Holzinger, A., Plass, M., Holzinger, K., Crişan, G.C., Pintea, C.-M., Palade, V.: Towards interactive Machine Learning (iML): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-ARES 2016. LNCS, vol. 9817, pp. 81–95. Springer, Heidelberg (2016). doi:10.1007/978-3-319-45507-5_6
24. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. AI Mag. **17**, 37–54 (1996)
25. Holzinger, A.: On topological data mining. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 331–356. Springer, Heidelberg (2014)
26. Ward, M., Grinstein, G., Keim, D.: Interactive Data Visualization: Foundations, Techniques, and Applications. AK Peters Ltd., Natick (2010)
27. Holzinger, A.: On knowledge discovery and interactive intelligent visualization of biomedical data - challenges in human computer interaction & biomedical informatics. In: Helfert, M., Fancalanci, C., Filipe, J. (eds.) DATA 2012, International Conference on Data Technologies and Applications, pp. 5–16 (2012)
28. Holzinger, A.: Human-computer interaction and knowledge discovery (HCI-KDD): what is the benefit of bringing those two fields to work together? In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds.) CD-ARES 2013. LNCS, vol. 8127, pp. 319–328. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40511-2_22
29. Holzinger, A., Jurisica, I.: Knowledge discovery and data mining in biomedical informatics: the future is in integrative, interactive machine learning solutions. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 1–18. Springer, Heidelberg (2014)
30. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518**, 529–533 (2015)
31. Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D.: How to grow a mind: statistics, structure, and abstraction. Science **331**, 1279–1285 (2011)
32. Burges, C.J.: Dimension reduction: a guided tour. Found. Trends Mach. Learn. **2**, 275–365 (2010)

33. Kandel, E.R., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., Hudspeth, A.: Principles of Neural Science, 5th edn. McGraw-Hill, New York (2012). (1760 pages)
34. McDermott, J.E., Wang, J., Mitchell, H., Webb-Robertson, B.J., Hafen, R., Ramey, J., Rodland, K.D.: Challenges in biomarker discovery: combining expert insights with statistical analysis of complex omics data. Expert Opin. Med. Diagn. **7**, 37–51 (2013)
35. Swan, A.L., Mobasheri, A., Allaway, D., Liddell, S., Bacardit, J.: Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. Omics- J. Integr. Biol. **17**, 595–610 (2013)
36. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Report, May 2011 (available online)
37. Goolsby, A.W., Olsen, L., McGinnis, M., Grossmann, C.: Clincial data as the basic staple of health learning - creating and protecting a public good. National Institute of Health (2010)
38. Holzinger, A.: Lecture 2 fundamentals of data, information, and knowledge. In: Biomedical Informatics: Discovering Knowledge in Big Data, pp. 57–107. Springer, Cham (2014)
39. Jeanquartier, F., Jean-Quartier, C., Schreck, T., Cemernek, D., Holzinger, A.: Integrating open data on cancer in support to tumor growth analysis. In: Renda, M.E., Bursa, M., Holzinger, A., Khuri, S. (eds.) ITBAM 2016. LNCS, vol. 9832, pp. 49–66. Springer, Heidelberg (2016). doi:10.1007/978-3-319-43949-5_4
40. Bleiholder, J., Naumann, F.: Data fusion. ACM Comput. Surv. (CSUR) **41**, 1–41 (2008)
41. Lafon, S., Keller, Y., Coifman, R.R.: Data fusion and multicue data matching by diffusion maps. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1784–1797 (2006)
42. Blanchet, L., Smolinska, A.: Data fusion in metabolomics and proteomics for biomarker discovery. In: Jung, K. (ed.) Statistical Analysis in Proteomics, pp. 209–223. Springer, New York (2016)
43. Bellegarda, J.R., Monz, C.: State of the art in statistical methods for language and speech processing. Comput. Speech Lang. **35**, 163–184 (2016)
44. Ricci, F., Rokach, L., Shapira, B.: Recommender systems: introduction and challenges. In: Ricci, F., Rokach, L., Shapira, B. (eds.) Recommender Systems Handbook, pp. 1–34. Springer, New York (2015)
