International Journal of Swarm Intelligence and Evolutionary Computation

OPEN ACCESS Freely available online

Editorial

Web Mining: Knowledge Discovery in Data Base

Gonapa Vasudha*

Department of Biotechnology, Andjra University, Visakhapatnam, Andhra Pradesh, India

EDITORIAL NOTE

Data mining is a process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

To most of us data mining goes something like this: tons of data is collected, then quant wizards work their arcane magic, and then they know all of this amazing stuff. But, how? And what types of things can they know? Here is the truth: despite the fact that the specific technical functioning of data mining algorithms is quite complex -- they are a black box unless you are a professional statistician or computer scientist -- the uses and capabilities of these approaches are, in fact, quite comprehensible and intuitive.

And these days, there's always more data. We gather far more of it then we can digest. Nearly every transaction or interaction leaves a data signature that someone somewhere is capturing and storing. This is, of course, true on the internet; but, ubiquitous computing and digitization has made it increasingly true about our lives away from our computers (do we still have those?). The sheer scale of this data has far exceeded human sense-making capabilities. At these scales patterns are often too subtle and relationships too complex or multi-dimensional to observe by simply looking at the data. Data mining is a means of automating part this process to detect interpretable patterns; it helps us see the forest without getting lost in the trees.

Data mining can unintentionally be misused, and can then produce results that appear to be significant; but which do not actually predict future behavior and cannot be reproduced on a new sample of data and bear little use. Often this results from investigating too many hypotheses and not performing proper statistical hypothesis testing. A simple version of this problem in machine learning is known as overfitting, but the same problem can arise at different phases of the process and thus a train/test split—when applicable at all—may not be sufficient to prevent this from happening.

Data mining requires data preparation which uncovers information or patterns which compromise confidentiality and privacy obligations. A common way for this to occur is through data aggregation. Data aggregation involves combining data together (possibly from various sources) in a way that facilitates analysis (but that also might make identification of private, individual-level data deducible or otherwise apparent).

For exchanging the extracted models—in particular for use in predictive analytics—the key standard is the Predictive Model Markup Language (PMML), which is an XML-based language developed by the Data Mining Group (DMG) and supported as exchange format by many data mining applications. As the name suggests, it only covers prediction models, a particular data mining task of high importance to business applications. However, extensions to cover (for example) subspace clustering have been proposed independently of the DMG.

Correspondence to: Gonapa Vasudha, Department of Biotechnology, Andjra University, Visakhapatnam, Andhra Pradesh, India, E-mail: gonapa.vasudha@gmail.com

Received date: November 28, 2020; **Accepted date:** December 21, 2020; **Published date:** December 28, 2020

Citation: Vasudha G (2020) Web Mining: Knowledge Discovery in Data Base. Int J Swarm Evol Comput 9:e194.