Big Data and Machine Learning in Finance: A Research Agenda

Rajesh T Krishnamachari, PhD
Research Topics

1. Impact of big/alternative data on financial markets
2. A grand unified theory of data science for financial time-series
3. New risk premia - Beyond behavioral finance and linear factors
New Data

**Taxonomy**

<table>
<thead>
<tr>
<th>Alternative Data</th>
<th>Individuals</th>
<th>Business Processes</th>
<th>Sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Media</td>
<td>Transaction Data</td>
<td>Satellites</td>
<td></td>
</tr>
<tr>
<td>News &amp; Reviews</td>
<td>Corporate Data</td>
<td>Geolocation</td>
<td></td>
</tr>
<tr>
<td>Web Searches &amp; Personal Data</td>
<td>Government Agencies Data</td>
<td>Other Sensors</td>
<td></td>
</tr>
</tbody>
</table>

1. Value in each data-set
2. Change in market structure
3. Limitations of Big Data
4. Other concerns?

- Using News Sentiment
- Using Email Receipts
- Using footfall data
New Tools

What can ML do for a Quant?

1. Clustering / Dimensionality Reduction
2. Feature Selection
3. Interaction Terms
4. Replacement of If-else

Statistics
- Combine intuition with data

Econometrics
- Causation from Correlation

Signal Processing
- Signal from noise

Machine Learning
- Patterns in data

Latest algorithms are in open-source domain. Implications?
New Risk Premia

Linear zoo → Non-linear ?

Behavioral Finance → Statistical

- Sector Rotation
- Stock Selection
- Regime Detection
- Style Rotation
- Portfolio Construction
- New Visualization

Using Random Forest / Xgboost

Risk Premia 2.0

Using hierarchical risk parity

FX Vol using PCA