Machine Translation: Summer Term 2021

Ex2: SMT Intuition; Probability and Noisy Channel.

Sangeet Sagar (7009050)
sasa00001@stud.uni-saarland.de

May 2, 2021

1. **What is great about RBMT?**
   RBMT is capable to achieve very high accuracy on a closed dataset. It does not need bilingual texts or parallel dataset. Another benefit of RBMT is that it is domain independent. Its rules are framed in such a way that it is valid for a sentence from any domain (certain exception might subject the need to write new rules). We also have a total control on the translations generated becuase all rules are hand made, hence easy to debug.

2. **What is not so good about RBMT?**
   RBMT is notoriously famous for the need of large number of rules that leads to a complex system. It can only capture local phenomena and any long range contexts are hard to capture. It is also hard to deal with rule interactions in big systems, ambiguity, and idiomatic expressions [Source: Wikipedia].

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

4. **In your own words, explain**

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

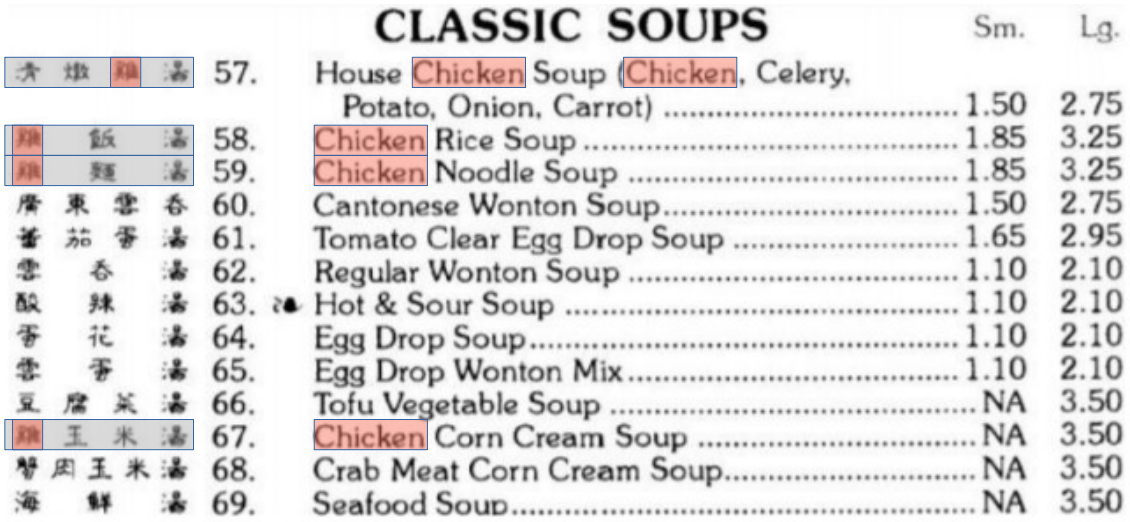
   This is the fundamental equation of Statistical Machine Translation (SMT) in terms of probability. The goal is to model a probability distribution P that finds the most probable English sentence e given a foreign language sentence f i.e. it should tell the probability of all possible English sentences that could be a translation of the given foreign language sentence. In the space of all possible English sentence, *argmax* searches for a sentence that maximizes this probability and the resulting (English) sentence is the best translation $\hat{e}$ of the input sentence.

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

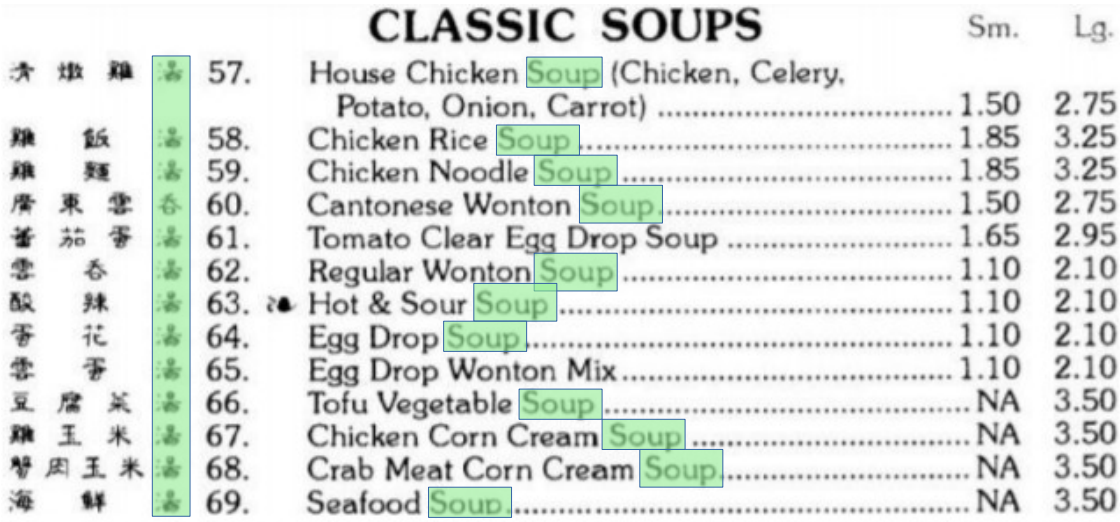
   - Human translation: Bi-text, parallel corpus of the language pair.
   - Monolingual data of the target language.

