

# Artificial Intelligence Homework

Information and Computation Science Department  
Peking University

This chapter-by-chapter homework is a syllabus that is subject to lecture-per-week-per according to the AI course.

<sup>+</sup> is required for graduate students.

## 1 Introduction

**1.1** Define in your own words the following terms: (a) AI, (b) intelligent agent, (c) algorithm, (d) commonsense, (e) rationality.

**1.2** Design a pseudocode of the goal-based agent.

**1.3** The vacuum environments have all been deterministic. Discuss possible agent programs for each of the following stochastic versions:

- (1) Murphy's law: twenty-five percent of the time, the Suck action fails to clean the floor if it is dirty and deposits dirt onto the floor if the floor is clean. How is your agent program affected if the dirt sensor gives the wrong answer 10% of the time?
- (2) Small children: At each time step, each clean square has a 10% chance of becoming dirty. Can you come up with a rational agent design for this case?

**1.4** Read Turing's original paper on AI (Turing, 1950). In the paper, he discussed several objections to his proposed enterprise and his test for intelligence. Which objections still carry weight? Are his refutations valid? Can you think of new objections arising from developments since he wrote the paper? In the paper, he predicted that, by the year 2000, a computer will have a 30% chance of passing a five-minute Turing Test with an unskilled interrogator. What chance do you think a computer would have today? Finally, do you think the Turing test is adequate for understanding intelligence?

## 2 Search Algorithms I

**2.1** The **missionaries and cannibals** problem is usually stated as follows. Three missionaries and three cannibals are on one side of a river, along with a boat that can hold one or two people. Find a way to get everyone to the other side without ever leaving a group of missionaries in one place outnumbered by the cannibals in that place.

- (1) Formulate the problem precisely, making only those distinctions necessary to ensure a valid solution. Draw a diagram of the complete state space.
- (2) Implement and solve the problem optimally using an appropriate search algorithm by designing the pseudocode. Is it a good idea to check for repeated states?

**2.2** Which of the following are true and which are false? Explain your answers.

- (1) Depth-first search always expands at least as many nodes as  $A^*$  search with an admissible heuristic.
- (2)  $h(n) = 0$  is an admissible heuristic for the 8-puzzle.
- (3)  $A^*$  is of no use in robotics because percepts, states, and actions are continuous.
- (4) Breadth-first search is complete even if zero step costs are allowed.
- (5) Assume that a rook can move on a chessboard any number of squares in a straight line, vertically or horizontally, but cannot jump over other pieces. Manhattan distance is an admissible heuristic for the problem of moving the rook from square  $A$  to square  $B$  in the smallest number of moves.

**2.3** Write down the pseudocode for the  $A^*$  algorithm.

**2.4** Prove that if a heuristic is consistent, it must be admissible. Construct an admissible heuristic that is not consistent.

## 3 Search Algorithms II

**3.1** Give the name of the algorithm that results from each of the following special cases:

- (1) Local beam search with  $k = 1$ .
- (2) Local beam search with one initial state and no limit on the number of states retained.
- (3) Simulated annealing with  $T = 0$  at all times (and omitting the termination test).
- (4) Simulated annealing with  $T = \infty$  at all times.

**3.2** Develop a formal proof of correctness for alpha-beta pruning. To do this, consider the situation of the alpha-beta pruning tree. The question is whether to prune node  $n_j$ , which is a max node and a descendant of node  $n_1$ . The basic idea is to prune it if and only if the minimax value of  $n_1$  can be shown to be independent of the value of  $n_j$ .

- (1) Node  $n_1$  takes on the minimum value among its children:  $n_1 = \min(n_2, n_{21}, \dots, n_{2b_2})$ . Find a similar expression for  $n_2$  as a child of  $n_1$  and hence an expression for  $n_1$  in terms of  $n_j$ .
- (2) Let  $l_i$  be the minimum (or maximum) value of the nodes to the left of node  $n_i$  at depth  $i$ , whose minimax value is already known. Similarly, let  $r_i$  be the minimum (or maximum) value of the unexplored nodes to the right of  $n_i$  at depth  $i$ . Rewrite your expression for  $n_1$  in terms of the  $l_i$  and  $r_i$  values.

