Future of Analytics – a definitive Roadmap

PG Madhavan, Ph.D.
Chairman, Syzen Analytics, Inc., Seattle, WA, USA
pgmad@syzenanalytics.com

HIGHLIGHTS: We describe a definitive roadmap for Analytics technology with examples of breakout applications in business.

1. An application today that exploits some of the methods described in this article.
2. Derive powerful new features from system parameters.
3. A practical framework for "closed-loop" and "real time" Analytics using Systems Theory approach.
4. Take the next step by realizing that customers are embedded in social influence networks and that the network actions are not instantaneous but have spatial and temporal extent.
5. When embedded in a spatio-temporal network, manage the high complexity of data by using a "source model" approach and focus on coupling among the sources.

This new framework is called "SYSTEMS Analytics" and lays the ground work for a new quantitative and powerful paradigm in Analytics.

Dr. PG Madhavan is the Founder of Syzen Analytics, Inc. He developed his expertise in Analytics as an EECS Professor, Computational Neuroscience researcher, Bell Labs MTS, Microsoft Architect and startup CEO. PG has been involved in four startups with two as Founder.

Major Original Contributions:

- Computational Neuroscience of Hippocampal Place Cell phenomenon related to the subject matter of 2014 Nobel Prize in Medicine.
- Random Field Theory estimation methods, relationship to systems theory and industry applications.
- Early Bluetooth, Wi-Fi, 2.5G/EDGE and Ultra-wideband wireless technology standards and products.
- Currently developing Systems Analytics bringing model-based methods into current Analytics practice.

PG has 12 issued US patents and over 100 publications & platform presentations to Sales, Marketing, Product, Industry Standards and Research groups. More at www.linkedin.com/in/pgmad

The future of Analytics is bright indeed! It is clearly more than a technology *du jour* – why? With Big Data, Analytics can and is already providing actionable insights to drive businesses success.

In a recent blog, we attempted to unify Machine Learning as we know it (since the 1970's) and pointed out some future directions ("Unifying Machine Learning to create breakthrough perspectives"). The current article explicitly lays out one of the many directions for Analytics evolution – specific technologies, frameworks and algorithms are described that a practitioner can start adopting today and grow in many fruitful directions.

In this article, we consider a specific yet broad application scenario of Retail Commerce (*notice that our roadmap will be particularly useful for automated Analytics necessary for Internet-of-Things (IoT) applications – we will mention some insights throughout this article*). We describe our Analytics Roadmap in the order of distance into the future (and hence with less and less specificity):

1. An application today that exploits some of the methods described in this article.
2. Derive powerful new features from system parameters.
3. A practical framework for "closed-loop" and "real time" Analytics using Systems Theory approach.
4. Take the next step by realizing that customers are embedded in social influence networks and that the network actions are not instantaneous but have spatial and temporal extent.
5. When embedded in a spatio-temporal network, manage the high complexity of data by using a "source model" approach and focus on coupling among the sources.

This new framework is called "SYSTEMS Analytics" and lays the ground work for a new quantitative and powerful paradigm in Analytics.

Introduction
The subject matter of this article goes by many names. We prefer to use "Analytics" and "Machine Learning or ML" interchangeably (without worrying too much about nuanced meanings). in fact, it has been noted that, "*Machine learning takes many different forms and goes by many different names: pattern recognition, statistical modeling, data mining, knowledge discovery, predictive analytics, data science, adaptive systems, self-organizing systems, and more …"* (from "The Master Algorithm", Pedro Domingos, 2015).

