Comprehensive Exam Questions for Mr. Dieudonne Ouedraogo

(Written Part: January 8, 2018 to January 22, 2018)

General Rules for the Written Comprehensive Exam:

- Answer all the questions in as much detail as possible. Use graphs/figures if needed. Provide citations from the literature where appropriate.
- You should give the best answers you can to all of the questions with the limited time you have available. Do NOT neglect some questions for the sake of giving perfect answers to some others.
- You are permitted to refer to any printed or electronic published materials, including, but not limited to, those on the reading list in your learning contract and those explicitly mentioned in the questions.
- You are permitted to use computers or other calculation devices as necessary limited only to the restrictions in software use as stated by specific problems.
- You may consult with members of your dissertation committee either for clarification on problems or for other reasons.
- You may NOT consult with any person not on your committee.
- Submit your typewritten answers by 5pm EST on January 22, 2018. Submission should be made via email to all of your committee members.

Comprehensive Exam Questions for Dieudonne Ouedraogo

By Dr. Hiroki Sayama

Question S-1. (Literature Review on Network Data Analysis)

Write a comprehensive review of recent research on network data analysis, with a particular focus on how it has been developed and applied to the research areas of your interest (e.g., consumer behavior analysis, healthcare, supply chain, finance, etc.). Discuss what has been accomplished, what is not, and then identify opportunities of further investigations, particularly in the context of your own research. Make sure to include references in your review. (Imagine that, for this question, you are writing a literature review chapter of your dissertation.)

Question S-2. (Technical Skills of Network Modeling and Analysis)

Download the Amazon ratings temporal network data from the following site:
http://konect.uni-koblenz.de/networks/amazon-ratings

Then do the following:

1. Import the data set using any network analysis/data analysis tools of your preference, and conduct some basic explorative analysis to obtain a "big picture" of how this data set "looks like." This can be done in various ways, including (but not limited to) descriptive statistics, (static) network analysis, data/network visualization, etc.
2. Identify "interesting" users/user groups using some justifiable criteria of your choice, and then study their temporal behaviors with regard to the products they rate and the scores they give to those products (and/or any other relevant behavioral traits). Try to obtain some novel discoveries that were not reported before.
3. Based on the results of the above steps, write a report of your data analytics work and findings. Imagine that you are writing this report as a professional recommendation to the Amazon executives.

Written Questions Submitted by Dr. Hal Lewis for Comprehensive Exam of Mr. Dieudonne Ouedraogo

Note: For this following questions, you are not permitted to use any special purpose software such as Fuzzy Logic Toolbox, special purpose entropy software, special purpose optimization software and the like. You may use general purpose software, including spreadsheets or general purpose programming languages.

Question L-1 (Fuzzy Models):

a.) Describe in detail your design for a Mamdani-style fuzzy inference system for a two-input/one-output application of interest to you (or it could be for an abstract application). Show at least two examples of numerical calculations (for given input values, determining what the output would be) showing the details for each step in your calculations.

b.) Describe in detail your design for a Fuzzy-singleton-style fuzzy inference system (also called zeroth-order Sugeno model) again for a two-input/one-output application of your choice (could be the same application as part a or a different one). Again show at least two examples of numerical calculations showing the details for each step in your calculations.

c.) Do the same as parts a and b, but this time for a TSK-style fuzzy inference system (also known as Takagi-style or as a first order Sugeno model).

d.) Describe in just general terms what a second order Sugeno model would look like.

e.) Describe any aspects of your research that could make use of fuzzy models of these types.

Question L-2 (Information Theory and Generalized Information Theory):

a.) Describe in your own words the conceptual justification for the maximum entropy principle.

b.) As a very simple example of the maximum entropy principle, consider the following situation. You are considering a particular two-dimensional cross section of a three-dimensional object, and you break up this cross section into just nine blocks in a 3x3 arrangement. You wish to come up with the best estimates for the amount of mass in each of the nine blocks based on information you gain from passing five rays through the cross section. Each ray provides an indication of the total mass to the three blocks it passes through as illustrated in the diagram below. For example, the sum of the masses in blocks 1, 2, and 3 is 800g. Find your estimate (to the nearest gram is sufficient) for the nine blocks in the cross section. Describe how you calculated these values.

