Examiners' report 2009

2910325 Data compression – Zone B

Important notice

From 2010 onwards the CIS325 examination paper will consist of just THREE questions. Candidates should attempt all three questions on the paper.

General remarks

CIS325 examination papers cover typically the most important topics of the course material. The questions may be different from year to year, but they can be classified as three types. The first type of questions are bookwork, asking you to explain a term or concept. The second type of questions are similar to what you have seen in the exercises or what you have done in the coursework, asking you to engage in certain activities; for example, to describe a technique, method, known algorithm, to analyse an output, or to explore certain characteristics. The third type of questions are unseen, allowing outstanding performance to shine.

As an examination candidate, you are expected to have fully understood the issues and mastered the main techniques learnt from the course. Excellent candidates are expected to have also developed certain analytical abilities in problem solving. When revising for the examination, you should focus on: bookwork, exercises, coursework and questions in the past examination papers, **in that order**. You should not try to memorise what you have seen but focus on understanding what you have done and why you can or cannot do it in a certain way.

Your grade depends crucially on how successfully you can demonstrate the required knowledge and skills in the examination. What you write on the script book is utterly important. It is hence essential for you to be able to present the solutions in the best possible way. For example, the answers on your scripts should be written in logical and coherent steps. You should show the entire process of deriving a solution, not a final result alone. The corresponding question numbers (e.g. Question 1, Question 3) and part numbers (e.g. (a), (b), ii., iii.) should be clearly marked in your answers. The solution to each question should begin on a new page if possible. The handwriting should be easy to read. You may use pens of any natural dark colour, preferably black or blue, but never red or green which are reserved for marking your examination scripts. It is a good practice sometimes to leave a note for the marker to clarify or highlight things, especially when solutions to

some parts of a question are presented in different places. Most importantly, candidates need to read the questions **carefully**.

Good timing in an examination is critical for your success. You should therefore draft a schedule quickly at the beginning of the examination to allocate the time for each question. This is to avoid the situation that you spend too much time on one question and do not have enough time for another. A similar timing plan should be made to each sub-question or part of the question.

There were a total of five questions in this year's examination paper. Candidates were required to attempt three out of five. We discuss the questions one by one in the following paragraphs.

Specific comments on questions

Question 1

a) There are two sub-questions in this part. A good answer would consist of two sections to address each sub-question separately and adopt the exact section numbers in the question.

   i. The answer to the first sub-question should include the proof. It is quite straightforward if you can write the entropy formula for each of the two given sources. Reorganising the terms of one formula, you can see the quantity relationship between the two entropies.

   ii. The answer to this part is actually one instance of the problem in (i).

b) The answer to this part is straightforward. A flowchart to demonstrate the adaptive Huffman algorithm is all that is required.

c) This continues the above parts and asks you to demonstrate an example of the decoding using the adaptive Huffman algorithm. An easy way to answer this part of the question is to organise your answer into two sections. First, give the decoding results in terms of the *input*, *output*, *alphabet* and *tree structure* for each step. Make sure that you highlight the final result from the final statement. Secondly, compute the compression ratio for the given example.

Question 2

a) A good answer to this part of the question would follow the requirements and is organised into two sections: The first section includes a definition of the prefix code. The second one includes two examples; one is a prefix code and the other is a non-prefix code that is uniquely decodable. For each example, explain why it is for a prefix code, or for a non-prefix code but uniquely decodable. A common mistake made by some candidates this year is to have missed one or two required parts.

b) This part of the question requires an account of explanation to the given statement. An easy way to answer the question is to use an approach where a false hypothesis is first established as a negative dual of the original claim, and a series of correct arguments followed

it would then lead to a self-contradiction. This would allow a conclusion in which the self-contradiction can only be avoided if the hypothesis were rephrased in the original way. Note in such an approach, the final conclusion is crucial and cannot be missed. Some candidates lost marks by eliminating the conclusion in this part.

c) This is an unseen question. A good way to answer this part of the question is to first group the requirements in the question and organise your answer before writing. You may, for example, organise your answer into two sections. The first section introduces the ideas of *progressive image compression* (PIC), and discusses the main advantages of the PIC technique. The second section gives an example that shows how the matrix of data is rearranged in hierarchical sets for data transmission, and how the data is displayed at the receiving end.

Question 3

a) A good answer to this part of the question would consist of two sections. The first section is used to demonstrate how a canonical minimum variance Huffman code can be derived for the given sequence of symbols. The second is used for analysis of the optimisation aspect of the Huffman code. A critical step is to show the gap between the average length of the Huffman code and the entropy of the given source. Please do not forget to describe a source that allows an optimal Huffman code to be derived from it. A few candidates this year forgot to do so and unfortunately lost marks.

b) Again, this is an unseen question. A good approach for this part of the question would include two sections. The first section describes the main ideas of the *predictive encoding* and *decoding* separately. The second section provides an example to demonstrate a full analysis of the two sources in terms of the entropy. For example, if the entropy of the residual source is smaller than that of the original source, given the same average length of a code, there would be more room for compression on the residual data.

Question 4

This is an unseen question. There are four parts and you should answer them separately.

a) This part of the question is straightforward. A numerical answer can be derived easily from a little calculation.

b) A good answer to this part of the question is a histogram that contains distributions of two sources.

c) This part of the question requires a good knowledge of Shannon's information theory. A good answer to this part of the question would simply compute and compare the entropies of the two given sources based on their probability distributions. The idea is that whichever source has a bigger entropy value contains more information on average.

d) An easy way to answer this part of the question is to first derive the canonical minimum variance Huffman codes for each source and then compare the gap for each source between the average length of the Huffman code and the entropy. The Huffman code for the source

with a zero gap is the optimal code and the one with a smaller gap is closer to the optimal.

Question 5

a) There are two requirements in this part of the question. A good answer would naturally consist of two sections. The first section would be a flowchart for the required Arithmetic decoding algorithm. The second section would include a trace of the execution of the algorithm in a diagram or table using the example given.

b) A good answer to this part of the question would consist of two sections. The first section explains the principle and the advantages of the *Reflected Gray code* for coding the colours in greyscale images. The second section derives such a code for decimal number 11. Note that you should include all the work in deriving the code. Some candidates lost marks for having missed intermediate steps.

c) This part of the question asks for an explanation of the concept of a cartoon-like image. A good answer would first explain the definition and then give the example required in the second section. You also need to provide a 5 x 5 matrix of data in the example.

d) In a similar setting, this part of the question requires a clarification of the concept of *data rate*. An easy answer to this part would consist of two sections. The first section would describe the definition and theory, and the second section would include an example of data or a diagram, with a clear indication of what can be measured by *data rate*.

Summary

Good performance in the examination depends upon not only your sound knowledge covering all the required topics, but also your abilities of demonstrating your knowledge and analytical skills. The importance of understanding the questions in the examination cannot be emphasised more. Examinees are, again, advised to read the questions carefully. You should make sure that you fully understand for each question what is required and what parts or sub-questions are involved explicitly and implicitly. You are encouraged to take notes, if necessary, while reading or attempting the questions. Above all, you should be completely familiar with the course materials. Good candidates should prepare to solve problems in some unseen form by applying their knowledge gained from the course.