Considering why Predictive Analytics is important briefly, it has been noted that a prerequisite for performance at a high level in business is the ability to understand and manage complexity. Complex systems to be managed properly requires a ton of data at the right time. BIG Data provide us the data we need; to put these data to work in order to take us to the high levels of complexity required while still managing it, ***we have to anticipate what is about to happen and react when it happens in a closed loop manner***. Predictive Analytics will allow us to push our "system" to the edge (without "falling over") in a managed fashion. ***This is why businesses embrace Predictive Analytics - to manage businesses at a high level of performance at the edge of complexity overload.***

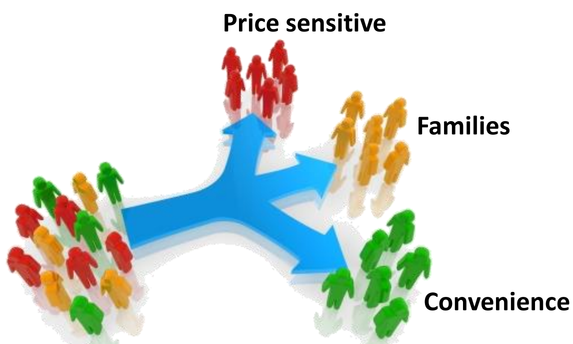
Now, let us consider a current Analytics business application to level-set and create a common-ground for further developments in Systems Analytics.

Milestone 0: Today – Taking on a "sacred cow"
Recommendation Engines are commonplace today – who has not heard of Amazon or Netflix? They provide recommendations to individual shopper (for goods or movies) based on a particular shopper's likes and dislikes as well as sellers' business priorities (push out movies in the "long tail" for Netflix and sell goods in the "fat front" from Amazon warehouses).

Let us consider a slightly more difficult problem – Recommendation Engine for a GROUP of shoppers. This scenario typically arises in brick-and-mortar retail merchandising – think of your nearby corner store. How does the shop owner decide how much of what products to put on the shelves? Note that shelf space is limited; this constraint applies equally to large eCommerce warehouses where the shelves are numerous but the products to stock are even more – the challenge is "product density". This is the problem of Optimal Product Assortment in retail merchandising.

The natural approach to such a problem is to "group" the hundreds or thousands of shoppers at the corner store – via "segmentation" or "clustering". Once grouped into a few "segments", the Recommendation engine can be optimized for each segment and optimum product assortment can be derived from the proportion of these few "segments" that shop at the store.

The state-of-the-art in Commerce is "behavioral" segmentation where *the market is divided into segments based on pre-selected characteristics which applies to all product categories in a store.*



Clearly, *bucketizing* shoppers into convenient segments such as "Price sensitive" or "Families" allows one level of meaningful abstraction. Instead of addressing millions of shoppers individually, one can tailor marketing, merchandising and loyalty efforts to a handful of labelled groups.

However, what is helpful at one level can be a flawed approach for some applications. Consider a case where a particular shopper, per behavioral segmentation, ended up in the Price Sensitive bucket. While this may be true in general for her, she may have specific preferences in certain product categories; for example, while Price Sensitive in general, her wine choice may be the expensive Châteauneuf-du-Pape brand. Such misallocations when multiplied by millions of shoppers lead to flawed product assortment decisions in the case of Behavioral Segmentation applied to Merchandising.

