SoSEdata3D

A process for mapping System of Systems data in 3d.

Dedicated to Prof Mo Jamshidi and in memory of Ahmed Jamshidi.

Special thanks to Emily Nolan and Peter Phillips for their contributions.

Anthony Nolan OAM JP
G3N1U5 Pty Ltd
Sydney, Australia
tony@g3n1u5.com

***Abstract*—**A challenge these days is how to extract and display core concepts in data. One solution is to view concepts and their relationships in three-dimensions (3D) especially data that has parent - child relationships between concepts. I see the future lying with visualizations and with understanding the differences between data points rather than the actual data points themselves. This helps with comprehending issues and understanding how one concept can interact and affect others such as the relationships between the concepts of humankind, the natural world, and the artificial world.

This paper addresses the issue of how to represent data in a way that shows relationships between concepts in three dimensions and can be viewed for different purposes. It provides a procedure for mapping data within a 3D environment using a Systems-of-System Engineering (SoSE) approach. This is called 'SoSEdata3d'. Each concept is assessed and quantified using relevant criteria and the results are represented in three dimensions using a combination of a data and metadata defined global model framework, a multi level matrix using the principles of fuzzy logic and granulation using interval and finite mathematics and a series of digital fingerprinting and relativity transformations of the data.

This produces a capstone system which displays the dynamic relationships between concepts in 3D. SoSEdata3d allows the user to see visually the links or connections between nodes which represent concepts and to analyse them dynamically by rotating the resulting conceptual relationship model.

Keywords—3D, Capstone Modeling, Digital Fingerprinting, Peer Transformations, Nth Dimensional Data, Mapping, Fuzzy Logic.

I. Introduction

SoSEdata3d is a procedure for establishing a framework where nodes and / or graphs are specifically plotted in 3d space by using dynamic plots, with coordinates representing clusters, metadata, and observations in a relative manner.

This process has been applied to social networking analysis, decision making mapping, surveys analysis, weather modelling, movement sensor, autonomous control, role play games master character mapping, and text mining analysis to gain a better understanding of the core issues found in the data and to see if one is varied in its consequences what effects has this on other issues such as changing conditions that affect decisions taken.

II. Background

A. The problem on hand

One of the challenges in this emerging data analysis age, is the modelling and visualization of complex data relationships that show the relationships between concepts, ideas, beliefs and similar in three dimensions and enables knowledge to be extracted and used for different purposes. This especially becomes apparent when you are modelling scenarios that include;

- parent-child relationships
- observations across disparate data sets
- capstone models.

Unfortunately, traditional focused science and mathematical approaches fall short because they are designed based on a range of premises which fail to understand the possibility that items can be;

- dynamic or have a dynamic relationship
- symbiotic or relativity focused
- part of a cause and effect chain

In short, many of our forefathers saw a colourful world through black and white lens. Over the last 50 years, we have come to a better understanding of just how much of the world either integrates or affects another aspect. The newer understandings of fuzzy logic, chaos theory, and the uncertainty effect, has opened up better opportunities to gain a greater understanding of how things work.

The current explosion of Data around the world, combined with the availability of open source software, is breaking down the barriers and silo style mentality towards data analysis. People, companies, and nearly every government is now starting to see data, and associated opportunities, in a totally different way.

The problem at hand, is how to represent data from a cause and effect scenario which has multiple sources, that exists at different levels, and has dynamic properties. Hence, any capstone model that includes the integration of a parent-child model hierarchy needs to be self-adjusting and dynamic across the entire model. This is the problem that I chose to try and solve.

I decided on using a detailed framework to generate reference points; a framework where graphs are positioned at each data coordination. The end result is the SoSEData3D process, an integrated suite of three approaches. These three approaches - Nolan's Matrix, Nolan's Peer Transformation and Nolan's Digital Hash – are described below.

B. The Software

When first developed in the mid 90's, the prototypes were originally developed in Microsoft Access and Excel. They where then written in Basic to be tested for accuracy. Around 2010, the calculations and transformations were programmed in R, and the results where being examined in Rattle. In 2014 the whole process is being developed in R, using a number of packages and the results are visualized using the RGL package. At the time this paper was being written, a web tool in java and d3 is in development.