- (3) Now reformulates the expression to show that to affect  $n_1$ ,  $n_j$  must not exceed a certain bound derived from the  $l_i$  values.
- (4) Repeat the process for the case where  $n_j$  is a min-node.

**3.3** Write down the pseudocode of Monte Carlo tree search (MCTS) algorithm as refinement as possible (more details than the ones in the AI slides).

**3.4<sup>+</sup>** Read the paper: *Silver, D, et. al., Mastering the game of Go with deep neural network and tree search, Nature, 529.7587, 2016, 484-489*. Can the methods of AlphaGo be transferred to Chinese Chess and Mahjong? Why?

(May refer: *Silver, D, et. al., A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science 07 Dec 2018: Vol. 362, Issue 6419, pp. 1140-1144.*)

## 4 Constraint Satisfaction Problems

**4.1** Write down the pseudocode of the path consistency algorithm for CSP.

**4.2** Consider the problem of constructing (not solving) crossword puzzles as follows: Giving a rectangular grid (as part of the problem), fitting words into the grid, and specifying which squares are blank and which are shaded. Assume that a list of words (i.e., a dictionary) is provided and that the task is to fill in the blank squares by using any subset of the list. Formulate this problem precisely and write the pseudocodes in two ways:

- (1) As a general search problem. Choose an appropriate search algorithm and specify a heuristic function. Is it better to fill in the blanks one letter at a time or one word at a time?
- (2) As a constraint satisfaction problem. Should the variables be words or letters?

**4.3** Show how a single ternary constraint such as " $A + B = C$ " can be turned into three binary constraints by using an auxiliary variable. You may assume finite domains. (*Hint*: Consider a new variable that takes on values that are pairs of other values, and consider constraints such as " $X$  is the first element of the pair  $Y$  .") Next, show how constraints with more than three variables can be treated similarly. Finally, show how unary constraints can be eliminated by altering the domains of variables. This completes the demonstration that any CSP can be transformed into a CSP with only binary constraints.

**4.4** Consider the following logic puzzle: In five houses, each with a different color, live five persons of different nationalities, each of whom prefers a different brand of candy, a different drink, and a different pet. Given the following facts, the questions to answer are "Where does the zebra live, and in which house do they drink water?"

The Englishman lives in the red house.

The Spaniard owns the dog.

The Norwegian lives in the first house on the left.

The green house is immediately to the right of the ivory house.

The man who eats Hershey bars lives in the house next to the man with the fox.

Kit Kats are eaten in the yellow house.

The Norwegian lives next to the blue house.

The Smarties eater owns snails.  
 The Snickers eater drinks orange juice.  
 The Ukrainian drinks tea.  
 The Japanese eat Milky Ways.  
 Kit Kats are eaten in a house next to the house where the horse is kept.  
 Coffee is drunk in the green house.  
 Milk is drunk in the middle house.

Discuss at least 2 different representations of this problem as a CSP. Why would one prefer one representation over another?

## 5 Logical Agents

**5.1** Consider the problem of deciding whether a propositional logic sentence is true in a given model.

- (1) Write a recursive algorithm  $\text{PL-TRUE?}(s, m)$  that returns *true* if and only if the sentence  $s$  is true in the model  $m$  (where  $m$  assigns a truth value for every symbol in  $s$ ). The algorithm should run in time linear in the size of the sentence.
- (2) Give three examples of sentences that can be determined to be true or false in a partial model that does not specify a truth value for some of the symbols.
- (3) Show that the truth value (if any) of a sentence in a partial model cannot be determined efficiently in general.

**5.2** This question considers representing satisfiability (SAT) problems as CSPs.

- (1) Draw the constraint graph corresponding to the SAT problem

$$(\neg X_1 \vee X_2) \wedge (\neg X_2 \vee X_3) \wedge \dots \wedge (\neg X_{n-1} \vee X_n)$$

for the particular case  $n = 5$ .

- (2) How many solutions are there for this general SAT problem as a function of  $n$ ?