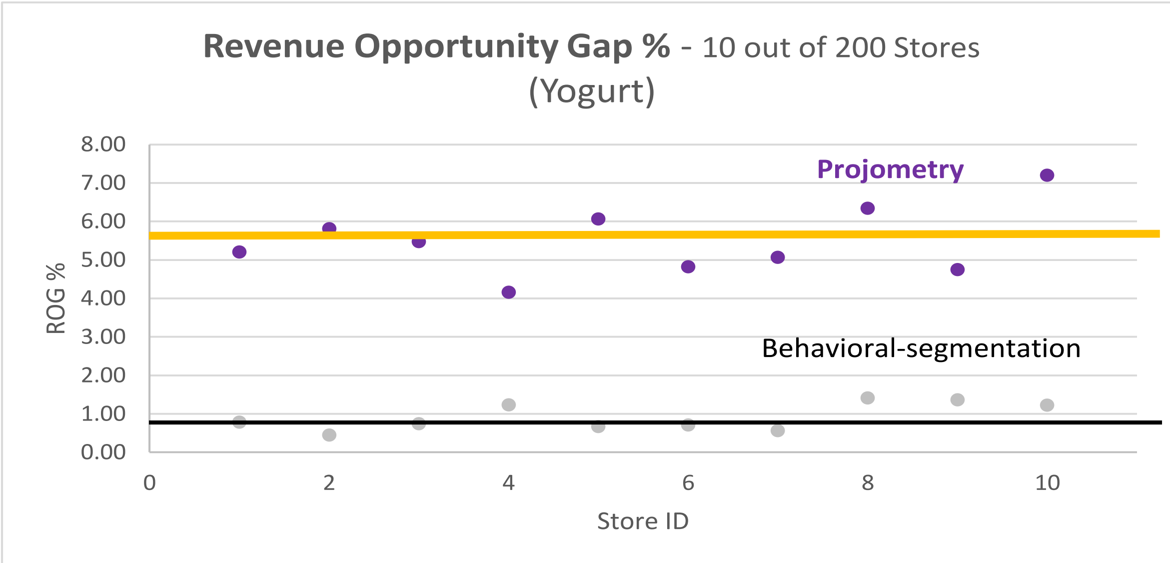
Let us consider a better approach using shopper big data and Machine Learning (ML) to create and identify "segments".

In ML segmentation, the shoppers (whatever their behavioral characteristics may be) fall into N Preference Groups based on what they actually buy (actual purchase pattern is a great proxy for true product preference). In essence, each product category is its own unique market.

Traditional behavioral segmentation would have predicted that Brand X will sell more in Store 123 because Price Sensitive shoppers prefer more of Brand X and Store 123 has more Price Sensitive shoppers.

In our ML method, we realize that since Store 123 shoppers are well-represented by N Preference Groups for a *particular* product category, the proportion of the N groups that shop at Store 123 ought to determine the assortment for that product category at Store 123. Such finer distinctions made with the aid of shopper data avoids the pitfall of employing the same behavioral groups across all product categories since shoppers' purchase propensities can vary across categories.

Comparing behavioral segmentation and ML method to optimize product assortments head-to-head, we obtained the following results. Consider Revenue Opportunity Gap (ROG) as an overall performance measure which indicates better product assortment optimization when they are high. Assume that the overall revenue

