BingPredicts: Predicting the Future with Big Data

Predictive models aim to extract inferences from large datasets which are unobservable at the surface, even to the best domain experts. Microsoft has access to a wealth of web data through its Bing search engine and through collaborations with Twitter and Facebook. Within Bing, we have been experimenting with useful ways to harness the power of the data we have to project outcomes of events. This could be anything from an election or sporting event to whether an actor will win the People's Choice Award.

This talk will provide an overview of the Bing Predicts system which was launched earlier this year. Bing Predicts utilizes this data to build predictive models for different events. So far, our predictions have included outcomes for Brazil World Cup, NFL, NBA Draft, Reality TV shows, the Scotland Referendum, and the Emmys.

We will outline the data that's available to us and discuss the predictive models we built for the National Football League (NFL) and reality voting shows. We will also show how our models can be extended to help solve real world problems, such as predicting a disease outbreak. This talk will examine how predictive models are used with some focus on the technical details of our data and predictive models. We will also discuss some of the machine learning and scale challenges involved in generating these predictions.

Bio

Saad Sheikh is a Data and Applied Scientist at Microsoft's Bing division. Dr. Sheikh received his Ph.D. from the Department of Computer Science at University of Illinois at Chicago in 2009. The title of his dissertation was "Combinatorial Methods in Kinship Analysis". He held postdoc appointments by INRIA and CNRS at École Polytechnique, Paris. After France, Dr. Sheikh worked as a postdoc at the University of Florida. His research interests include algorithms and complexity in computational biology, population genetics, structural bioinformatics, parallel algorithms, optimization and combinatorial stability. He has published papers in prestigious peer-reviewed conferences and journals including ISMB, IPDPS, Bioinformatics, INFORMS Journal of Computing, Journal of Bioinformatics and Computational Biology, and Endangered Species Research.