Optimal Cancer Prognosis Under Network Uncertainty

Mohammadmahdi R. Youssefi
Department of Electrical and Computer Engineering
The Ohio State University
Columbus, Ohio
yousefi@ece.osu.edu

Lori A. Dalton
Department of Electrical and Computer Engineering
Department of Biomedical Informatics
The Ohio State University
Columbus, Ohio
dalton@ece.osu.edu

ABSTRACT
Typically, a vast amount of experience and data is needed to successfully determine cancer prognosis in the face of (1) the inherent stochasticity of cell dynamics, (2) incomplete knowledge of healthy cell regulation, and (3) the inherent uncertain and evolving nature of cancer progression. There is hope that models of cell regulation could be used to predict disease progression and successful treatment strategies, but there has been little work focusing on the third source of uncertainty above. In this work we investigate the impact of this kind of network uncertainty in predicting cancer prognosis. In particular, we focus on a scenario in which the precise aberrant regulatory relationships between genes in a patient are unknown, but the patient gene regulatory network is contained in an uncertainty class of possible mutations of some known healthy network. We optimistically assume that the probabilities of these abnormal networks are available, along with the best treatment for each network. Then, given a snapshot of the patient gene activity profile at a single moment in time, we study what can be said regarding a patient’s treatability and prognosis. Our methodology is based on recent developments on optimal control strategies for probabilistic Boolean networks and optimal Bayesian classification. We show that in some circumstances, prognosis prediction may be highly unreliable, even in this optimistic setting with perfect knowledge of healthy biological processes and ideal treatment decisions.

Categories and Subject Descriptors
G.3 [Mathematics of Computing]: PROBABILITY AND STATISTICS—Markov processes, Multivariate statistics; G. 2.2 [Mathematics of Computing]: Graph Theory—Network problems; J.3 [Computer Applications]: LIFE AND MEDICAL SCIENCES—Biology and genetics, Health

General Terms
Theory

Keywords
Cancer prognosis, network modeling, Bayesian classification

1. INTRODUCTION
Cancer prognosis is defined as an estimate of the likely course and outcome of a cancer, which is often viewed as the chance that the disease will be treated successfully and that the patient will recover [1]. A central problem in translational medicine is thus to decide, given biological knowledge and a collection of observations, whether a cancer patient will bear any chance of successful treatment.

There are a myriad of approaches to model both normal (healthy) and aberrant (cancerous) cell dynamics, including biological pathways, co-expression networks, Bayesian networks, (probabilistic) Boolean networks (BNs), Petri nets, differential equation-based networks, etc. It is believed that these may be used to predict disease progression and successful treatment strategies, which has led to much work on the identification and analysis of biological networks in genomics and biomedicine.

There remain two questions regarding prognosis. First, even if the underlying network of a patient were perfectly known, and the best drug to use for the patient were also known, would a patient necessarily be curable? Second, suppose the precise network of a patient were unknown, but probabilities of an uncertainty class of networks, for instance all possible mutations of some healthy network, were available along with the best drug to use for each abnormal network. Then based on available measurements, say genomic or proteomic profiles of the patient, what could be said regarding a patient’s treatability and prognosis? That is, might the uncertain progression and unique characteristics in each individual make it impossible to predict prognosis, even given perfect knowledge of all biological processes and ideal treatment decisions?

PBNs are a class of dynamical models for functional gene regulatory networks (GRNs) [6]. They can capture the intrinsic uncertainty of gene interactions and measurement error, rendering GRN dynamics as Markov chains. They also provide a systematic way of modeling intervention scenarios, where the theory of discrete-time Markov processes can be applied to determine optimal intervention strategies [8, 9]. The steady-state distribution (SSD) of the model Markov chain reflects the long-term behavior (phenotypes) of the underlying network, and changes imposed on the SSD through various types of network intervention serve as a guide for developing beneficial treatment strategies so that the gene activity profiles (GAPs) evolve in a desired manner.
Managing uncertainty is especially important in modeling biological networks, where there is inherent uncertainty in the state of a network due to immeasurable latent variables, as well as uncertainty due to a lack of knowledge or partial knowledge of the relationships between observable variables even in a healthy network [5, 7, 4]. Here we focus on a third source of model uncertainty due to the inherent unpredictability of somatic gene mutations or aberrant pathway malfunctioning that may arise in a cancer. This corresponds to listing plausible scenarios in which a healthy network may undergo a functional disruption in normal gene regulation. It is imperative to take into account this uncertainty to provide a robust decision regarding cancer prognosis.