45. Spinrad, N.: Google car takes the test. Nature **514**, 528 (2014)
46. Wilson, A.G., Dann, C., Lucas, C.G., Xing, E.P.: The human kernel. arXiv preprint arXiv:1510.07389 (2015)
47. Settles, B.: From theories to queries: active learning in practice. In: Guyon, I., Cawley, G., Dror, G., Lemaire, V., Statnikov, A. (eds.) Active Learning and Experimental Design Workshop 2010, vol. 16, pp. 1–18. JMLR Proceedings, Sardinia (2011)
48. Hund, M., Sturm, W., Schreck, T., Ullrich, T., Keim, D., Majnaric, L., Holzinger, A.: Analysis of patient groups and immunization results based on subspace clustering. In: Guo, Y., Friston, K., Aldo, F., Hill, S., Peng, H. (eds.) BIH 2015. LNCS (LNAI), vol. 9250, pp. 358–368. Springer, Heidelberg (2015). doi:10.1007/978-3-319-23344-4_35
49. Lathrop, R.H.: The protein threading problem with sequence amino-acid interaction preferences is np-complete. Protein Eng. **7**, 1059–1068 (1994)
50. Aggarwal, C.C.: On k-anonymity and the curse of dimensionality. In: Proceedings of the 31st International Conference on Very Large Data Bases VLDB, pp. 901–909 (2005)

51. Gigerenzer, G., Gaissmaier, W.: Heuristic decision making. Annu. Rev. Psychol. **62**, 451–482 (2011)
52. Harary, F.: Structural Models. An Introduction to the Theory of Directed Graphs. Wiley, New York (1965)
53. Strogatz, S.: Exploring complex networks. Nature **410**, 268–276 (2001)
54. Dorogovtsev, S., Mendes, J.: Evolution of Networks: From Biological Nets to the Internet and WWW. Oxford University Press, New York (2003)
55. Dehmer, M., Emmert-Streib, F., Pickl, S., Holzinger, A. (eds.): Big Data of Complex Networks. CRC Press Taylor & Francis Group, Boca Raton, London, New York (2016)
56. Holzinger, A., Ofner, B., Dehmer, M.: Multi-touch graph-based interaction for knowledge discovery on mobile devices: state-of-the-art and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 241–254. Springer, Heidelberg (2014)
57. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
58. Kleinberg, J.: Navigation in a small world. Nature **406**, 845 (2000)
59. Koontz, W., Narendra, P., Fukunaga, K.: A graph-theoretic approach to nonparametric cluster analysis. IEEE Trans. Comput. **100**, 936–944 (1976)
60. Wittkop, T., Emig, D., Truss, A., Albrecht, M., Boecker, S., Baumbach, J.: Comprehensive cluster analysis with transitivity clustering. Nat. Protoc. **6**, 285–295 (2011)
61. Holzinger, A., Malle, B., Bloice, M., Wiltgen, M., Ferri, M., Stanganelli, I., Hofmann-Wellenhof, R.: On the generation of point cloud data sets: the first step in the knowledge discovery process. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 57–80. Springer, Heidelberg (2014)
62. Canutescu, A.A., Shelenkov, A.A., Dunbrack, R.L.: A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. **12**, 2001–2014 (2003)
63. Jiang, C., Coenen, F., Sanderson, R., Zito, M.: Text classification using graph mining-based feature extraction. Knowl. Based Syst. **23**, 302–308 (2010)
64. Washio, T., Motoda, H.: State of the art of graph-based data mining. ACM SIGKDD Explor. Newsl. **5**, 59 (2003)
65. Cook, D.J., Holder, L.B.: Substructure discovery using minimum description length and background knowledge. J. Artif. Int. Res. **1**, 231–255 (1994)
66. Yoshida, K., Motoda, H., Indurkhya, N.: Graph-based induction as a unified learning framework. Appl. Intell. **4**, 297–316 (1994)
67. Dehaspe, L., Toivonen, H.: Discovery of frequent DATALOG patterns. Data Min. Knowl. Discov. **3**, 7–36 (1999)
68. Windridge, D., Bober, M.: A kernel-based framework for medical big-data analytics. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 196–207. Springer, Heidelberg (2014)
69. Zhou, X., Han, H., Chankai, I., Prestrud, A., Brooks, A.: Approaches to text mining for clinical medical records. In: Proceedings of the 2006 ACM Symposium on Applied Computing - SAC 2006, New York, USA, p. 235. ACM, New York (2006)
70. Corley, C.D., Cook, D.J., Mikler, A.R., Singh, K.P.: Text and structural data mining of influenza mentions in Web and social media. Int. J. Environ. Res. Public Health **7**, 596–615 (2010)
