Data Driven Profiling of Dynamic System Behavior using Hidden Markov Model based Combined Unsupervised and Supervised Classification

Cen Li

Department of Computer Science
Vanderbilt University
Nashville, TN 37235
cen.li@vanderbilt.edu

Dynamic systems are often best characterized by a combination of static and temporal features, with the static features describing time-invariant properties of the system, and the temporal features capturing dynamic aspects of the system. Our goal is to construct context based temporal behavior models of dynamic systems using information from both types of features.

Our dynamic system profiling framework consists of three main steps: (i) model generation, (ii) model validation, and (iii) model interpretation. Model generation step can be further decomposed into two components: (ia) temporal model generation, and (ib) context generation.

Based on temporal feature values of the systems, temporal model generation step constructs K models to account for dynamic behavior patterns. We choose Hidden Markov Model(HMM)(Rabiner 1989) representation for temporal models. One important and desirable characteristic of HMM is that the hidden states of a HMM can effectively be used to model the set of potentially valid stages going through by a dynamic system and the directed probabilistic links between states be used to model its transition patterns among the set of stages. Our HMM clustering scheme tries to improve upon existing methods in two ways: First, existing HMM clustering systems assume fixed, pre-specified HMM topology. To obtain better fit models, we propose a dynamic and automatic HMM refinement procedure that interleaves with the clustering process and constructs HMMs of appropriate topologies for individual clusters. Bayesian model selection criteria(Chichering & Heckerman 1997) are employed in this process. Second, existing HMM clustering systems rely on predefined threshold values to determine number of clusters, i.e., the value of K, in the final partition. We take a model based approach(Cheeseman & Stutz 1996) Our clustering model is composed of clusters in the current partition and one hidden state that assigns cluster membership for each object. Given this clustering model structure, the number of clusters in the final partition is one that gives the highest model posterior probability.

The K clusters derived from temporal model generation step provide class labels for data objects. Supervised classification can then be applied to induce contexts, or pre-conditions, of each temporal model based on information from static features. We plan to use C4.5(Quinlan 1993), a decision-tree based classifier, for this process. Given a set of labeled static data, C4.5 generates a classification tree and a set of decision rules characterizing each class. The end result of this two-step model generation procedure is K context based dynamic behavior models.

The models will be validated through likelihood tests. Given an object from test data, we first determine its most probable temporal model based on its static feature values and the context definitions of the set of models. Then we compare the likelihood of its temporal data given its appointed model against the likelihood of the data given the other models. If the percentage of all test cases that obtain higher likelihood from their appointed model is greater than certain confidence level, then we accept the set of models as validated.

Once validated, we then incorporate domain knowledge to interpret the context-based models. Based on HMM state emission probabilities, characterized by multi-variant normal distributions associated with individual states of a HMM, state definitions may be assigned in a domain and task specific manner. Dynamic behavior of the system can be interpreted in terms of the probabilistic transitions between pairwise states.

References

Cheeseman, P., and Stutz, J. 1996. Bayesian classification(autoclass): Theory and results. In Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R., eds., *Advances in Knowledge Discovery and Data Mining.* AAAI-MIT press. chapter 6, 153–180.

Chichering, D. M., and Heckerman, D. 1997. Efficient approximations for the marginal likelihood of bayesian networks with hidden variables. *Machine Learning* 29:181–212.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257–285.