MT Summer Term 2021 Ex2: SMT Intuition; Probability and Noisy Channel.

1. What is great about RBMT?

2. What is not so good about RBMT?

3. Why do we try to use machine learning or statistical estimation from data for MT?

4. In your own words, explain $\hat{e} = \underset{e}{\operatorname{argmax}} P(e|f)$.

5. What kinds of data do we need for SMT?

6. Given this bitext, which symbol is the likely Chinese symbol for "chicken"? Which symbol is the likely symbol for "soup"?

		CLASSIC SOUPS	Sm.	Lg.
清燉雞湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
雞飯湯	58.	Chicken Rice Soup	1.85	3.25
雞麵湯	59.	Chicken Noodle Soup	1.85	3.25
廣東雲吞	60.	Cantonese Wonton Soup	1.50	2.75
蕃茄蛋湯	61.	Tomato Clear Egg Drop Soup	1.65	2.95
雲吞湯	62.	Regular Wonton Soup	1.10	2.10
酸辣湯	63.	Hot & Sour Soup	1.10	2.10
蛋花湯	64.	Egg Drop Soup	1.10	2.10
雲蛋湯	65.	Egg Drop Wonton Mix	1.10	2.10
豆腐菜湯	66.	Tofu Vegetable Soup	NA	3.50
雞玉米湯	67.	Chicken Corn Cream Soup	NA	3.50
蟹肉玉米湯	68.	Crab Meat Corn Cream Soup	NA	3.50
海鮮湯	69.	Seafood Soup	NA	3.50

7. Given the following bitext and word alignment (indicated in terms of colour codes), (i) estimate a word based probabilistic translation dictionary (a translation model), (ii) find the best

I love the boy.
J'aime le garçon.

I love the dog.
J'aime le chien.

They love the dog.
Ils aiment le chien.

They talk to the girl.
Ils parlent à la fille.

I talk to the mother.
Je parle à la mère.

"translations" of

- They love the girl
- I talk to the dog

into French under the model and (iii) compute the probabilities for the best translations under the model based on the word-based translation probabilities, assuming that the probabilities are independent of each other: $P(f_1\, f_2 \ldots\, f_n | e_1 e_2 \cdots e_n) = \prod_{i=1}^{n} P(f_i | e_i)$.

8. Given a sequence of n numbers/measurements/numerical observations $x_1 x_2 \cdots x_n$, please define

- Population mean, sample mean
- Population variance, sample variance
- Population standard deviation, sample standard deviation

9. Explain the notions of sample space, outcome and event in set-based formalisations of probability.

10. What is a Laplace experiment and a Laplace probability?

11. Probabilities can be estimated from counts (relative frequencies). Give the following observations, estimate the probabilities:

	Car	Lorry	M-cycle	Cycle	Pedes.	n
$\#(x)$	124	49	7	64	271	515
$\frac{\#(x)}{n}$						
$P(x)$						

12. Please complete the following table describing the Boolean Algebra of events:

$A \cap B = B \cap A$	$A \cup B =$
$A \cap (B \cap C) =$	$A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cup C) =$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cap A = A$	$A \cup A =$
$A \cap (A \cup B) = A$	$A \cup (A \cap B) =$
$\overline{A \cap B} =$	$\overline{A \cup B} = \bar{A} \cap \bar{B}$
$A \cap \emptyset = \emptyset$	$A \cup \Omega = \Omega$
$A \cap \bar{A} =$	$A \cup \bar{A} =$
$\bar{\bar{A}} = A$	

13. Define conditional probability:

$$P(A|B) =_{def} \cdots$$

14. Given a fair 6-sided dice, what are

$$P(odd) = \cdots$$

$$P(odd|prime) = \cdots$$

$$P(prime|odd) = \cdots$$

$$P(even|prime) = \cdots$$

$$P(\{5\}|odd) = \cdots$$

$$P(\{2,5\}|odd) = \cdots$$

$$P(\emptyset|odd) = \cdots$$

15. When are two events mutually exclusive, when are two events independent?

16. Give the specific and the general version of the addition rule of probabilities:

$$P(A \cup B) = \cdots$$

$$P(A \cup B) = \cdots$$

17. Give the specific and the general version of the multiplication rule of probabilities:

$$P(A \cap B) = \cdots$$

$$P(A \cap B) = \cdots$$

18. Give the complement rule of probability: $P(\bar{A}) = \cdots$

19. Expand the following using the chain rule of probability: $P(w_1\, w_2 \ldots w_n) = \cdots$

20. What is the prior, the likelihood and the posterior in Bayes rule:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

21. Prove Bayes Rule.

22. Why is Bayes rule useful?

23. In your own words, relate the fundamental rule of statistical machine translation (SMT) to the noisy channel model (NC):

$$\hat{e} = \underset{e}{\operatorname{argmax}}\, P(e|f) = \underset{e}{\operatorname{argmax}} \frac{P(f|e) \times P(e)}{P(f)} = \underset{e}{\operatorname{argmax}}\, P(f|e) \times P(e)$$

What is the translation model, what is the language model, what is the source model, what is the channel model, what is the prior, the likelihood and the posterior? In what sense is this a MAP (maximum a posteriori) decision rule?

24. Rule of total probability: can you show in terms of a drawing why $P(A) = P(A \cap B) \cup P(A \cap \overline{B})$

25. Total probability: why is $(A \cap B) \cup P(A \cap \overline{B}) = P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B})$