**5.3** Consider a knowledge base containing just two sentences:  $P(a)$  and  $P(b)$ . Does this knowledge base entail  $\forall x P(x)$ ? Explain your answer in terms of models.

**5.4<sup>+</sup>** The question involves formalizing the properties of mathematical groups in FOL. Recall that a set is considered to be a group relative to a binary function  $f$  and an object  $e$  if and only if

- (1)  $f$  is associative;
- (2)  $e$  is an identity element for  $f$ , that is, for any  $x$ ,  $f(e, x) = f(x, e) = x$ ; and
- (3) every element has an inverse, that is, for any  $x$ , there is an  $i$  such that  $f(x, i) = f(i, x) = e$ .

Formalize these as sentences of FOL with two nonlogical symbols, a function symbol  $f$ , and a constant symbol  $e$ , and show using interpretations that the sentences logically entail the following property of groups:

For every  $x$  and  $y$ , there is a  $z$  such that  $f(x, z) = y$ .

Explain how your answer shows the value of  $z$  as a function of  $x$  and  $y$ .

## 6 Automated Reasoning

**6.1** Suppose you are given the following axioms:

- (1)  $0 \leq 3$ .
- (2)  $7 \leq 9$ .
- (3)  $\forall x \quad x \leq x$ .
- (4)  $\forall x \quad x \leq x + 0$ .
- (5)  $\forall x \quad x + 0 \leq x$ .
- (6)  $\forall x, y \quad x + y \leq y + x$ .
- (7)  $\forall w, x, y, z \quad w \leq y \wedge x \leq z \Rightarrow w + x \leq y + z$ .
- (8)  $\forall x, y, z \quad x \leq y \wedge y \leq z \Rightarrow x \leq z$ .

- (A) Give a backward-chaining proof of the sentence  $7 \leq 3 + 9$ . (Be sure, of course, to use only the axioms given here, not anything else you may know about arithmetic.) Show only the steps that lead to success, not the irrelevant steps.
- (B) Give a forward-chaining proof of the sentence  $7 \leq 3 + 9$ . Again, show only the steps that lead to success.

**6.2** Convert the following set of sentences to clausal form:

- $A \Leftrightarrow (B \vee E)$ ,
- $E \Rightarrow D$ ,
- $C \wedge F \Rightarrow \neg B$ ,
- $E \Rightarrow B$ ,
- $B \Rightarrow F$ ,
- $B \Rightarrow C$ .

Use resolution to prove the sentence  $\neg A \wedge \neg B$  from the clauses.

**6.3** Victor has been murdered, and Arthur, Bertram, and Carleton are the only suspects (meaning exactly one of them is the murderer). Arthur says that Bertram was the victim's friend, but that Carleton hated the victim. Bertram says that he was out of town the day of the murder, and besides, he did not even know the guy. Carleton says that he saw Arthur and Bertram with the victim just before the murder. You may assume that everyone – except possibly for the murderer – is telling the truth.

- (1) Use resolution to find the murderer. In other words, formalize the facts as a set of clauses, prove that there is a murderer, and extract his identity from the derivation.
- (2) Suppose we discover that we were wrong – we cannot assume that there was only a single murderer (there may have been a conspiracy). Show that in this case, the facts do not support anyone's guilt. In other words, each suspect presents a logical interpretation that supports all the facts but where that suspect is innocent and the other two are guilty.

**6.4<sup>+</sup>** Here are two sentences in the language of first-order logic:

- (A)  $\forall x \exists y (x \geq y)$ ,
- (B)  $\exists y \forall x (x \geq y)$ .

Using resolution, try to prove that (A) follows from (B). Do this even if you think that (B) does not logically entail (A); continue until the proof breaks down and you cannot proceed (if it does break down). Show the unifying substitution for each resolution step. If the proof fails, explain exactly where, how, and why it breaks down.