C. What is not a system of systems?

This brings me to the focus of the paper Systems of Systems. Wikipedia cites "**System-of-Systems Engineering** (**SoSE**) as a set of developing processes, tools, and methods for designing, re-designing and deploying solutions to System-of-Systems challenges". My main interest here is really focusing on how to model the data at specific points contained within a nth dimensional framework. Different terminology calls the same approach a Model of Models.

To me a system is an integration of any three or more facets (physical items, attributes, concepts), which produce a repeatable result through cause and effect. A System of Systems is any capstone system, which is the result of the combination of three or more systems, which produce a repeatable result through cause and effect. Such an open definition, allows me to see Systems of Systems, both natural and unnatural every where. What I am interested is the parent / child integration symbiotic relationship, which gains in complexity as it builds in the number of feeds and outcomes. Hence, I am interested in how mathematics can be used to model the processes.

When looking at systems, it is importing to map their influences on each other, and there cause and effect. Often people shout out about uncertainty and of course randomness.

I prefer the term pseudo randomness, as I do not believe that randomness occurs in nature, and that every thing is a product of cause and effect. Pseudo randomness is really any cause and effect which happens under the threshold that can be observed. For example, when you roll a dice, the resulting distribution is normal. However, if you control for drop height, starting position, etc, and you repeat the process, with the same dice, then you get more of an unnatural distribution.

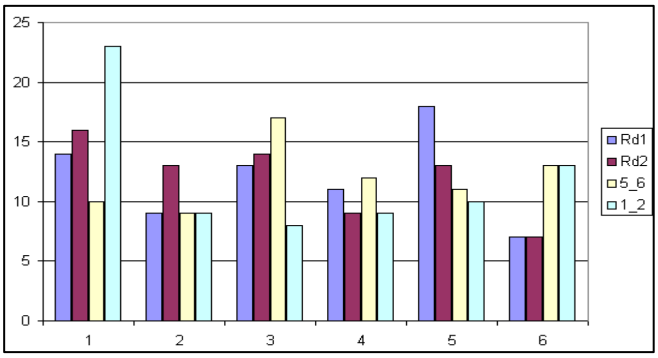

Fig. 1. Bar chart of dice throwing expriment results.

The bar graph (figure 1) shows the number of times a six sided dice produced a result. The experiment was to roll a 6 sided dice 72 times. Rd1 means the dice was released from any starting position at a set height, where every possible top and forward facing combination was repeated 3 times. Rd2 means that the dice was released from any height and starting position. 5_6 means the dice was released down a shoot at a set height, where the number 5 was pointing up and the 6 was facing forward. 1_2 is a repeat of 5_6 but with 1 facing the top and 2 facing forwards. As can be observed the two non fixed starting positions have a more diverse distribution. The two fixed starting locations have a much tighter distribution. The purpose of this experiment is to demonstrate, that what is believed to be random can become more predictable when a number of controlling factors can be applied.

While it is not possible to control for all the possible factors in cause and effect, allowances can be made to have an input of uncertainty, or black swan events in the capstone system, but being listed as unknown.

One of the advantages of this type of approach is how you can take the complexity and represent it in a simpler way. Complex systems are often very dynamic, have lots of interdependencies and or symbiotic relationships. While consolidation or compression can remove some of the detail, the number of chain observations in sequence, can be more telling, and make up for the loss in detail. In this type of capstone model, the distance between the data points, can be as equally informative as the data points themselves.

III. DIFFERENT FACETS

A. Nolan's Matrix

A Nolan's Matrix is a multi level matrix originally launched in 1995. It comprises of a number of different overlays. One level is pure data, where each cell has a number of observations which are encoded by a distance preserving digital hash. Another level applies fuzzy logic and granulation, within a combination of interval and finite mathematics, to make this matrix dynamic. Each observation remains relative to all others in the matrix, and is represented in a proportional manner. This gives the process the dynamic and self adjusting properties required. The final level of the matrix is the staging and graphing phase, which gives each observation a set of coordinates in nth dimensional space, that allows for comparison and filtering when applied to a framework.