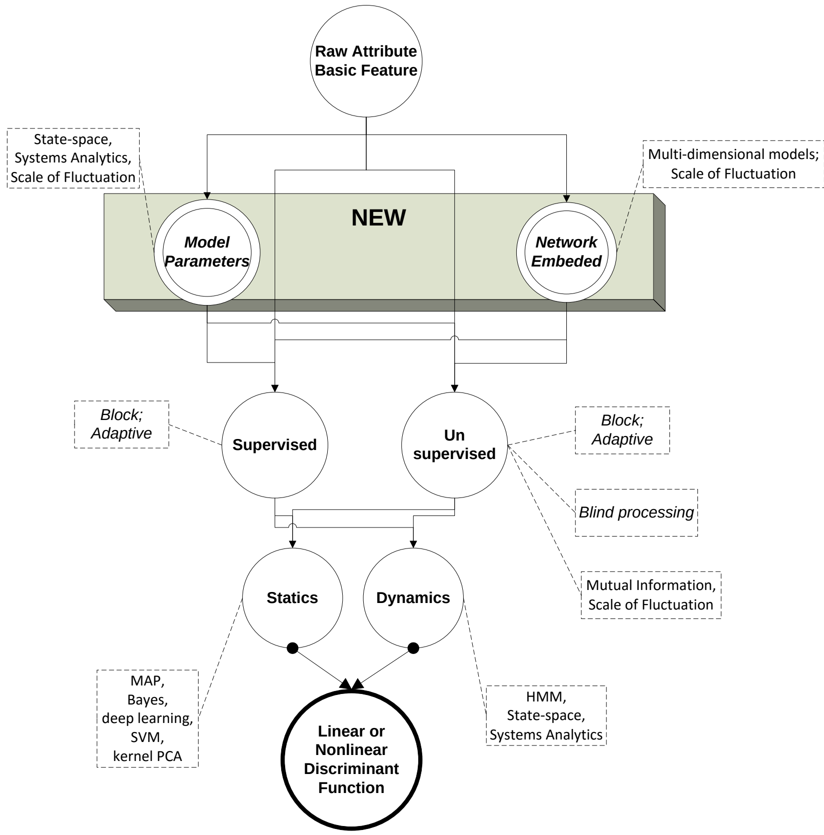


for a category (yogurt, in this example) was $100. While Behavioral Segmentation shows an average of 1% or $1 of ROG (improvement possibility), Projometry which is an ML Segmentation method shows a 5% or $5 improvement possibility. In other words, improvement due to our data-driven method is FIVE times higher than that due to Behavioral Segmentation.

Why is ML-based approach better than Behavioral Segmentation? The following table captures the reasons. Looking at more dimensions than just the head-to-head outcome comparison above, it is clear that ML method has several advantages when it comes to Merchandising. Whenever the data itself determine groups rather than being externally imposed, data analysis history has shown that results will be superior. Another nice feature is that human labor for and subjectivity in "bucketizing" can be avoided which makes analysis fast, inexpensive and repeatable. The fact that separate preference groups are generated for every product category and that ML method acts on **what** people purchase rather than **why** has led to breakout applications of this ML method in Retail Merchandising product assortment optimization.

Behavioral Segmentation	ML Method
Simplified model of demand, pre-determined number of segments	Machine determined model of demand
Sample based	Population based
Same across all product categories	Differs by product category
Subjective (behavioral assumptions)	Objective (no behavioral assumptions)
Costly and time consuming to create and maintain	Cheap, fast. No segmentation required.
Based on WHY people purchase	Focused on WHAT people purchase

Milestone 1: Unification of Machine Learning – System parameters as high-value features
From the early days of "ML", we consider Pattern Recognition and Classification as a unifying perspective for the efforts that have gone on under different names. In particular, the classic textbook of Duda & Hart, "Pattern Classification & Scene Analysis", published in 1973 is our starting point. From this perspective, given labelled samples, one obtains a class description and a decision rule that specifies a decision boundary in feature space among classes, which is called a ***"Discriminant Function". Then, most if not all current ML techniques can be seen as competing methods to derive Discriminant Functions.***



Further details of this path of integration are described in "Unifying Machine Learning to create breakthrough perspectives"). **Let us pull all of the notions discussed there into a diagram.**

Once the patterns have been recognized and classes identified, the resulting classes can be used for all sorts of applications such as Recommendation Engine, Language Translation, Fraud Detection and many others. The approach outlined above allows you to use a single framework till the application development stage.

In all of the existing ML bags of tricks, we are still staying at the surface level - we are **modeling the attributes or data DIRECTLY.** What if we went one level deeper? ***Model the SYSTEM that generates the data and then use model parameters as high-value features.*** In the retail commerce application discussed earlier, we can hypothesize that there is a system, either explicit or implicit, behind the scenes generating customer purchase behaviors and purchase propensities.

Behavioral segmentation is an approach to "modeling" the shoppers at the corner store. We were able to do better with our ML-based "Preference Groups" which is another parametrization of the shopper model; some models are better for some applications than others. As we go down our Systems Analytics Roadmap, a more sophisticated understanding of Systems approach will be essential. The elegant introduction and varied applications of Systems Theory in John Casti's books is an excellent starting point – references to his key volumes are available at Syzen Analytics, Inc.'s website (Reality Rules: I is the recommended starting point).

