CS134 Project Proposal:

Application of Support Vector Machines to Separate Dose of Radiation Exposure in Human Fingernails

Xiaoming He

April 13, 2010

Rapid and accurate retrospective dosimetry is of critical and strategic value in the emergency medical response to an accidental or mass casualty radiological/nuclear event. The assessment of individual exposure doses in these situations is one crucial parameter in the triage process for determining the levels of ionizing radiation. One technique that has the features required for performing rapid and accurate dosimetry assessment is based on the ex vivo measurement of a relatively stable radiation-induced signal in human fingernails by electron paramagnetic resonance (EPR) spectrometry methods. One advantage of this approach is fingernails are readily available and sample collection is non-invasive.

One of the major goals is to achieve a simple screening into “significantly radiation exposed” and “not significantly radiation exposed” categories which can be used to make clinical decisions. Experimental data collected based on the irradiated or non-irradiated human fingernails in a simulated situation can be analyzed to see whether and how these two categories can be separated. The model built from experimental data can be used to guide the diagnosis in the triage situation to a large population.

One challenge here is the generation of unwanted background signals by clipping itself due to the reason of ex vivo measurement. This unwanted signal can largely interrupt the radiation induced signal which we are interested. Another reason is the complexity of the signal itself because fingernails are composed of alpha keratin which is a complicated large protein molecule that can give rise to multiple products when breaking chemical bonds. Several other confounding factors include the temporal changes of signal intensity, relatively low signal to noise ratio and so on. So these two classes of data are hard to separate by a simple linear model.

I will employ support vector machine (SVM) to build a reliable model for the classification of irradiated samples and non-irradiated samples. SVM is a supervised learning technique which is capable to distinguish two categories of data (pattern recognition) using a boundary defined by a complex hypersurface. The SVM boundary is only minimally influenced by outliers while most commonly used analysis method is largely influenced by outliers. SVM has the characteristic to do qualitative and quantitative predication for several kinds of data in chemometrics which is the science of extracting information from chemical systems by data-driven means. SVM model can be used for a more homogeneous and automatic evaluation.

Two large datasets of training data (100 samples) and test data (50 samples) can be provided by my current research lab. Machine learning approach has not been applied in this research area yet. My previous study is chemical mechanism oriented in the ionizing

radiation process. Although no existing published research results can be obtained on these kinds of data from SVM, I think there is still a chance to get some exciting results.

There are two problems I need to solve in order to extract useful results from SVM approach. The first one is the descriptor selection. Determination of which features are useful to classify the two categories is very crucial. The spectroscopic data I will use are spectra scanned in the dimension of magnetic field and in the dimension of time. Some work has already been done to characterize the physical and chemical features of these spectra. Based on these experiences, optimal selection of the multiple dimensional vectors is likely to be reached. Another problem is to determine which kernel functions can best capture the characteristics of the data. So I will test several kernels to identify the best functions.

I plan to use the training set to build few simple SVM models based on some selected descriptors by the time of milestone. After the milestone, I will optimize the descriptors and kernel functions to get the best separation.

References:

[1] O. Ivanciuc, Applications of Support Vector Machines in Chemistry. *Reviews in Computational Chemistry*. Volume 23, 2007, pp. 291-400
[2] H. Li, Y. Liang, Q. Xu, Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems*. Volume 95, 2009, pp. 188-198
[3] http://www.isis.ecs.soton.ac.uk/resources/svminfo/
[4] X. Chen, X. Yu, D. Wu, Y. He, Application of Least-square Support Vector Machines in Qualitative Analysis of Visible and Near Infrared Spectra: Determination of Species and Producing Area of Panax. *Fourth International Conference on Natural Computation*. IEEE, 2008, pp. 108-111