International Journal of Approximate Reasoning 73 (2016) 76–86

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Random walks on graphs with interval weights and precise marginals

Damjan Škulj

University of Ljubljana, Faculty of Social Sciences, Kardeljeva ploščad 5, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form 26 February 2016
Accepted 29 February 2016
Available online 7 March 2016

Keywords:
Weighted graph
Random walk
Markov chain
Imprecise Markov chain
Reversible Markov chain
Local optimization

ABSTRACT

We propose a model of random walks on weighted graphs where the weights are interval valued, and connect it to reversible imprecise Markov chains. While the theory of imprecise Markov chains is now well established, this is a first attempt to model reversible chains. In contrast with the existing theory, the probability models that have to be considered are now non-convex. This presents a difficulty in computational sense, since convexity is critical for the existence of efficient optimization algorithms used in the existing models. The second part of the paper therefore addresses the computational issues of the model. The goal is finding sets of weights which maximize or minimize expectations corresponding to multiple steps transition probabilities. In particular, we present a local optimization algorithm and numerically test its efficiency. We show that its application allows finding close approximations of the globally best solutions in reasonable time.

1. Introduction

1.1. Modelling uncertainty in Markov chains and weighted graphs

Markov chains with the property that every sequence of states is equally likely no matter whether the process runs forwards or backwards are said to be *reversible*. Reversible Markov chains are often interpreted and modelled with *random walks* on weighted graphs [1–7] where the states of the chain are the vertices of the graph and transition probabilities are proportional to the weights of the edges incident to the initial vertex. Reversible Markov chains are often used in Monte Carlo methods [8–10]. Random walks on graphs have become very popular in network analysis [11–14], social networks [15–18] and web recommender systems [19].

Modelling real world phenomena with Markov chains requires estimating a large number of parameters. Even with ever growing amounts of data at disposal this task is often impossible to achieve without serious uncertainty in the estimates. Ignoring this fact and regarding the parameters as precise leads to overprecise unreliable results. The need for more robust models for probability has led to various models known under the common name as *theory of imprecise probabilities* [20]. In particular, for Markov chains the theory of *imprecise Markov chains* has been developed for discrete [21–23] as well as continuous case [24]. Most of the existing models are based on the theory of lower previsions [25].

In the core of the theory of imprecise Markov chains is the idea that transition probabilities at each step are modelled with convex sets of probability distributions rather than single transition probabilities. Equivalently, all relevant probability

E-mail address: damjan.skulj@fdv.uni-lj.si.

http://dx.doi.org/10.1016/j.ijar.2016.02.008
0888-613X/

distributions can then be modelled by superadditive functionals called *coherent lower previsions*, which are defined as lower envelopes of sets of linear functionals.

Weights in graphs often also reflect some relation between vertices obtained on the basis of imperfect data. One way of expressing the resulting uncertainty is to use intervals instead of precise weights. While being a compelling generalization, the related optimization problems seem to be generally hard [26,27]. Up until now finding minimum spanning tree and shortest paths in graphs with weighted intervals has received a lot of attention, while random walks have not yet been explored, as it seems. The lack of appropriate models of imprecise Markov chains and the apparently high complexity of the general model might be among the reasons for this. The high complexity is also the main reason for our decision to keep our model simple by not allowing weights to vary completely freely within interval bounds, but instead assuming the sum of weights of edges incident to a given vertex to be constant. This could only be efficiently achieved by allowing self-loops, which then contain the non-allocated weight mass.

1.2. Model

The aim of the present article is to extend the theory of imprecise Markov chains to the case of reversible chains; more specifically, random walks on weighted graphs with interval weights. Interval weights are interpreted as sets containing the precise weights that will actually set the probabilities of transitions. We also assume that weights are not constant in time but rather at every time step an unknown mechanism selects a new set of weights, for which the only information we have is that they belong to the given intervals. Once the weights at certain time step are selected, transitions are calculated in the usual way. In our model we restrict the set of weights by requiring that the total sum of weights of edges incident to a vertex is constant and precisely known. This is achieved by assigning the remaining mass to the self-loops. This restriction will allow an efficient local optimization for calculation of multiple steps probability bounds. Actually a similar effect is the result of the rate of leaving a state when modelling continuous time Markov chains. Having precisely given marginals while dependencies are imprecise is not that uncommon since usually there is a lot more data available about marginal values than about dependencies.

In comparison to the existing models of imprecise Markov chains the most important differences are that probability models behind our model are not necessarily convex and that in general they do not satisfy Bellman's principle of optimality (see e.g. [28]). Consequently calculating bounds for multiple steps transition probabilities is a much more computationally intensive task.

We give the detailed description of the model in Sections 2 and 3.

1.3. Results

While our theoretical model is not very different from other models of imprecise Markov chains, there are substantial differences when it comes to computations. We will investigate computing n-step transition probabilities, which are the basis for any analysis with Markov chains. As imprecision is involved, we cannot speak about single precisely given transition probabilities, but rather their lower and upper bounds. Moreover, in the case of imprecise probabilities, bounds for elementary events are not sufficient to specify the corresponding probability models. Therefore we have to consider computing bounds for more general expectations.

The existing models of imprecise Markov chains allow setting transitions from one state to others independently from one another. This ensures convexity of the underlying probability models and possibility to apply Bellman's principle of optimality. These properties then imply existence of a single local and therefore also global optimum, which is found by sequentially maximizing expectations via linear programming. Complexity of the problem thus remains linear in the number of time steps. The problem of finding extremal expectation in our settings becomes considerably more complicated. In general the problem is not convex and neither it satisfies Bellman's principle. Consequently, in general multiple local optima exist, and backwards induction is not applicable. This means that the irreducible dimensionality of the problem grows exponentially with the number of time steps.

Our main numerical result is a local optimization algorithm which we propose in Section 4. Given an initial weight function it returns a local optimal solution. Global extrema, though, are still sought by taking various starting points and by doing local optimization. As the size of the space of all feasible points is far too big to be tractable by any reasonable computer, we cannot provide a criterion that would definitely ensure that obtained solution is global maximum. But numerical testing shows that in most cases a reasonable approximation of the global solution can be obtained by taking a moderate number of starting points. Even more convincingly it shows that if weight functions were chosen at random, without applying the local optimization, then it would almost certainly take incomparably larger samples to get results comparably close to the optimal solution. While, as far as we are aware, no other algorithms exist for optimization of random walks on graphs with interval weights, we can only compare our method to random choice, which is therefore by far outperformed.

2. Model settings

Let $\mathcal{X}$ be a nonempty set of *states*. We will usually denote the number of states by s. We consider random walks on the graph with vertices $\mathcal{X}$ and weighted edges that are given in the form of an interval *weight function*. The probabilities of