B. Nolan's peer transformation

$$X = \frac{(x - xmin)}{(xmax - xmin) * n}$$

Fig. 2. Formula for Nolan Peer Relativity Transformation

Nolan's peer transformation is a rescaling transformation between set upper and lower limits, which redistributes the variables into a finite number of classes. This bin allocation process transforms the observations into like bins, making them relative to other observations. When the bins are transformed within a metadata grouping classification and chained together in a sequence, the result is a profile of an entity across all relevant characteristics. This is similar to a DNA sequence, but rather than 4 nodes, there can be many more. This rescaler can also be used to cut across different maximum and minimum values per variable, to give relative proportions. This type of measure is not affected by skewness.

C. Nolan's Digital Hash

$$Score = \sum_{i=1}^{n} X_i * Cos(Y*i)$$

Fig. 3. Formula for Nolan's Digital Hash.

One of the challenges in this emerging data analysis age, is the modeling and visualization of complex data relationships that show the relationships between concepts, ideas, beliefs and similar in three dimensions and enables knowledge to be extracted and used for different purposes.

This is a distance preserving discrete cosine transformation used to establish a digital fingerprint. A digital fingerprint is an index value which can only be calculated by a unique combination of observations, either within the same variable, or across a number of variables. While it is not a perfect digital hash, it is reliable within the limitations defined by the matrix. To date, the hash has worked across over 200 observations, where the traditional belief is that a hash can work across no more than 14 variables.

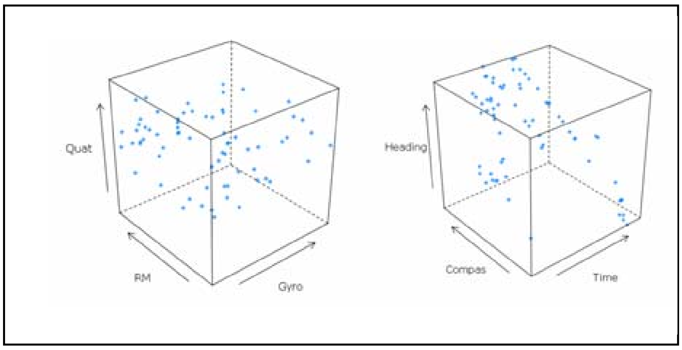

Fig. 4. 3d plots of multiple sensor data compressed into 3 axis.

Figure 4, shows data from a sensor board, which provides rotation, accretion, compass, etc. Each data point on the left graph represents 14 observations on three sensors compressed into the single data point. The right graph is an example of each data point represents five observations in two variables, transformed to a time sequence variable.

D. The Framework

The frame work is essential to making this process work. The framework provides the structure for the mapping of data points, the context for the results, and the comparison across of graphs. The capstone framework is quasi fractural also, and the capstone framework has a value of 1, where any child frames have a proportion of the value relative to other child framework at the same level. When a child framework becomes a parent framework, its value becomes 1, for and child frameworks below to be re-portioned, however it maintains its original proportion, relative to its level or above.

IV. PUTTING IT ALL TOGETHER

The capstone approach is to build an overarching model, which is greater than the sum of all its parts. Original data is transformed using both techniques, in different sequences and some times repeated, which are dependent on the piece of analysis being undertaken. The sequence can be used to follow either a crisp logic or fuzzy logic transformation, depending on the order. The digital hash gives a very specific result, while the peer binning gives a grouped result. By grouping the hashes, you can create an index which represents any number of observations. The end result a way of graphing any number of observations on a single graph axis. By Hashing the groups, you can create an index of grouped results, where aligns specific memberships of sub populations in nearest neighbor combinations. Once the results are incorporated within a framework and 3d spatial coordinates are applied, the dynamic modelling recalculates the relationships of all the hierarchy parent and child nodes and models. The framework generates a quasi fractal environment where any observation can be changed, inserted or deleted at any level. The adjustment has a cascading effect upon the entire model, and depending on the significance and sensitivity threshold of its contribution to the model, can elicit a noticeable change as all elements readjust.

Once the final model of data, framework and coordinates is established, clusters of data are defined by the sub-systems they belong to and are used to generate graphs. While the values within the graphs are specific to that sub system, the axis and categories are general across the rest of the whole model. Some graphs not only plot data observations, but also plot metadata as well. When you inspect a number of the graphs, not only can you examine the data and its relationship within that system, you can also compare the different characteristics of the metadata. This allows cross visual comparison within and between the different sub systems.

