Ievgen Gomolskyi, Bogdan Kyryliuk
IASA, NTUU "KPI"

EVOLUTIONARY DATA MINING ALGORITHMS

Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. The main tasks that data mining is usually called upon to accomplish are Description, Estimation, Prediction, Classification, Clustering, Association.

Evolutionary Algorithms (EAs) are stochastic search algorithms inspired by the process of neo-Darwinian evolution. Evolutionary algorithms are controlled by several inputs, such as the size of the population, and the rates that control how often mutation and crossover are used. In general, there is no guarantee that the evolutionary algorithm will find the optimal solution to an arbitrary problem, but a careful manipulation of the inputs and choosing a representation that is adequate to the problem increase the chances of success. EAs are often used as optimization algorithms, and this is the role that they play in most data mining applications.

The **motivation** for applying EAs to data mining is that EAs are robust, adaptive search methods that perform a global search in the space of candidate solutions. In contrast, several more conventional data mining methods perform a local, greedy search in the space of candidate solutions. As a result of their global search, EAs tend to cope better with attribute interactions than greedy data mining methods. Hence, intuitively EAs can discover interesting knowledge that would be missed by a greedy method. Also, EAs are a very flexible algorithmic paradigm. This gives the data miner a considerable freedom in the design of the individual representation, the fitness function and the genetic operators.

Like any other data mining paradigm, EAs also have some **disadvantages**. One of them is that conventional genetic operators – such as conventional crossover and mutation operators – are "blind" search operators in the sense that they modify individuals (candidate solutions) in a way independent from the individual's fitness (quality). This characteristic of conventional genetic operators increases the generality of EAs, but intuitively tends to reduce their effectiveness in solving a specific kind of problem. Another disadvantage of EAs is that they are computationally slow, by comparison with greedy search methods.

Evolutionary algorithms can complement many existing **data mining algorithms**. They can extract and select features, train neural networks, find classification rules, and build decision trees. Evolutionary algorithms are particularly useful when the problems involve the optimization of functions that are not smooth and differentiable, or functions where the objective value changes over time, which can happen in data mining as more data becomes available or if sampling is used to reduce the computation time. While evolutionary algorithms enable us to solve some difficult problems, they come at a price, namely a need for high computational resources. However, with processors becoming faster and the increasing acceptance of parallel systems, we hope that this problem will be minimized in the future.

References:

1. http://www.wikipedia.org/
2. Goldberg, D. E. - Genetic algorithms in search, optimization, and machine learning. 1989
3. Hornick, Mark F. - Java Data Mining: Strategy, Standard, and Practice. 2006
4. http://www.kdnuggets.com/

Pavlo Gorbenko

Department of Informatics and CT

NTUU "KPI"

WiMAX

WiMAX, meaning Worldwide Interoperability for Microwave Access, is a telecommunications technology that provides wireless transmission of data using a variety of transmission modes, from point-to-multipoint links to portable and fully mobile internet access. The technology provides up to 10 Mbps broadband speed without the need for cables. The technology is based on the IEEE 802.16 standard (also called Broadband Wireless Access).

The bandwidth and range of WiMAX make it suitable for the following potential applications: connecting Wi-Fi hotspots to the Internet; providing a wireless alternative to cable and DSL for "last mile" broadband access; providing data, telecommunications and IPTV services; providing a source of Internet connectivity as part of a business continuity plan; providing portable connectivity.

WiMAX subscriber units are available in both indoor and outdoor versions from several manufacturers. Self-install indoor units are convenient, but radio losses mean that the subscriber must be significantly closer to the WiMAX base station than