Fully distributed data mining

One of the group's main activities involves distributed data mining. The GOLF framework developed earlier was extended with the capability to adapt to dynamically changing environments and concept drift. The basic idea behind this method is to maintain a fixed age distribution in the network among the models. Young models help in adapting to sudden changes, while older models represent stability. The proposed method targets the maintenance of the age distribution and makes use of age during model propagation. The group members also invented new distributed solutions for bandit algorithms. The model they proposed has identical bands at each point in the network. The goal here is to allow the peers to exchange as much information as possible about the arms with as little communication as possible. The group's solution achieves a linear speedup in the number of network nodes using a gossip protocol. An additional result is related to storing graph structured databases in an optimal way over multiple sites. Graphs should ideally be stored in such a way that the number of edges between any two sites is minimal, which leads to a partitioning problem. The research group proposed a fully distributed simulated annealing heuristics to tackle the problem, and demonstrated its favorable performance over several real-world networks. The paper describing this research won the best paper award at IEEE SASO 2013. The team's researchers also studied peer sampling algorithms, which are key components of peer-to-peer systems. They focused on environments where creating new connections is expensive, and where NAT-ed nodes are present, and proposed a solution in which a dynamically, but very slowly changing network is maintained, over which short random walks are performed.