While the output of this process can be incorporated into known modeling techniques and other analytical tools such as a random forest, Social Network Analysis, Self Organising Maps, decision trees, etc, the final stage of the process is the visual exploration and the ‘what if’ analysis capability of the interactive viewing engine. The interactive viewing engine has the ability to rotate in pitch – roll – yaw directions, select multiple nodes, filter results and zoom in and out to inspect positions and graphs.

So far, this capstone approach has been used with Social Networking Analysis, Decision Making Mapping, Survey Analysis, Weather Modelling, Movement Sensor, Autonomous Control, Role Play Games Master Character Mapping, and Text Mining.

V. Expand on the Examples

The follow examples are drawn from some of the different applications where the procedure has been tested in, or is in current development.

A. Decision Mapping

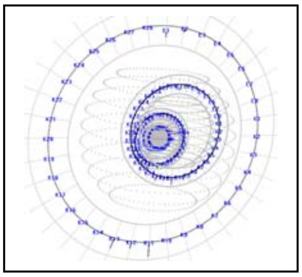

Fig. 5. Hierachial chart of parent – child relationships in 3d space.

In this example, I have modeled a decision being represented by a series of 3d spheres (see figure 5). The halo around the outside represents different types of metadata, which covers both emotions and different knowledge attributes. Each sphere is positioned relative to each other, along an array which represents a specific type of subject use in that decision. The distance from the centre represents the importance, and the size represents the number of sub-systems included at further child levels. As the model is quasi fractal, each globe is using the same framework. The viewing platform allows for the model to be rotated in 3 axes, and can be magnified to examine the lowest levels.

B. Weather Modeling

Weather is key factor in our daily decision making that affects us in the areas of our leisure activities, dress and fashion coverings, transportation and communications; and impacts on shopping, working, health, and even survival. By definition the weather is the current and predicted near future state of the atmosphere for a particular location. In practicable terms, weather is the term given to describe a series of environmental systems characterized by variables and measurements, which can be categorized according to their impact on how we conduct our human activities. For instance, the level of temperature, the speed and direction of the wind, the amount of light, the chance of getting wet, and the amount of solar radiation we receive at any time, influences our behaviours. Such as, how we dress, which mode of transportation to use, what time we undertake certain activities, and the risk to our completion of our activities in a safe and desired outcome.

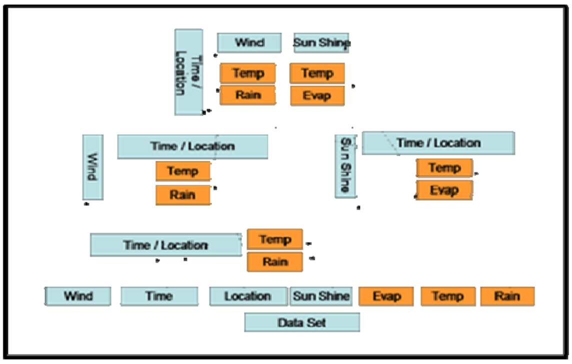

Fig. 6. Concept map of the weather dataset analysis.

By using weather data as an example of a systems of systems approach, factors like wind, air pressure, sun shine hours, cloud cover, rain and evaporation, can be combined with other datasets from other observations like different locations, wave height, seasonal variations, sun solar effects and gravity. Hence, a very complex model can be developed to analyze trends in weather events

The weather dataset comes from data provided from the Australian Bureau of Metrology

C. Position Sensor data

With the growing usage of sensor data for sporting, model planes, health, and recreational industry, data received from a onboard sensor chip which transmits a Bluetooth signal, was attached to a robot following a set course, a remote control quad drone, to various pieces of sporting equipment, and a push bike. The purpose of this model was to track the variations of movement and quickly process it, to send back instructions, or feed back.