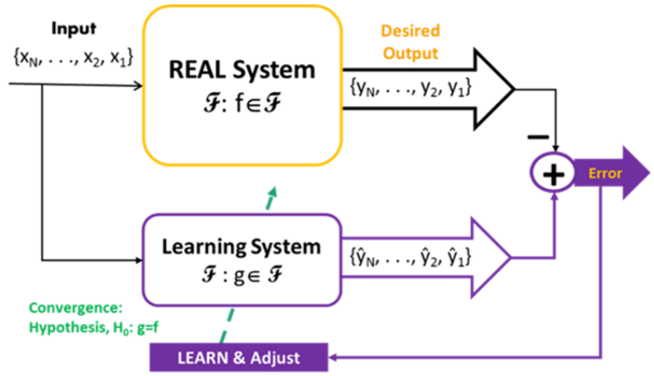
Milestone 2: Systems Analytics – Framework for closed-loop & real-time Analytics
SYSTEMS Analytics is a new extension to Machine Learning process outlined above. Traditionally, "systems" approach has been the realm of engineering control systems, process control, system identification, etc. but that

need not prevent us from exploiting the large body of powerful techniques and tools (such as Kalman Filtering) that have been developed in the past for our “machine learning” purposes. In fact, ***system models that naturally arise in ML provide the framework for SYSTEMS Analytics - a new model-based paradigm in Analytics or Machine Learning.***

<u>Basic Formulation:</u>
Consider machine learning purely in the realm of supervised learning and define it as algorithms to optimize a performance criterion using example data or past experience. Here is a figure that captures the essentials (from [Yaser Abu-Mostafa, 2012](#)).

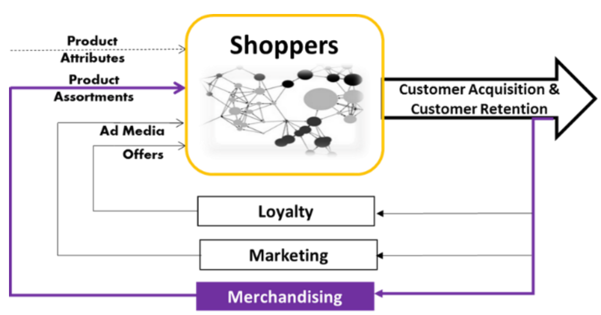
Using Training examples, Learning Algorithm tries to “learn” Target Function, f, to a good approximation, g, using a model or formula. To be explicit,

- We have a set of Models.
- Each model in the set has a Model Structure and an associated Learning Algorithm.
- Pick a Model Structure & associated Learning Algorithm; from the infinite set of functions, $\mathcal{F}$, possible within these 2 choices, there is the TRUE Target Function, f, which is in unknowable and its good approximation, g, that we try to find using our choice of Model Structure & associated Learning Algorithm and Training Data.



Learning – in Systems Analytics, there is nothing new other than a new toolset that we bring to ML but with far reaching implications. ***Systems Analytics as a new toolset for ML brings with it a vast array of tried and tested mathematical and statistical methods from Systems Theory, Random Process theory and System Identification***. Judicious use of these methods will enhance the power of ML as well as extend it in new directions!

Let us consider the Product Assortment selection problem from Milestone 0 within this framework.



Model:	Structure	Learning Algorithm	Objective
Perceptron	Perceptron nodes	Perceptron Algorithm	Find g ≈ f
Logistic Regression	Multiple Regression	Gradient Descent	
Neural Network	Multiplayer Perceptron	Back Propagation	
SYSTEMS	***State-space***	***Kalman Filter***	

The table above provides a partial list. ***The last row that we have introduced in the table defines SYSTEMS Analytics by example.*** From this table, it is clear how SYSTEMS Analytics fits in with the traditional framework of Machine

Retail Commerce from a demand-chain perspective is captured in this canonical model. Note that in Milestone 0, we considered the Merchandising part only – here we show all the main activities in the demand-chain such as