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.
57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
58.	Chicken Rice Soup	1.85	3.25
59.	Chicken Noodle Soup	1.85	3.25
60.	Cantonese Wonton Soup	1.50	2.75
61.	Tomato Clear Egg Drop Soup	1.65	2.95
62.	Regular Wonton Soup	1.10	2.10
63.	Hot & Sour Soup	1.10	2.10
64.	Egg Drop Soup	1.10	2.10
65.	Egg Drop Wonton Mix	1.10	2.10
66.	Tofu Vegetable Soup	NA	3.50
67.	Chicken Corn Cream Soup	NA	3.50
68.	Crab Meat Corn Cream Soup	NA	3.50
69.	Seafood Soup	NA	3.50



		Sm.	Lg.
	CLASSIC SOUPS		
57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
58.	Chicken Rice Soup	1.85	3.25
59.	Chicken Noodle Soup	1.85	3.25
60.	Cantonese Wonton Soup	1.50	2.75
61.	Tomato Clear Egg Drop Soup	1.65	2.95
62.	Regular Wonton Soup	1.10	2.10
63.	Hot & Sour Soup	1.10	2.10
64.	Egg Drop Soup	1.10	2.10
65.	Egg Drop Wonton Mix	1.10	2.10
66.	Tofu Vegetable Soup	NA	3.50
67.	Chicken Corn Cream Soup	NA	3.50
68.	Crab Meat Corn Cream Soup	NA	3.50
69.	Seafood Soup	NA	3.50



		Sm.	Lg.
	CLASSIC SOUPS		
57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50	2.75
58.	Chicken Rice Soup	1.85	3.25
59.	Chicken Noodle Soup	1.85	3.25
60.	Cantonese Wonton Soup	1.50	2.75
61.	Tomato Clear Egg Drop Soup	1.65	2.95
62.	Regular Wonton Soup	1.10	2.10
63.	Hot & Sour Soup	1.10	2.10
64.	Egg Drop Soup	1.10	2.10
65.	Egg Drop Wonton Mix	1.10	2.10
66.	Tofu Vegetable Soup	NA	3.50
67.	Chicken Corn Cream Soup	NA	3.50
68.	Crab Meat Corn Cream Soup	NA	3.50
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 "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:

$$P(f_1 f_2 \dots f_n) = \prod_{i=1}^{n} P(f_i|e_i)$$

Given the above aligned data, we prepare the collated statistics-

English word	Foreign word	Frequency	Probablity
I	J'	2	0.67
	Je	1	0.33
love	aime	2	0.67
	aiment	1	0.33
the	le	3	0.60
	la	2	0.40
boy	garçon	1	1
girl	fille	1	1
mother	mère	1	1
dog	chiene	3	1
they	lls	3	1
talk	parlent	2	0.67
	parle	1	0.33
to	à	2	0.67
	au/_the	1	0.33

1)

They	love	the	girl
lls	aime	le	fille
1	0.67	0.60	1
lls	aiment	la	fille
1	0.33	0.40	1

But from the parallel data, we see that lls occurs with aiment. Hence the choice aime is ruled out by the language model. Therefore the appropriate translation would be:

They	love	the	girl
lls	aiment	le	fille
1	0.33	0.60	1

P(lls aiment le fille) = 1*0.33*0.60*1 = 0.198

2)

I	talk	to	the	dog
J'	parlent	à	le	chiene
0.67	0.67	0.67	0.60	1

Je	parle	au	la	chiene
0.33	0.33	0.33	0.40	1

From the parallel data, we see that Je occurs with parle. Hence the choice parlent is ruled out by the language model. So, Je parle is a good translation. Now, parle occurs together with à twice and with au just once. Hence, parle à is a more suitable choice than parle au. Also, à la is seen to occur together in the aligned data, hence the other possibility i.e. à le is also ruled out. Therefore, the most appropriate translation would be-

I	talk	to	the	dog
Je	parle	à	la	chiene
0.33	0.33	0.67	0.40	1

P(Je parle à la chiene) = 0.33*0.33*0.67*0.40*1 = 0.0291

8. **Given a sequence of n numbers/measurements/numerical observations $x_1, x_2, \ldots x_n$, 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:**
12. **Please complete the following table describing the Boolean Algebra of events:**
13. **Define conditional probability:**
14. **Given a fair 6-sided dice, what are**
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:**

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

18. **Give the complement rule of probability:**

19. **Expand the following using the chain rule of probability**

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

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):**
    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**

25. **Total probability: why is**