71. Chen, H., Sharp, B.M.: Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinform. **5**, 147 (2004)

72. Barabási, A., Gulbahce, N., Loscalzo, J.: Network medicine: a network-based approach to human disease. Nat. Rev. Genet. **12**, 56–68 (2011)
73. Cannon, J.W.: The recognition problem: what is a topological manifold? Bull. Am. Math. Soc. **84**, 832–866 (1978)
74. Zomorodian, A.: Computational topology. In: Atallah, M., Blanton, M. (eds.) Algorithms and Theory of Computation Handbook. Applied Algorithms and Data Structures Series, vol. 2, 2nd edn, pp. 1–31. Chapman and Hall/CRC, Boca Raton (2010). doi:10.1201/9781584888215-c3
75. Epstein, C., Carlsson, G., Edelsbrunner, H.: Topological data analysis. Inverse Prob. **27**, 120–201 (2011)
76. Edelsbrunner, H., Mucke, E.P.: 3-dimensional alpha-shapes. ACM Trans. Graph. **13**, 43–72 (1994)
77. Wagner, H., Dlotko, P.: Towards topological analysis of high-dimensional feature spaces. Comput. Vis. Image Underst. **121**, 21–26 (2014)
78. Kobayashi, M., Aono, M.: Vector space models for search and cluster mining. In: Berry, M.W. (ed.) Survey of Text Mining: Clustering, Classification, and Retrieval, pp. 103–122. Springer, New York (2004)
79. Holzinger, A., Schantl, J., Schroettner, M., Seifert, C., Verspoor, K.: Biomedical text mining: open problems and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Knowledge Discovery and Data Mining. LNCS, vol. 8401, pp. 271–300. Springer, Heidelberg (2014)
80. Wagner, H., Dłotko, P., Mrozek, M.: Computational topology in text mining. In: Ferri, M., Frosini, P., Landi, C., Cerri, A., Fabio, B. (eds.) CTIC 2012. LNCS, vol. 7309, pp. 68–78. Springer, Heidelberg (2012). doi:10.1007/978-3-642-30238-1_8
81. Carlsson, G.: Topology and data. Bull. Am. Math. Soc. **46**, 255–308 (2009)
82. Zhu, X.: Persistent homology: an introduction and a new text representation for natural language processing. In: Rossi, F. (ed.) IJCAI. IJCAI/AAAI (2013)
83. Cerri, A., Fabio, B.D., Ferri, M., Frosini, P., Landi, C.: Betti numbers in multidimensional persistent homology are stable functions. Math. Methods Appl. Sci. **36**, 1543–1557 (2013)
84. Bubenik, P., Kim, P.T.: A statistical approach to persistent homology. Homology Homotopy Appl. **9**, 337–362 (2007)
85. Mowshowitz, A.: Entropy and the complexity of graphs: I. An index of the relative complexity of a graph. Bull. Math. Biophys. **30**, 175–204 (1968)
86. Körner, J.: Coding of an information source having ambiguous alphabet and the entropy of graphs. In: 6th Prague Conference on Information Theory, pp. 411–425 (1973)
87. Holzinger, A., Ofner, B., Stocker, C., Calero Valdez, A., Schaar, A.K., Ziefle, M., Dehmer, M.: On graph entropy measures for knowledge discovery from publication network data. In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds.) CD-ARES 2013. LNCS, vol. 8127, pp. 354–362. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40511-2_25
88. Dehmer, M.: Information theory of networks. Symmetry **3**, 767–779 (2011)
89. Adler, R.L., Konheim, A.G., McAndrew, M.H.: Topological entropy. Trans. Am. Math. Soc. **114**, 309–319 (1965)
90. Adler, R., Downarowicz, T., Misiurewicz, M.: Topological entropy. Scholarpedia **3**, 2200 (2008)
91. Hornero, R., Aboy, M., Abasolo, D., McNames, J., Wakeland, W., Goldstein, B.: Complex analysis of intracranial hypertension using approximate entropy. Crit. Care Med. **34**, 87–95 (2006)

92. Pincus, S.M.: Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. **88**, 2297–2301 (1991)
93. Mueller, H., Reihs, R., Zatloukal, K., Holzinger, A.: Analysis of biomedical data with multilevel glyphs. BMC Bioinform. **15**, S5 (2014)
94. Toderici, G., Aradhye, H., Paca, M., Sbaiz, L., Yagnik, J.: Finding meaning on youtube: tag recommendation and category discovery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010), pp. 3447–3454. IEEE (2010)