## 7 Automated Planning

**7.1** Describe the differences and similarities between problem solving and planning.

**7.2** The **monkey-and-bananas** problem is faced by a monkey in a laboratory with some bananas hanging out of reach from the ceiling. A box is available that will enable the monkey to reach the bananas if he climbs on it. Initially, the monkey is at  $A$ , the bananas at  $B$ , and the box at  $C$ . The monkey and box have a height *Low*, but if the monkey climbs onto the box he will have a height *High*, the same as the bananas. The actions available to the monkey include *Go* from one place to another, *Push* an object from one place to another, *ClimbUp* onto or *ClimbDown* from an object, and *Grasp* or *Ungrasp* an object. The result of a *Grasp* is that the monkey holds the object if the monkey and object are in the same place at the same height.

- (1) Write down the initial/goal state descriptions and the six action schemas in the PDDL.
- (2) Your schema for the *Push* is probably incorrect because if the object is too heavy, its position will remain the same when the *Push* schema is applied. Fix your action schema to account for heavy objects.
- (3) Construct two-level (hierarchical) planning in a PDDL description.

**7.3** <sup>+</sup> Imagine that we have a collection of blocks on a table and a robot arm that is capable of picking up blocks and putting them elsewhere. We assume that the robot arm can hold at most one block at a time. We also assume that the robot can only pick up a block if there is no other block on top of it. Finally, we assume that a block can only support or be supported by at most one other block, but that the table surface is large enough that all blocks can be directly on the table. There are only two actions available: *puton*( $x, y$ ), which picks up block  $x$  and moves it onto block  $y$ , and *putonTable*( $x$ ), which moves block  $x$  onto the table. Similarly, we have only two fluents: *On*( $x, y, s$ ), which holds when block  $x$  is on block  $y$ , and *OnTable*( $x, s$ ), which holds when block  $x$  is on the table.

- (1) Write the precondition axioms for the actions.
- (2) Write the effect axioms for the actions.
- (3) Show how successor state axioms for the fluents would be derived from these effect axioms. Argue that the successor state axioms are not logically entailed by the effect axioms by briefly describing an interpretation where the effect axioms are satisfied but the successor state ones are not.
- (4) Show how frame axioms are logically entailed by the successor state axioms.

**7.4**<sup>+</sup> Consider the **Sussman anomaly** problem. The problem was considered anomalous because the noninterleaved planners of the early 1970s could not solve it. Write a definition of the problem and solve it, either by hand or with a planning program. A noninterleaved planner is a planner that, when given two subgoals  $G_1$  and  $G_2$ , produces either a plan for  $G_1$  concatenated with a plan for  $G_2$ , or vice versa. Explain why a noninterleaved planner cannot solve this problem.

## 8 Uncertain Knowledge and Reasoning

**8.1** Consider two medical tests, A and B, for a virus. Test A is 95% effective at recognizing the virus when it is present, but has a 10% false positive rate (indicating that the virus is present, when it is not). Test B is 90% effective at recognizing the virus, but has a 5% false positive rate. The two tests use independent methods of identifying the virus. The virus is carried by 1% of all people. Say that a person is tested for the virus using only one of the tests, and that test comes back positive for carrying the virus. Which test returning positive is more indicative of someone really carrying the virus? Justify your answer mathematically.

**8.2** It is quite often useful to consider the effect of some specific propositions in the context of some general background evidence that remains fixed, rather than in the complete absence of information. The following questions ask you to prove more general versions of the product rule and Bayes's rule, concerning some background evidence  $\mathbf{e}$ :

- (1) Prove the conditionalized version of the general product rule:

$$\mathbf{P}(X, Y | \mathbf{e}) = \mathbf{P}(X | Y, \mathbf{e}) \mathbf{P}(Y | \mathbf{e}).$$

- (2) Prove the conditionalized version of Bayes's rule.
- (3) Suppose you are given a bag containing  $n$  unbiased coins. You are told that  $n - 1$  of these coins are normal, with heads on one side and tails on the other, whereas one coin is a fake, with heads on both sides.
- (4) Suppose you reach into the bag, pick out a coin at random, flip it, and get ahead. What is the (conditional) probability that the coin you chose is the fake coin?
- (5) Suppose you continue flipping the coin for a total of  $k$  times after picking it and see  $k$  heads. Now what is the conditional probability that you picked the fake coin?
- (6) Suppose you wanted to decide whether the chosen coin was fake by flipping it  $k$  times. The decision procedure returns *fake* if all  $k$  flips come up heads; otherwise, it returns *normal*. What is the (unconditional) probability that this procedure makes an error?