2. METHODOLOGY

We assume that a patient’s network belongs to an uncertainty class of networks, each derived from a known healthy network that contains some structure essentially common to all networks. Each network in the uncertainty class possesses one or more “mutations” of the healthy network, representing various possible subtypes or stages of the cancer. Some networks in the uncertainty class may be very treatable (good prognosis), while others may be difficult or impossible to treat (bad prognosis). In fact, we will partition the space of networks into four classes based on the severity of the disease with treatment, and the benefit of treatment. We measure the severity of disease by the long-run probability that cancerous cells visit certain known undesirable states, or equivalently, the SSD mass of these undesirable states. We measure the benefit of treatment by the difference between the steady-state mass of undesirable states before and after treatment, which we call the steady-state shift.

Our objective is to optimally classify patients into our four prognosis categories, and to study the impact of network uncertainty on predicting prognosis. Recent work on optimal Bayesian classification (OBC) furnishes an elegant framework for designing optimal classifiers and optimally estimating their error [2, 3]. In the general setting, it is assumed that the true underlying sampling distribution belongs to a parameterized uncertainty class of distributions associated with a known prior probability distribution. Closed form solutions are available for several models with conjugate priors.

In this work, we assume a single GAP is observed from the patient, which is essentially a snapshot of the state of the patient’s network at the moment the sample is drawn. A patient’s sampling distribution is thus equivalent to the steady-state distribution of their network without control, giving a correspondence between the uncertainty class of networks and the uncertainty class of sampling distributions. We impose a prior distribution over the uncertainty class of networks, with the interpretation that certain mutation events are more or less likely with known probabilities. We can therefore cast our classification problem in a discrete Bayesian setting and directly apply closed-form optimal OBC classification and Bayesian error analysis. Note there is no training data per se, since we are modeling uncertainty in the progression of cancer itself while assuming a perfect understanding of cell regulation, as opposed to modeling uncertainty due to ignorance of biological relationships between genes, where knowledge could be enriched with training examples.

While the optimal classifier makes optimal prognosis predictions under network uncertainty, obtaining the GAP or any other relevant information from a patient has the effect of reducing uncertainty. A key point in this work is that we study performance with respect to prognosis only. Although one must overcome network uncertainty, it is not necessary to be able to actually infer the network or any mutations, rather, for our purposes one only needs enough relevant data to make good predictions regarding prognosis. Thus, a second major question we address is whether it is possible to successfully predict prognosis with a relatively small amount of data. For instance, here we base predictions on a single GAP measurement, whereas network identification typically requires a massive collection of time-series measurements. Prognosis performance depends on many factors, including the type of cancer (the original healthy network and its associated uncertainty class), the individual patient’s network, and the particular sample drawn from the patient. Very often prognosis prediction from a single GAP is highly unreliable, even in this optimistic setting with perfect knowledge of healthy biological processes and ideal treatment decisions. In this case, the remedy is to collect more data, for instance time-series GAP measurements, to help identify the patient’s network, or at least ensure reliable prognosis.

3. REFERENCES

[1] Understanding cancer prognosis (www.cancer.gov/cancertopics/factsheet/support/prognosis-stats).
[2] L. A. Dalton and E. R. Dougherty. Optimal classifiers with minimum expected error within a Bayesian framework – part I: Discrete and Gaussian models. Pattern Recogn., 46(5):1301–1314, May 2013.
[3] L. A. Dalton and E. R. Dougherty. Optimal classifiers with minimum expected error within a Bayesian framework – part II: Properties and performance analysis. Pattern Recogn., 46(5):1288–1300, May 2013.
[4] M. S. Esfahani and E. R. Dougherty. Incorporation of biological pathway knowledge in the construction of priors for optimal bayesian classification. IEEE/ACM Trans. Comp. Bio. and Bioinf., 11(1):202–218, Jan. 2014.
[5] M. S. Esfahani, J. Knight, A. Zolla, B.-J. Yoon, and E. R. Dougherty. Classifier design given an uncertainty class of feature distributions via regularized maximum likelihood and the incorporation of biological pathway knowledge in steady-state phenotype classification. Pattern Recogn., 46(10):2783–2797, Oct. 2013.
[6] I. Shmulevich, E. R. Dougherty, S. Kim, and W. Zhang. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics, 18(2):261–274, Feb. 2002.
[7] B.-J. Yoon, X. Qian, and E. Dougherty. Quantifying the objective cost of uncertainty in complex dynamical systems. IEEE Trans. Signal Proc., 61(9):2256–2266, May 2013.
[8] M. Yousefi, A. Datta, and E. Dougherty. Optimal intervention in Markovian gene regulatory networks with random-length therapeutic response to antitumor drug. IEEE Trans. Biomedical Eng., 60(12):3542–3552, Dec. 2013.
[9] M. R. Yousefi and E. R. Dougherty. Intervention in gene regulatory networks with maximal phenotype alteration. Bioinformatics, 29(14):1758–1767, July 2013.

ACM-BCB 2014

683