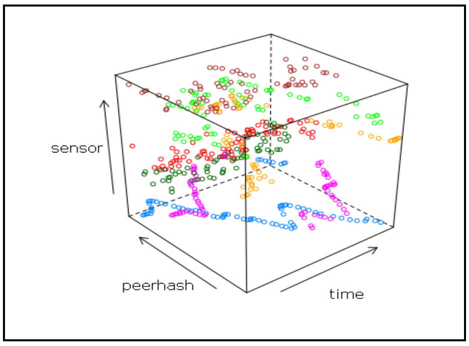

Fig. 7. 26 motion sensors variables transformed by hash into 7.

Figure 7 is an example of a robot following a set loop track, and each sensor is processed with both peer transformation and digital hash. The problem was how the sensor data dealt with the compression, and which sensors worked better in partnership with each other for the best result. The intention at the time of the writing of this paper, is to feed back flight instructions to a quad drone, to follow a set path.

D. 3d Dendrogram

Neural Nets, Decision Trees, Random Forests, etc are all useful tools, but they fail in visual comparisons because they do not have a standardized framework and scale to compare against each other. The image below is an example of a 3d dendrogram, where the distance and bearing between node points, represents the topic and strength of relationship. The framework is based on a standardized subject classification system.

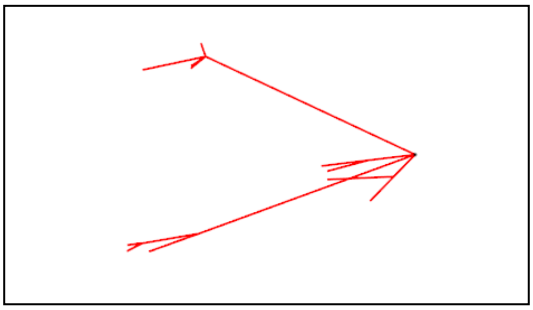

Fig. 8. 3d dendrogram of subjects used in decision making.

Represented in Figure 8, are four models which are indicated at each branch point; one model has a child and grand child hierarchy, on the lower left branch. As can be seen in the example, there are similar subject areas represented in the various parent and child models. The sum of the capstone model equals one, and each model's contribution is proportional to all other models. Each sub model is proportional to its parent, as if its parent was one also, and relative to all other models in the chain and at the same level.

E. Lotto Gambling

Lotto games are a series of balls mixed up by some form of mechanical process and then drawn out in an order, which people try to guess and gamble on. There is a consideration that it is total random, and can not be predicted. At first glance it is easy to believe, but when you take a scientific approach, you can see that it is not true.

For instance, change the perspective on the output of the model. Rather than trying to pick the numbers which are drawn each week, it is far easier to pick the numbers which have less of a chance of being picked. So rather than having to pick 7 or 9 out of 45 numbers, you design your model to locate numbers with less chance to be drawn. Then you remove them from the population. This is achieved by breaking down various aspects of the Lotto process into its different sub systems, and building a model on each one. As well as building a model on the history of numbers drawn.

At this point, we get the discussion that every number has an equal chance to be drawn out at each draw time. However when an historical analysis model is undertaken on the OZLotto game in Australia, patterns start to appear. For instance, over a 18 month sample of draws, 6 to 7 of the 9 numbers drawn in the previous 11 draws, appeared 83% of the time. Even running a single model of only choosing numbers which have been drawn in the 11 week period, generally cuts the population from 45 numbers to 31 numbers. In 68% of the time, the 5 or 6 of the winning numbers were present in the reduced population.

The purpose of this model is not to win OZLotto. It is required to have at least five numbers present in the cut down population. Of course to have a model that aimed to actually win money would require a different model for the spread betting options. Of course on a single model operation, it would take more money to place the bets, than the average return on investment - unless of course you are extremely lucky.

The process then went on to explore other possible models. For instance, the ink on the surface area of the ball, the starting location of the balls, the time difference between combinations of numbers being drawn, etc. Seven different models were consider viable, and were input into the capstone model. One of the interesting outputs of the model was not that patterns were detected, but that the patterns detected occurred in a narrow time frame. In some cases, if you were able to predict the first two balls drawn, and place a bet on that sequence, you increased your winnings and reduce your investment by half.