**8.3** Suppose you are a witness to a nighttime hit-and-run accident involving a taxi in Athens. All taxis in Athens are blue or green. You swear, under oath, that the taxi was blue. Extensive testing shows that, under dim lighting conditions, discrimination between blue and green is 75% reliable.

- (1) Is it possible to calculate the most likely color for the taxi? (Hint: distinguish carefully between the proposition that the taxi *is* blue and the proposition that it *appears* blue.)
- (2) What if you know that 9 out of 10 Athenian taxis are green?

**8.4<sup>+</sup>** Consider the following example:

*Metastatic cancer is a possible cause of a brain tumor and is also an explanation for an increased total serum calcium. In turn, either of these could cause a patient to fall into an occasional coma. Severe headaches could also be explained by a brain tumor.*

- (1) Represent these causal links in a belief network. Let  $a$  stand for “metastatic cancer”  $b$  for “increased total serum calcium,”  $c$  for “brain tumor,”  $d$  for “occasional coma,” and  $e$  for “severe headaches.”
- (2) Give an example of an independence assumption that is implicit in this network.
- (3) Suppose the following probabilities are given:
 
$$P(a) = 0.2$$

$$P(b|a) = 0.8 \quad P(b|\neg a) = 0.2$$

$$P(c|a) = 0.2 \quad P(c|\neg a) = 0.05$$

$$P(e|c) = 0.8 \quad P(e|\neg c) = 0.6$$

$$P(d|b, c) = 0.8 \quad P(d|\neg b, c) = 0.8$$

$$P(d|b, \neg c) = 0.8 \quad P(d|\neg b, \neg c) = 0.05$$
 and assume that it is also given that some patients are suffering from severe headaches but have not fallen into a coma. Calculate joint probabilities for the eight remaining possibilities (that is, according to whether  $a$ ,  $b$ , and  $c$  are true or false).
- (4) According to the numbers given, the a priori probability that the patient has metastatic cancer is 0.2. Given that the patient is suffering from severe headaches but has not fallen into a coma, are we now more or less inclined to believe that the patient has cancer? Explain.

## 9 Making Decisions

**9.1** The Allais paradox is stated as follows: People are given a choice between lotteries  $A$  and  $B$  and then between  $C$  and  $D$ , which have the following prizes:

$A$  : 80% chance of \$4000       $C$  20% chance of \$4000  
 $B$ : 100% chance of \$3000       $D$ : 25% chance of \$3000

Most people consistently prefer  $B$  over  $A$  (taking the sure thing), and  $C$  over  $D$  (taking the higher EMV). The normative analysis disagrees.

Prove that the judgments  $B \succ A$  and  $C \succ D$  in the Allais paradox violate the axiom of substitutability.

**9.2** Consider a student who has the choice to buy or not buy a textbook for a course. We’ll model this as a decision problem with one Boolean decision node,  $B$ , indicating whether the agent chooses to buy the book, and two Boolean chance nodes,  $M$ , indicating whether the student has mastered the material in the book, and  $P$ , indicating whether the student passes the course. Of course, there is also a utility node,  $U$ . A certain student, Sam, has an additive utility function: 0 for not buying the book and -\$100 for buying it; and \$2000 for passing the course and 0 for not passing. Sam’s conditional probability estimates are as follows:

$P(p|b, m) = 0.9$   
 $P(m|b) = 0.9$   
 $P(p|b, \neg m) = 0.5$   
 $P(m|\neg b) = 0.7$   
 $P(p|\neg b, m) = 0.8$   
 $P(p|\neg b, \neg m) = 0.3$

You might think that  $P$  would be independent of  $B$  given  $M$ , but this course has an open – book final – so having the book helps.

- (1) Draw the decision network for this problem.



- (2) Compute the expected utility of buying the book and of not buying it.
- (3) What should Sam do?

