On the personalized estimation of cancer risk:

Bohao Tang

Background. Cancer is one of the most complex and deadly challenges in today's world. Since it has no efficient cure in most situations, predicting and finding it in time become very important questions. There are lots of study on the cancer risk, but most of them are focused on the case study of population based risk, like in [1], [2]. On the other side, researchers have found and still be finding the factors that influence the cancer risk, which have certain phenomenon that can help us to predict cancer. For example, it's well known that smoking will increase probability for certain kind of mutation and therefore increase risk for certain cancer. There are also researches finding out that many cancers are caused only by some random replicative errors, which is called "Bad Luck Theory" [3]. After all these findings, a natural question arises: Since we have known some factors that will influence cancer risk and we nowadays can collect more and more data from a single person, can we develop methods to estimate cancer risk from that specific person? This proposal is to suggest these kinds of researches. We can build models that contain current factors and find new factors, which in the end will be used to do personalized estimation, also we can collect personal activity data to study whether certain act will influence the risk and how long will it need to be significant.

Hypothesis. The influence of factors and body activities to the cancer risk for a specific person is stable so that we are able to find them out. Also, the personalized risk is just a small random fluctuation from the population risk, so that we can use the data of population to do inference of a specific person.

Approach. First, we need to specific a kind of cancer, say breast cancer. Then we need to collect data, like gene expression, environment, behavior … of population that are related to breast cancer and which are not. Second can used methods from supervised learning to factorize the factors that influence gene mutation distribution and related them to certain behavior or gene pattern. Then do cross validation to valid them. Third, we build stochastic process models to model the mutation and the relation between mutation and cancer, we can use the hidden Markov model. Finally we build a mobile software that will automatically collect data from specific person and use the model to inference his cancer risk.

Summary. To make the public more aware of the risk of cancer and make them go to do cancer test more efficient, this kind of mobile based personalized cancer risk estimation is important and worthy for study. Also we can find more factors that will influence cancer and know deeper about how cancer happens, which will on the other hand help us to understand cancer and finally beat it.

Reference.

[1] Kelsey, Jennifer L., Marilie D. Gammon, and Esther M. John. "Reproductive factors

and breast cancer." Epidemiologic reviews 15.1 (1993): 36.
[2] Chan, June M., et al. "Plasma insulin-like growth factor-I and prostate cancer risk: a prospective study." Science 279.5350 (1998): 563-566.
[3] Tomasetti, Cristian, and Bert Vogelstein. "Variation in cancer risk among tissues can be explained by the number of stem cell divisions." Science 347.6217 (2015): 78-81.