F. Games and simulation

This is under current development. The procedure here is applied to every character in a gaming environment. Players would work in teams, and each team is represent by a globe, and the meta data on each player contains their health, equipment, skills, healing ability, intelligence, morale, and strength. The team has some extra metadata in wealth, route through the environment, etc. The game has an optimum pathway, and looks at the deviation from that path. Due to the nature of the gaming capstone model, all characters, pathways and scenarios are plotted in the matrix, and then the language used, the number of clues, and the complexity of the problems solved are all adjustable from the gaming model.

The idea of this approach is to design a game which is self adjusting to adapt to a skill set slightly higher than the players current skills set, to make the game a challenge, yet still good playable experience.

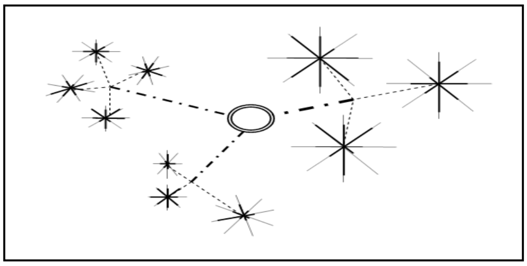

Fig. 9. Example of a figure caption. *(figure caption)*

The above diagram demonstrates a series of star graphs, where the position of each player in their teams, with skills and game attributes as metadata.

G. Network detection in law enforcement, non compliance, marketing and social networking analysis.

Network detection and mapping of all the connections and relationships between entities, is a useful tool in any analyst toolbox. In general, most linkage mapping occurs in 2 dimensional, and rarely it is representational of the strength of the connection, but where the data points are plotted for visual exploration than a distance measure. While this is highly useful for reading a chart, it is not so useful for representing networks and sub-networks.

By simply identifying targets and then looking at who is connected to them, by either address, relationship, business partners, social memberships, etc, it becomes possible by assigning a simple degrees of separation index, and using a frequency count percentage, risk rating that level of subpopulation, and number of targets represented in that level of subpopulation. Each entity is clustered into a network, and then all networks which are treated the same way as a system, are mapped in relationship to each other in a 3d scatter plot.

The data derived from the various networks can then be imported into other software packages for analysis. Considering the amazing amount of open source data which is available and the mash-up competitions, there is lots of scope for exploring different systems and using capstone modeling.

In Australia, there is the Australian Bureau of Statistics, the NSW Bureau of Crime Statistics and Research, and the Bureau of Metrology. Each of these government agencies produces data that is geocoded to a location zone. Simply by adding in a grouping code, clustering on the data, and recoding each data stream into equal number of bins with an assigned value, it becomes possible to map these different variables. These can then be mapped and even digital hashed to explore similarities and networks.

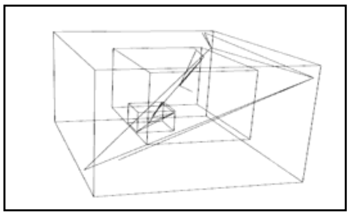

Fig. 10. Sub networks in a proportional relationship.

Figure 10 represents a diagram of a number of sub networks that sit proportionally against all others in the network. Each box represents a sub population in a parent / child hierarchy.

VI. In Conclusion

In conclusion, I have found that there is lots of opportunity to explore new techniques in complex systems analysis, and capstone system modeling. While the SoSEdata3d is still being developed and applied to different scenarios, it has shown a new way to visually explore data, as well as adding value as a transformer tool, to feed data into other tools for analysis.

VII. References

[1] A. Nolan, Getting underneath the decision making process, In Modsim 95 proceedings, Modelling and Simulation Association, Newcastle, 1995.

[2] A. Nolan, Modelling and Applications of Multi-Dimensional Interval Data to Artificial Life, Control Systems, Decision Making, In Proc. Fourth World Automation Congress, pp. 345-360, Hawaii, USA 2000.

[3] http://cran.r-project.org/

[4] Nolan, A. “An Action Philosophy Model” *Alar* 1(1) (July1996).

[5] Nolan A.WAC 2000 SAHIS: Simulated Artificial Hybrid Intelligent System

[6] Nolan, A. Nolan, E. “Using Nth Dimensional Fuzzy Logic for Complexity Analysis, Decision Analysis and Autonomous Control Systems” FORGING NEW FRONTIERS 40th of Fuzzy Pioneers (1965-2005)