**9.3** We illustrate the current stalemate in the Donbas of the Russia-Ukraine conflict in the following table, which is a version of the prisoner's dilemma (called "Donbas game"). The table depicts a static position in the Donbas game with strategies for Ukraine and Russia respectively. Ukraine's options are given by the two rows: *Accept separation* (A) on the part of DL (the separatist Donetsk and Luhansk People's Republics) and *Regain territory* (R) by diplomatic or military means. Given that Ukraine would not agree to accept the separation, its strictly dominant strategy is given by the second row (Regain territory). Russian strategies are given by the two columns, *Keep intervening* (K), and *Stop intervening* (S). Given that Russia would realize the separation, its strictly dominant strategy is given by the first column (Keep intervening). Four outcomes in the table are  $(A, K)$ ,  $(A, S)$ ,  $(R, K)$ , and  $(R, S)$ , respectively.

Players	Strategies	Russia	
		Keep intervening	Stop intervening
Ukraine	Accept separation	$(A, K)$	$(A, S)$
	Regain territory	$(R, K)$	$(R, S)$

- (1) Of the four outcomes, point out the best and the worst for Ukraine and Russia, respectively.
- (2) Of the four outcomes, assign the payoff matrix. Which is a Nash equilibrium that provides the maximum payoff to both players. Why?

**9.4<sup>+</sup>** Prove that a dominant strategy equilibrium is a Nash equilibrium, but not vice versa.

## 10 Machine Learning

**10.1** Consider the following data, each consisting of three input bits and a classification bit:  $(111, 1)$ ,  $(110, 1)$ ,  $(011, 1)$ ,  $(010, 0)$ ,  $(000, 0)$ .

- (1) Draw a decision tree consistent with this data.
- (2) Draw a neural network using a threshold activation function consistent with this data.
- (3) Consider a learning program  $LP$  as taking a set of classified examples as input and returning a function that is supposed to calculate the appropriate classification given an unclassified example. Show how to use  $LP$  to construct a learning agent  $LA$ . The agent should learn from percepts that include the correct action to take, as well as doing the action. When a percept arrives that does not include the correct action, it should respond with the action that its past experience indicates might be appropriate. Write down the  $LA$  in pseudocode, being as precise as possible.

**10.2** Construct by hand a neural network that computes the XOR function of two inputs. Make sure to specify what sort of units you are using.

**10.3** Consider the following standard CNN architecture:

- Input data size: 32x32x3 (e.g. RGB image).
- Convolutional layer 1:
  - Number of filters: 16,
  - Filter size: 3x3,
  - Stride: 1,
  - Padding: same,
  - Activation function: ReLU.
- MaxPooling layer 1:
  - Pool size: 2x2,
  - Stride: 2
- Convolutional layer 2:
  - Number of filters: 32,
  - Filter size: 3x3,
  - Stride: 1,
  - Padding: same,
  - Activation Function: ReLU.
- MaxPooling layer 2:
  - Pool size: 2x2,
  - Stride: 2.
- Fully connected layer:
  - Number of neurons: 128,
  - Activation function: ReLU.
- Output layer:
  - Number of neurons: 10 (say for classification into 10 classes)
  - Activation function: Softmax

(1) Draw the picture of this CNN architecture.

(2) Calculate the total number of parameters in this CNN architecture.

**10.4<sup>+</sup>** Read the paper: LeCun&Bengio&Hinton, Deep learning, Nature 521, 436-444, 2015 ([www.nature.com/nature/journal/v521/n7553/full/nature14539.html](http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html)). Why is the deepness of deep learning (so-called very deep learning) critical for recognition systems? Are there some limitations of deep learning in the principle of intelligence?

## 11 Natural Language Understanding

**11.1** Consider the following context-free grammar (where  $X^*$  means zero or more occurrences of  $X$ ):

$S \rightarrow NP \ VP$   
 $S \rightarrow first \ S \ then \ S$   
 $NP \rightarrow Determiner Modifier Noun | Pronoun | Proper Noun$   
 $Determiner \rightarrow a | the | every$   
 $Pronoun \rightarrow she | he | it | him | her$   
 $Modifier \rightarrow Adjective^* | Noun^*$   
 $Adjective \rightarrow red | violet | fragrant$   
 $Noun \rightarrow rose | dahlia | violet$   
 $VP \rightarrow Verb NP$   
 $VP \rightarrow Intransitive Verb$   
 $VP \rightarrow Copula Adjective$   
 $Verb \rightarrow smelled | watered | was$   
 $Intransitive Verb \rightarrow smelled | rose$   
 $Copula \rightarrow was | seemed | smelled$   
 $Proper Noun \rightarrow Spike$

