High Fidelity Finite Length Markov Chain Walks

J. Genoff

Given an irreducible markov chain model, what is the complete set of the highest fidelity state sequences of a given length (finite length walks), that the model is able to produce? Here, highest fidelity property means : 1) all sequences in the set have the same relative transition frequency matrix ; 2) according to some measure, the relative transition occurrence frequency matrix is as close to the model stochastic matrix as possible ; 3) each sequence starts with a state that has nonzero initial probability. A three stage approach for generation of such complete set is discussed : 1) calculation of the exact absolute transition occurrence numbers matrix, which in general, consists of non-integer values ; 2) application of controlled matrix rounding to the latter in order to obtain integer values for all transition occurrence numbers, which introduces a bias leading to the fidelity issue ; 3) generation of all sequences that satisfy the constraints for state transition occurrence numbers from the integer matrix, which can be done by any constraint preserving permutation group generating technique. This paper is focused on the first stage an original linear algebraic solution to the problem is proposed. Three distinct and mutually complementary intuitive constraint aspects of the finite length markov chain walks are formulated and formalized as three independent systems of linear equations with the exact absolute state transition occurrence numbers being the unknowns. Each system suffers from linear dependency among its equations. After proper elimination of certain part of each system and union of what is left of the three, one aggregated system is constructed. The conditions for the existence of such system's solution are investigated and its uniqueness is proven. The influence of solution's values over the existence of respective walks is considered. Several interesting properties of the system and its solution are discussed, some of them with proofs. Since the coefficient matrix possesses a kind of sparseness, a specific optimized parallel solving procedure is devised and presented.