Dynamic Programming

by Carmine Casciato, McGill University

Dynamic programming is a mathematical technique for dealing with multi-dimensional maximization problems. Its inventor, Richard Bellman, introduced it in 1957 as a way of solving time-varying "physical, economic or engineering problems" (Bellman 1957). In particular, it provides a method for dealing with systems of that evolve over time. His perspective is to categorize them as being of a certain number N stages with M decisions made at each stage. The brute force method would suggest computing N*M decisions, a situation which may soon become computationally intractable as N and M increase. This problem is further compounded if the system is stochastic in nature. Bellman's solution lies in defining the structure of an optimal policy. That is, define a framework that assigns a cost metric to each decision such that the optimal solution (maximum or minimum cost) is derived at every stage, and is related to each previous stage. Chang (1972) uses this conceptual framework to outline a recursive algorithm for selecting the best subset of features from a larger set.

This idea of optimality resurfaces in music information retrieval (MIR) research involving dynamic programming. It takes another guise, however, namely the one posited by Rabiner and Huang (1993) in their seminal text on speech recognition. The problem involved a way to compare two utterances of the same word. These utterances were in the form of two vectors of short-time spectral features. Practically speaking, the two utterances will be of a different temporal length and spectral profile, so the first problem in comparing them arises in their time-alignment and normalization. "Finding the "best" alignment between a pair of patterns is functionally equivalent to finding the "best" path through a grid mapping the acoustic features of one pattern to the acoustic features of the other pattern." (Rabiner and Huang 1993) This amounts to finding two warping functions that temporally align the two spectral feature vectors. Furthermore, path constraints can be added to refine the alignment, allowing for only certain path choices. This alignment measure can then be used to determine how similar the two utterances are. It should be noted that both the warping functions as well as the constraints are heuristically determined and therefore bias any derived measurement. This form of dynamic programming, called Dynamic Time Warp (DTW), has found extensive use within MIR research.

Paulus and Klapuri (2002) use DTW to measure the similarity of rhythmic patterns. Their approach is similar to Rabiner and Huang's (1993). Nishimura et al. (2001), Heo et al. (2003), Guo and Siegelman (2004), and Adams et al. (2004) all employ DTW to align humming queries with songs in the database. Raphael (2002) uses dynamic programming techniques to make a Gaussian mixture model of rhythm and tempo computationally feasible.

References

Adams, N., M. Bartsch, J. Shifrin, and G.Wakefield. 2004. Time series alignment for music information retrieval. In *Proceedings of the International Conference on Music Information Retrieval*: 303–10

Bellman, R. 1957. *Dynamic Programming*. Princeton: Princeton University Press.

Chang, C. 1972. Dynamic programming as applied to feature subset selection in a pattern recognition system. In *Proceedings of the ACM annual conference* 1: 94–103.

Guo, A., and H. Siegelman. 2004. Time-warped longest common subsequence algorithm for music retrieval. In *Proceedings of the International Conference on Music Information Retrieval*: 258–61.

Heo, S., M. Suzuki, A. Ito, and S. Makino. 2003. Three dimensional continuous DP algorithm for multiple pitch candidates in music information retrieval system. In *Proceedings of the International Conference on Music Information Retrieval*.

Nishimura, T., H. Hashiguchi, J. Takita, J. Zhang, M. Goto, and R. Oka. 2001. Music signal spotting retrieval by a humming query using start frame feature dependent continuous dynamic programming. In *Proceedings of the International Conference on Music Information Retrieval*.

Paulus, J., and A. Klapuri. 2002. Measuring the similarity of rhythmic patterns. In *Proceedings of the International Conference on Music Information Retrieval*.

Rabiner, L., and B-H. Huang. 1993. *Fundamentals of Speech Recognition*. Prentice Hall: New Jersey.

Raphael, C. 2002. A hybrid graphical model for rhythmic parsing. *Artificial Intelligence* 137: 217–38.