- (1) Which of the following sentences are generated by the grammar?
  - (a) first first Spike smelled fragrant then he smelled then he watered the violet violet
  - (b) the red red rose rose rose
  - (c) she was a violet violet violet
- (2) Show the parse tree for the sentence “First she watered the rose then it smelled”.
- (3) How many ways can the sentence “First the violet violet rose then the violet violet violet smelled” be parsed?
  - (a) 0, (b) 1, (c) 2, (d) 4, (e) more than 4.
- (4) What type of ambiguity is causing this multiplicity?
  - (a) lexical, (b) semantic, (c) referential.
- (5) In English, one can say “first A then B then C” whereas nested constructions such as “first first A then B then first C then” are not usually allowed. Show how to replace the rule “ $S \rightarrow first \ S \ then \ S$ ” by one or more new rules to reflect this.
- (6) *True/False*: A sentence that has exactly one parse tree is unambiguous.

**11.2** Consider the classification of spam email. Create a corpus of spam email and one of non-spam mail. Examine each corpus and decide what features appear to be useful for classification: unigram words, bigrams, message length, sender, and time of arrival.

- (1) Design a classification algorithm.
- (2) Write down the pseudocode for using an LM for spam email detection.

**11.3** Select five sentences and submit them to an online translation service. Translate them from English to another language and back to English. Rate the resulting sentences for grammaticality and preservation of meaning. (1) Repeat the process; does the second round of iteration give worse results or the same results? (2) Does the choice of intermediate language make a difference to the quality of the results? (3) If you know a foreign language, look at the translation of one paragraph into that language. Count and describe the errors made, and conjecture why these errors were made. (4) Can the performances mentioned above be improved by prompting in an LLM?

**11.4<sup>+</sup>** Read the paper: Vaswani A et al., Attention Is All You Need, [arxiv.org/abs/1706.03762](https://arxiv.org/abs/1706.03762).

- (1) Calculate the total number of parameters in this CNN architecture.
- (2) Write down the pseudocode of a pretrained language model based on the transformer. Explain the training processes of self-supervised learning and fine-tuning.

## 12 Robotics

**12.1** Write down the pseudocode of implementing the ViT (Vision Transformer) model by the Transformer. Explain the two modal visual-language models in the pertaining-finetuning paradigm.

**12.2** Humans are so adept at basic household tasks that they often forget how complex these tasks are. In this exercise, you will discover the complexity and recapitulate the last 30 years of developments in robotics. Consider the task of building an arch out of three blocks. Simulate a robot with four humans as follows:

**Brain.** The Brain directs the hands in the execution of a plan to achieve the goal. The Brain receives input from the Eyes, but *cannot see the scene directly*. The brain is the only one who knows what the goal is.

**Eyes.** The Eyes report a brief description of the scene to the Brain: “There is a red box standing on top of a green box, which is on its side.” Eyes can also answer questions from the Brain such as, “Is there a gap between the Left Hand and the red box?” If you have a video camera, point it at the scene and allow the eyes to look at the viewfinder of the video camera, but not directly at the scene.

**Left hand and right hand.** One person plays each Hand. The two Hands stand next to each other, each wearing an oven mitt on one hand, Hands execute only simple commands from the Brain — for example, “Left Hand, move two inches forward.” They cannot execute commands other than motions; for example, they cannot be commanded to “Pick up the box.” The Hands must be *blindfolded*. The only sensory capability they have is the ability to tell when their path is blocked by an immovable obstacle such as a table or the other Hand. In such cases, they can beep to inform the Brain of the difficulty.

**12.3** Do you think existing autonomous vehicles, say Tesla, have reached the SAE Level higher 4-5? List three scenarios to illustrate the differences between autonomous driving and human driving. If similar SAE intelligent grading is applied to humanoid robots, you provide a five-level grading standard.