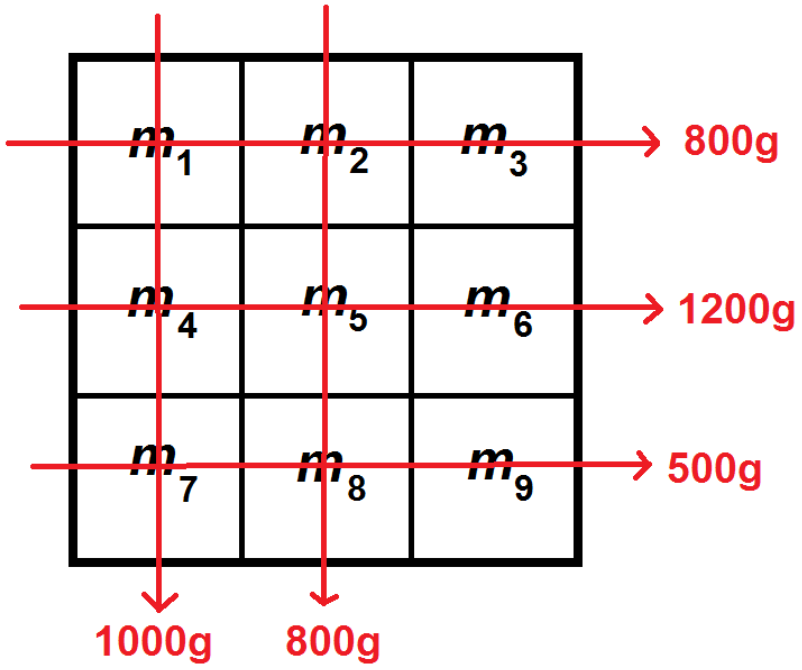

Diagram of cross section for part b of Question L-2

c.) According to Klir's generalized information theory, describe one example of another way you could have measured the total uncertainty for the situation described in part b. How would the principle of maximum uncertainty have worked in that case? Come up with the best estimates you can for the mass of each block based on maximizing the alternative form of uncertainty measurement you chose. How does it compare to your results in part b?

Question L-3 (Hybrid Machine Intelligence):

a.) Describe a hybrid of one of more traditional AI methodologies (e.g. heuristic tree search, logic programming, expert systems) combined with one or more soft computing methodologies (fuzzy logic, artificial neural networks, evolutionary computing) that could be used to solve a practical problem.

b.) How might this level of hybrid machine intelligence play a role in your research?

Questions from Prof. Daehan Won

Section 1: Course works related problems

W1. (SSIE 605, Applied Multivariate Data Analysis): Data Analysis

You can answer the following questions as True/False with proper evidences (supports)

i. When variables are uncorrelated, are they independent as well? If it is false, provide a counter example
ii. When *(Y,X)* has a bivariate normal distribution, does *X* has a normal distribution?
iii. If the data come from a multivariate normal distribution, then will the Chi Square q-q plot show an upward curve?
iv. If a 95% confidence interval is stated as $-1.4 < \mu_1 - \mu_2 < -.2$, then can we state with 95% confidence that μ_1 is somewhere between .2 and 1.4 units smaller than μ_2?

W2. (SSIE 660, Stochastic Systems): Probability and Statistics

One of the main research in the complex system is to investigate the structural and intrinsic characteristics. Weibull distribution, which is one of the continuous probability distribution, has been widely used to capture such intrinsic characteristics of the systems. For example, it has been introduced to measure the life-time of the component as well as degree distribution fitting to characterize the evolutionary process of the systems.
Also, Weibull distribution is known as a generalization of exponential distribution, which is one of the popular model to describe the life time in reliability engineering. Solve the following questions about Weibull distribution.

(i) Assume that you have multiple i.i.d. Xs following exponential distribution with parameter α. How can we estimate the α?
(ii) Once you obtain the α from (1), show that $X^{\frac{1}{\beta}}$ is a Weibull random variable with a shape parameter β using α.
(iii) Conversely, you are given multiple i.i.d. Ys with parameter α (scale parameter) and shape parameter β. How can we estimate α from Ys? Is it possible to estimate β? If yes, show the detail process.

W3. (SSIE 500, Computational Tools and SSIE 501 Introduction to Systems Science): Information Theory of Complex Systems

You have 13 almost identical balls. The only difference between them is that one of them have a slightly deviating weight, it could be heavier, or lighter. You are tasked to find which of the balls that deviates, with only the means of a balance scale (that tips to the side with the heaviest load, or remains in balance if the two sides are of equal weight) and three measurements

i. How large is the entropy of the system? How large is the entropy of which of the 13 balls that deviates?
ii. Assuming ideal measurements, how many measurements would you need at least to find the deviating one?
iii. Find a procedure to identify the deviating ball using a balance scale and only three measurements. You need only to describe the procedure assuming the worst case outcome for each measurement (the most probable outcome), i.e. not all possible branches of measurements.

Section 2: Future Research's related problems

W4. (Network Science) Let consider a star network, where a single node is connected to all other $N-1$ nodes. The rest of the nodes have degree one, and are connected only to the superstar at the center. Assume that $N \gg 1$

i. What is P_k, the degree distribution of this graph?
ii. What is $\langle k \rangle$, the average degree of the graph?
iii. What is the clustering coefficient C?
iv. What is the probability q_k that moving along a randomly picked edge, we find at its end a node with degree k?
v. Calculate the degree correlation coefficient for this network.

W5. Describe the academic problem in the area of combining network science and data mining (e.g., one representative problem is a link prediction). You should elaborate it in detail including required references. (Note: summarize everything up to 3 pages)