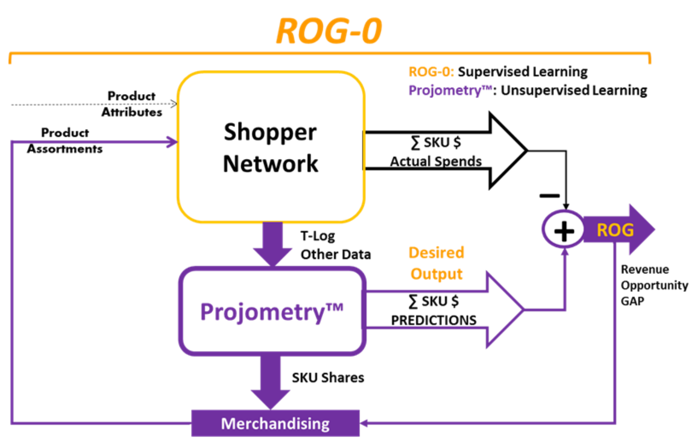
Marketing & Loyalty. Limiting ourselves to finding optimal product assortment for the corner store, we can redraw the canonical model as a closed-loop systems model as shown.

In choosing a closed-loop model, we recognize the fact that real business applications are not "one and done" solutions! First, we have to provide Product Assortment recommendations; its effect on sales over time has to be monitored; minor corrections and periodic fine-tuning have to be done to obtain business results over time. **We can think of the overall solution as "goal-seeking" over time** – somewhat like "flu shots" where the formulation has to be tweaked and shots administered frequently!

In the "ROG-0" goal-seeking solution developed by Syzen Analytics, Inc., transaction log (or T-log) data from the Retailer provides a wealth of information about shopper purchases. We use this information to group shoppers using our ML method using our Projometry™ algorithm. Once we group their activities, it is used to characterize their shopper preferences at a particular store in a chain leading to predictions of their purchase propensities. Knowing shopper preferences, we can then transform them into "SKU $ Spends". This is our Desired Output; the difference from what they actually spent in the past is the error signal used to drive the supervised learning in ROG-0. Shopper preferences that we learn using Projometry™ are utilized to prescribe the adjustments to be made to the product assortment so that the Desired Output will be achieved thus minimizing the ROG error over time.

From this brief discussion of a Systems Analytics business example, the following must be clear:

- We have a framework for a closed-loop Analytics solution.
  - Closing the loop can be via a human who takes the intermediate results (product assortment recommendations, say) and changes shelf layout.
- We have a framework for real-time Analytics.
  - Goal-seeking solution that improves and delivers results over time.
  - Time period may be days or milliseconds depending on the application.
- IoT applications that require automated Analytics can exploit this Systems Analytics framework.
  - *Applications to IoT must be obvious – think of process control in a large oil refinery or a paper mill; automatic communication pathways close the loop and electro-mechanical actuators affect the changes necessary.*



Milestone 3: Social Network – Accounting for influences

In Milestone 0, we considered individual shopper's purchase propensities which we aggregated appropriately to find product assortments. As businesses push to higher levels of performance, higher fidelity models are going to be needed to produce valuable predictions and recommendations for business operations. In the case of a shopper, she is not an isolated entity but influenced by her social and "influence" networks. In other words, **data exist in embedded forms in preference and influence networks which are distributed in time and space**.