95. Sturm, W., Schreck, T., Holzinger, A., Ullrich, T.: Discovering medical knowledge using visual analytics a survey on methods for systems biology and omics data. In: Bühler, K., Linsen, L., John, N.W. (eds.) Eurographics Workshop on Visual Computing for Biology and Medicine, Eurographics EG, pp. 71–81 (2015)
96. Müller, E., Assent, I., Krieger, R., Jansen, T., Seidl, T.: Morpheus: interactive exploration of subspace clustering. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD 2008, pp. 1089–1092. ACM (2008)
97. Shepard, R.N.: The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika **27**, 125–140 (1962)
98. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Privacy aware learning. J. ACM (JACM) **61**, 38 (2014)
99. Malle, B., Kieseberg, P., Weippl, E., Holzinger, A.: The right to be forgotten: towards machine learning on perturbed knowledge bases. In: Buccafurri, F., Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-ARES 2016. LNCS, vol. 9817, pp. 251–266. Springer, Heidelberg (2016). doi:10.1007/978-3-319-45507-5_17
100. Bloice, M.D., Holzinger, A.: A tutorial on machine learning and data science tools with python. In: Holzinger, A. (ed.) ML for Health Informatics. LNCS (LNAI), vol. 9605, pp. xx–yy. Springer, Heidelberg (2016)
101. Gordon, A.D., Henzinger, T.A., Nori, A.V., Rajamani, S.K.: Probabilistic programming. In: Proceedings of the on Future of Software Engineering, pp. 167–181. ACM (2014)
102. Jeanquartier, F., Jean-Quartier, C., Kotlyar, M., Tokar, T., Hauschild, A.C., Jurisica, I., Holzinger, A.: Machine learning for in silico modeling of tumor growth. In: Holzinger, A. (ed.) ML for Health Informatics. LNCS (LNAI), vol. 9605, pp. xx–yy. Springer, Heidelberg (2016)
103. Valiant, L.G.: A theory of the learnable. Commun. ACM **27**, 1134–1142 (1984)
104. Baxter, J.: A model of inductive bias learning. J. Artif. Intell. Res. **12**, 149–198 (2000)
105. Evgeniou, T., Pontil, M.: Regularized multi-task learning. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 109–117. ACM (2004)
106. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**, 207–244 (2009)
107. Parameswaran, S., Weinberger, K.Q.: Large margin multi-task metric learning. In: Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., Culotta, A. (eds.) Advances in Neural Information Processing Systems (NIPS 2010), vol. 23, pp. 1867–1875 (2010)
108. McCloskey, M., Cohen, N.J.: Catastrophic interference in connectionist networks: the sequential learning problem. In: Bower, G.H. (ed.) The Psychology of Learning and Motivation, vol. 24, pp. 109–164. Academic Press, San Diego (1989)
109. French, R.M.: Catastrophic forgetting in connectionist networks. Trends Cogn. Sci. **3**, 128–135 (1999)

110. Goodfellow, I.J., Mirza, M., Xiao, D., Courville, A., Bengio, Y.: An empirical investigation of catastrophic forgeting in gradient-based neural networks arXiv:1312.6211v3 (2015)
111. Pan, S.J., Yang, Q.A.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**, 1345–1359 (2010)
112. Taylor, M.E., Stone, P.: Transfer learning for reinforcement learning domains: a survey. J. Mach. Learn. Res. **10**, 1633–1685 (2009)
113. Sycara, K.P.: Multiagent systems. AI Mag. **19**, 79 (1998)
114. Lynch, N.A.: Distributed Algorithms. Morgan Kaufmann, San Francisco (1996)
115. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**, 118–121 (1974)
116. Benediktsson, J.A., Swain, P.H.: Consensus theoretic classification methods. IEEE Trans. Syst. Man Cybern. **22**, 688–704 (1992)
117. Weller, S.C., Mann, N.C.: Assessing rater performance without a gold standard using consensus theory. Med. Decis. Making **17**, 71–79 (1997)
118. Olfati-Saber, R., Fax, J.A., Murray, R.M.: Consensus and cooperation in networked multi-agent systems. Proc. IEEE **95**, 215–233 (2007)
119. Roche, B., Guegan, J.F., Bousquet, F.: Multi-agent systems in epidemiology: a first step for computational biology in the study of vector-borne disease transmission. BMC Bioinform. **9**, 435 (2008)
120. Kok, J.R., Vlassis, N.: Collaborative multiagent reinforcement learning by payoff propagation. J. Mach. Learn. Res. **7**, 1789–1828 (2006)