The Zebra Puzzle

Carlo Tomasi
Computer Science
Duke University

1 Introduction

In the class textbook, the Zebra puzzle is offered as an exercise for Section 1.1 on logic (Exercise 61, page 20). Here is the puzzle:

> Five men with different nationalities and with different jobs live in consecutive houses on a street. The houses are painted different colors. The men have different pets and have different favorite drinks. Determine who owns a zebra and whose favorite drink is mineral water (which is one of the favorite drinks) given these clues[1]:
>
> 1. The Englishman lives in the red house.
> 2. The Spaniard owns a dog.
> 3. The Japanese man is a painter.
> 4. The Italian drinks tea.
> 5. The Norwegian lives in the first house on the left.
> 6. The green house is on the right of the white one.
> 7. The photographer breeds snails.
> 8. The diplomat lives in the yellow house.
> 9. Milk is drunk in the middle house.
> 10. The owner of the green house drinks coffee.
> 11. The Norwegian's house is next to the blue one.
> 12. The violinist drinks orange juice.
> 13. The fox is in a house next to that of the physician.
> 14. The horse is in a house next to that of the diplomat.

The textbook also gives the following hint:

> Make a table where the rows represent the men and columns represent the colors of their houses, their jobs, their pets, and their favorite drinks and use logical reasoning to determine the correct entries in the table.

This is a fun puzzle to think about. However, it has to do with "logic" only marginally. The real difficulty is in knowing where to start, and in how to think about the problem.

When addressing real-life problems as computer scientists, we will be in this situation often: someone poses a problem (usually less well defined than this one), and we are asked to solve it. This is different from solving a homework assignment in a class. For the homework, we are either told what technique to apply, or we can make an

[1] I have numbered the clues for later reference.

educated guess by picking a solution method from the part of the textbook that the problem refers to. In real life, all we have is the problem itself, and whatever knowledge and experience we have accumulated over time. So how do we come up with a solution? We will use the zebra puzzle to explore some heuristics for thinking about complex problems.

2 How Hard Is the Problem?

Why is it useful to know how hard the problem is? Mostly because we want to know how many resources to throw at it: Do I just scribble a couple of lines on the back of an envelope? Do I need to write a computer program? If I do, will the program run in a reasonable amount of time?

For puzzles, understanding its complexity is usually a simple exercise of combinatorics, a set of counting techniques that we will explore in this course. The web page http://mathforum.org/library/drmath/view/55627.html suggests a solution to the zebra puzzle, and starts with the following observation:

> Of course one could just enumerate the $5^5 = 3125$ different possible answers, and scratch off all of those which didn't fit the clues, but this is the hard way.

First, this may not be as hard if done with a computer program: 3000 or so alternatives can be explored in microseconds. If the program is not too complex to write, we may be done more quickly this way. An advantage of a systematic approach like this is that we can also answer ancillary questions: is the solution unique? If not, how many other solutions are there?

Another advantage of a software solution is that if the code is simple we are more confident that our solution is correct. Going through the many steps in the "logical" solution given in the web page mentioned above is error prone, and writing a short program may be safer.

Thirdly, the solution on the web page goes through a chain of arguments of which here is a snippet (that version of the problem uses flowers instead of jobs, and Ukrainians instead of Italians):

> By No.7, the geranium grower owns snails. Thus the geranium grower does not own the dog, fox, horse, or zebra, and the snail owner does not grow roses, marigolds, lilies, or gardenias. Thus the geranium owner is not the Spaniard.

We can all follow arguments of this type, given the clues. The difficulty is not in verifying whether a logical implication is valid or not. The real trouble is in coming up with a sequence of arguments that leads us to a conclusion quickly. So in a sense the given solution is not a solution unless we are also told how the solver came up with the proper sequence of arguments.

In this regard, if we can write a program that solves this puzzle, perhaps we learn how to write programs to solve other puzzles, and maybe even programs that in some sense can be said to "think."

However, the problem is harder than exploring 3000 or so alternatives. The different possible answers are not 3125, but many more. To see this, let us do what is a useful first step in all cases: let us *visualize* the problem.

Drawings always help. We could draw five little Victorian houses with windows and chimneys, paint them in five different colors, and draw a gondolier for an Italian, and so forth. Needless to say, we want something quicker. The textbook hint to use tables is good: tables organize information in a concise way, and let us view the whole situation at a glance. This is a broad view of things that is very useful to us, just as it is mostly useless to computers, which in a sense can only look at one data item at a time.