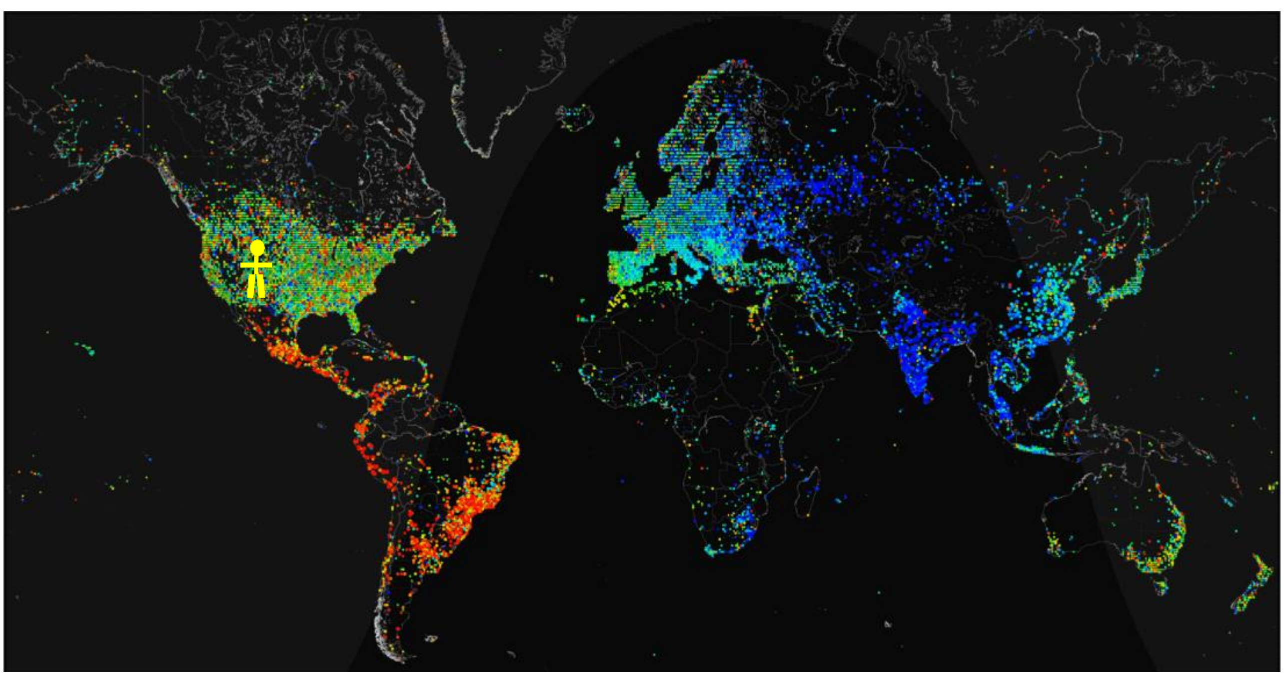
In this sequence of pictures, starting from the left, we recognize the following:

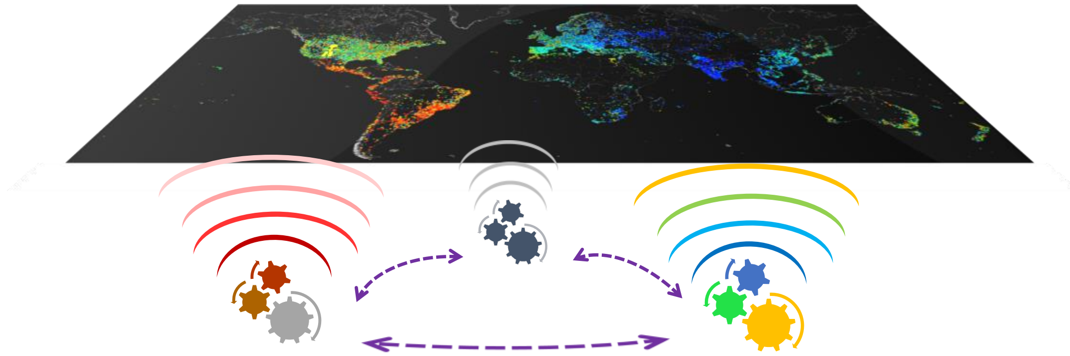
- Shopper can be considered isolated for early models but truly exists in a social network.
- In current network graph theory, it is often the case that the space and time dimensions are ignored; in other words, influences happen instantaneously – "spooky action at a distance!"
- With global connectivity, shopper is influenced by events around the world.

Brining such practical considerations into a mathematical framework is devilishly difficult. There are some tools and techniques – for example, our work in [Space-Spatial Frequency distributions] which estimate spatial frequency at a point in space may be applicable to some cases. This is an area for future development.

