Executive Summary: EOL Data Mining Project

Enron has been enjoying significant revenue growth after the successful launch of Enron's on-line commodity trading platform EnronOnLine (EOL) in November of 1999. In fact EOL is the largest e-commerce website in the world with daily notional value of transaction of 3 billion US dollars and life-to-date transactions of 1.4 millions. There are many listed products traded on EOL ranging from liquid natural gas and power contracts to less liquid or illiquid products such as steel and freight.

EOL is a principal based trading platform, meaning Enron is the buyer (seller) when there is seller (buyer) who wants to transact on EOL. EOL provides market liquidity by making the bid-ask spread. However making the spread is not the only revenue source for running EOL. There is certain information asymmetry beneficial to Enron as the market maker:

- Enron owns EOL trading database that contains detailed information about each transactions; trades can be aggregated according to different categories, for example, by commodity, by contract maturity, by counter party, by trading time interval, just to name a few. The informational advantage will allow us to explore market inefficiency and arbitrage across different products.
- The time series recorded in EOL data base contains valuable information about supply-demand balance, market directions and volatilities, market correlations and cross-market correlations, trading habits and patterns.

The EOL Data Mining project is aimed at taking the advantage of the information asymmetry and market inefficiency so as to predict the market conditions. The benefit of predictability is obvious, especially in the following aspects:

- Predictability means profit. The ability to predict (even in a statistical sense) will give us an edge in trading and risk management.
- Predictability will enable us to control and reduce the risk of market making.

Data Ming is a new field that combines the business insights with computer learning capabilities. The business insights are translated into certain quantitative state space models [1] (most likely non-linear time series models). Then the best model is selected deductively to fit the reality the most. A different approach is gaining popularity, that is the inductive methodology [2].   The inductive method takes advantage cheap computation power and artificial intelligence, builds the prediction model by learning the patterns contained in the time series.   In the EOL Data Mining project, we will exploit both type approaches to build our ultimate Enron Perdition Models.

[1] Weigend, A. S. & Gershenfeld, N. A. (eds) *Time Series Prediction: Forecasting the Future and Understanding the past* (Santa Fe Inst. Studies in the Sciences of Complexity, Addision-Wesley, Reading, MA, 1993).

[2] *Introduction to Data Mining and Knowledge Discovery*, 3rd Ed., Two Crows Corp. 1999.