The specifics of the hint, on the other hand, are misleading. Why should we put the men on the rows, and put jobs, colors, pets, and drinks on the columns? What makes men (or their nationalities) any different, in the abstract problem, from jobs or colors? This is always a key question when you make a table: what to put on rows and columns, and should the table have two dimensions (rows, columns), or more?

Think about this for a minute. I mean it: put this document down, and actually *do* think for a while about whether the nationality of a house occupant is to be treated any differently from the job he holds, or the pet he has, for the purpose of solving the zebra puzzle.

There is really no reason for this distinction. All we have are five houses, and five "attributes" for each house: its color, the nationality, job, and favorite drink of its occupant, and the pet that lives in it. Assigning a job and an occupant to a house also automatically assigns that job to that occupant. More importantly, if we were to use the table suggested in the hint, how would we use it to express the fact that *The green house is on the right of the white one* (clue 6)? The suggested table does not say where the houses are relative to each other, while clues 5, 9, 11, 13, 14 all refer to house position.

So instead we think of five columns (not rows, merely because houses are vertical), one per house, and of five attributes that we stack in each house. One possible assignment of attributes (not consistent with the clues) is captured by Table 1. This table captures also positional information, what is where and next to what, so this type of table is more likely to help.

Color	Blue	Green	Red	White	Yellow
Nationality	Englishman	Italian	Japanese	Norwegian	Spaniard
Job	Diplomat	Painter	Photographer	Physician	Violinist
Pet	Dog	Fox	Horse	Snails	Zebra
Drink	Coffee	Milk	Orange Juice	Tea	Mineral Water
House	1	2	3	4	5

Table 1: One possible assignment of color, nationality, job, pet, and drink to the five houses.

How many such assignments can we come up with? There are five colors, so there are five possible color assignments to house 1. For each of these, there are four remaining colors that can be assigned to house 2, for a total of 4×3 combinations. For each of these, there are three colors left for house 3, two for house 4 and one for house 5. The total number of color assignments to houses is therefore $5 \times 4 \times 3 \times 2 \times 1 = 5!$ (pronounced "five factorial"), that is, 120 assignments.

Suppose that we pick one of the 120 color assignments. For this assignment, we can again choose $5!$ assignments for the nationalities of the occupants, for a total of $5! \times 5! = (5!)^2$ different combinations of colors and nationalities. If we repeat this reasoning for jobs, pets, and drinks, we see that the total number of possible assignments is $(5!)^5$, not just 5^5. That is, we have 24,883,200,000 possible assignments (that's almost 25 billion).

This rules out any exhaustive solution by hand, but not necessarily by computer: With computer clocks running in the gigahertz range, a well written program could conceivably be done in seconds or minutes.

However, this counting exercise also shows that if we had a slightly modified problem with six homes and six categories of attributes (let's add favorite flowers, for instance) instead of five, then we would have $(6!)^6$ possible assignments, which is about 1.4×10^{17}. Even if it took a single machine clock cycle to check the consistency of one assignment with the given clues, a 10GHz machine would take $1.4 \times 10^{17}/10^{10} = 1.4 \times 10^7$ seconds, that is, almost 116 days (divide seconds by $60 \times 60 \times 24$ to obtain days), to check all assignments and find one (the only one?) that is compatible with the clues.

Thus, the zebra puzzle is an example of a problem whose exhaustive solution is feasible on a single computer as given, but does not scale to any bigger problem of the same type. If we cannot come up with anything better, and if all we need to do is to solve the zebra puzzle itself, we may take this approach (and perhaps be left with an aftertaste of inelegance). If we want to gain experience with solving similar puzzles, or solve bigger ones, we need to try harder.

3 Tables and Clues

Going through the clues, we realize that some of them are more directly applicable than the others. Specifically, clues 5 and 9 let us assign the Norwegian to the first house and milk to the third. Clue 11 can also be applied right after clue 5: since the Norwegian is in the first house and he is next to a blue house, the second house is blue. After this, things become harder, as no other clue seems to be directly applicable. How do we proceed further?

We could start by drawing the compulsory assignments so far, and leave everything else blank, as in Table 2.

We could then continue adding assignments, checking for consistency with the clues as we do this. This is a natural strategy: start with what you know for sure, and grow from there. However, we would get stuck soon with this: Say that we put the Spaniard in house 2. Then we must put the dog in that house as well, because of clue 2. Then let us put the Japanese in house 3 (and the painter in house 3 because of clue 3), and the Italian in house 4 (and tea in house