In general, engineers like to reduce the parameters of interest to a few scalars! In a typical IoT application, we may track the variation of a scalar over time, put thresholds or "bands" around it; when exceeded, raise an alarm to alert a human to take some automatic corrective action. Measures such as space-spatial frequency distributions simply do not lend itself to such an approach.

In fact, considering only the spatial extent is a gross simplification of the real situation. Data is changing over space and time – the following [animated GIF] captures the true complexity! Instead of attacking the spatio-temporal data analysis problem head-on, it is time to take a different approach!





We hypothesize that the "surface" activity is generated by a few real or virtual "sources" and coupling among them creates the spatio-temporal dynamics that the surface data exhibit.

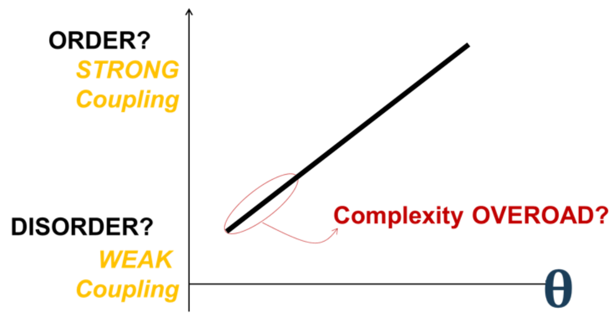
In the case of earthquake and scalp EEG studies, one can find many articles that develop "dipole" models that hypothesize deep virtual sources that oscillate producing the data recorded on the surface. Then the data description problem resolves to estimating parameters of the dipole models, a much more tractable problem. Once these "source model" parameters are estimated, they can be used as high-level features for the type of problems discussed in Milestone 0.

Milestone 4: Spatio-Temporal Dynamics – Coupling

If you think of the global pictures in the previous section as showing the preference for a product or the dynamics of a tweet or the effects of a marketing campaign, we start to realize that for certain levels of business analysis, the surface data themselves may not be as relevant as the coupling among the sources. Your business may be interested in manipulating the couplings such that desired surface activity is produced over space and time!

In the 1980's, Erik Vanmarcke developed a new fundamental formulation of the theory of Random Fields. Subsequent work has extended it to systems applications and estimation.

An important parameter in Vanmarcke's formulation is called "Scale of Fluctuation" or "θ". θ turns out to be a many-splendored thing with some amazing properties. There is an accessible discussion in my blog, "Network Dynamics & Coupling: Shannon's Reverie Reprised" (2012). Relevant to our discussion on "source coupling", here is a diagram of interest.



θ is on the X-axis. What the figure shows is that as θ increases, coupling strength increases – on the surface, its effect will be seen as nearly uniform colors in the global map in the last section; this is akin to a slow "swell" spreading across the world in a very orderly fashion. On the other hand, small θ indicates a chaotic wave pattern with local "disturbances" over the surface.

As you can tell, the explanations above are very suggestive of some aspects of a marketing campaign which can be tracked by a single scalar. θ has much deeper and more complex interpretations than as a "coupling" measure which may be valuable in Analytics applications in the future.

Summary of "Future of Analytics – a definitive Roadmap" by PG Madhavan:

- We have sketched out a roadmap of increasingly complex tools that can be brought to bear on Analytics or Machine Learning of today.
- These tools provide us with high-value features in terms of system parameters, a framework for closed-loop real-time Analytics and ways to possibly accommodate the networked nature of data sources.
- Theories of all the techniques introduced are fully or partially developed but will require thoughtful additional development to reach their full potential for Analytics applications.

I have been fortunate in being involved in the development of the basic theory of each of the five milestones described above (more at www.JinInnovation.com); I have also personally developed the base algorithms for each case. Breakthrough business applications of the later milestones will require significantly more development in collaboration with business domain experts. Syzen Analytics, Inc., the first SYSTEMS Analytics company, is following the roadmap described in this article.