Second Question Set

David Barbas Rebollo

Statistical Structured Prediction
January, 2021

1 Theoretical Questions

To complete the next exercises, the following equation has been used. It is avaible on the lecture notes and is used to calculate the rule estimation for a PCFG given a training dataset:

$$p(A \rightarrow \alpha) = \frac{\Sigma_{x \epsilon D} \dfrac{1}{P_\theta(x, \Delta_x)} \Sigma_{t_x \epsilon \Delta_x} N(A \rightarrow \alpha, t_x) * P_\theta(x, t_x)}{\Sigma_{x \epsilon D} \dfrac{1}{P_\theta(x, \Delta_x)} \Sigma_{t_x \epsilon \Delta_x} N(A, t_x) * P_\theta(x, t_x)}$$

1.1 Question 1

The objective of the first exercise is to analyze the behaviour of the Inside-Outside algorithm when the initial probabilities of all rules of a grammar are equiprobable.

For this analysis, the grammar, dataset and used are the ones available in slide 50 and 51. The rule referred to is: p (Suj → Art Nom Adj).

P_θ ((la vieja)(demanda ayuda)) $= \dfrac{49}{60} + \dfrac{1}{240} = \dfrac{197}{240}$

P_θ (la mujer oculta pelea) $= \dfrac{49}{60} + \dfrac{1}{240} = \dfrac{197}{240}$

P_θ (la vieja ayuda) $= \dfrac{1}{120}$

$$p\ (\text{Suj} \to \text{Art Nom Adj}) = \frac{\dfrac{1/240}{197/240}}{\dfrac{49/60 + 1/240}{197/240} + \dfrac{49/60 + 1/240}{197/240} + \dfrac{1/120}{1/120}} = \frac{1}{591}$$

Given the results from above it can be concluded that in this escenario rules with more elements in the right hand side would be penalized with a lower probability.

1.2 Question 2

Question 2 asks to repeat exercise one but with Viterbi algorithm instead of Inside-Outside algorithm. All other elements used will be the same.

Given the information calculated in the previous exercise the following calculation is developed:

$$p\ (\text{Suj} \to \text{Art Nom Adj}) = \frac{\dfrac{1/240}{1/240}}{\dfrac{49/60}{49/60} + \dfrac{1/240 + 49/60}{197/240} + \dfrac{1/120}{1/120}} = \frac{1}{3}$$

It can be observed that estimating the probability of a rule of a set of equiprobable probabilities via Viterbi algorithm has no effect.

1.3 Question 3

Using the example of page 50, the estimation of the rule $(Suj \to ArtNomAdj)$ is computed via the Inside-Outside algorithm. The training samples, which include brackers, are: $D = \{$(*la vieja*)(*demanda ayuda*), *la mujer oculta pelea*, *la vieja ayuda*$\}$.

P_θ ((la vieja)(demanda ayuda)) $= 0.9 * 10^{-3}$
P_θ (la mujer oculta pelea) $= 12.66^{-3}$
P_θ (la vieja ayuda) $= 7 * 10^{-3}$

$$p\ (\text{Suj} \to \text{Art Nom Adj}) = \frac{\dfrac{0.01176}{0.01266}}{\dfrac{0.0009}{0.0009} + \dfrac{0.0009 + 0.01176}{0.01266} + \dfrac{0.007}{0.007}} = 0.3096$$

1.4 Question 4

In this exercise the same problem is proposed, as in Question 3, but doing the estimation via Viterbi. However, in this case only the tree with maximum probability is considered.

$$p\ (\mathrm{Suj} \rightarrow \mathrm{Art\ Nom\ Adj}) = \frac{\frac{0.01176}{0.01176}}{\frac{0.0009}{0.0009} + \frac{0.0009 + 0.01176}{0.01266} + \frac{0.007}{0.007}} = \frac{1}{3}$$

1.5 Question 5

This questions asks to calculate the rule (Suj → Art Nom Adj) using the Inside-Outside algorithm and the training set $D = \{$*la vieja demanda ayuda*), *la mujer oculta pelea*, *la vieja mujer oculta demanda ayuda*$\}$.

As the current grammar used cannot generate de phrase "*la vieja mujer oculta demanda ayuda*" a new rule must be created, (Suj → Art Adj Nom Adj). To assign this new rule a probability value, the other rules' probability are reduced by 0.1. The other rules are (Suj → Art Nom) and (Suj → Art Adj Nom). This is done so all rules with the same left side, Suj, give a sumed probability of 1.

After applying Inside-Outside the new estimation for this rule is:

$$p(Suj \rightarrow ArtNomAdj) = \frac{\frac{0.01176}{0.01248}}{\frac{0.00072 + 0.00168}{0.0024} + \frac{0.00072 + 0.01176}{0.01248} + \frac{0.0002268}{0.0002268}} = 0.3141$$

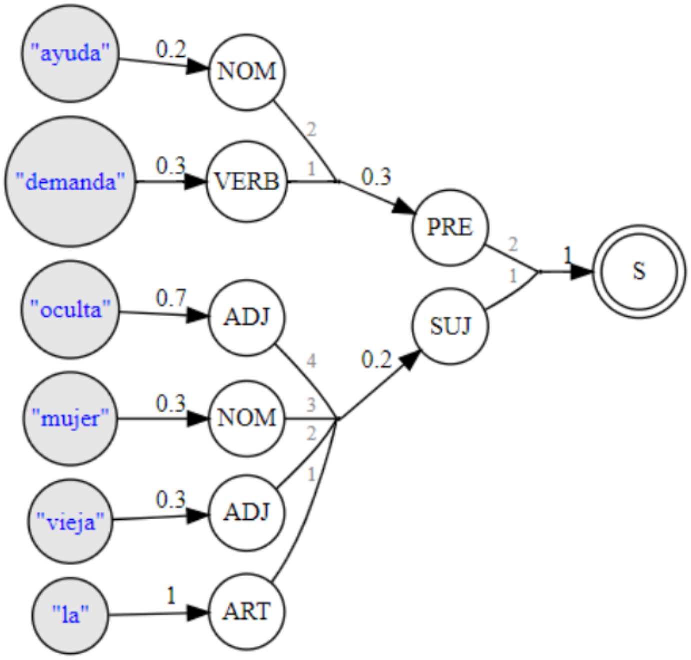

Figure 1: Tree generated with the new phrase

1.6 Question 6

Question 6 asks to repeat question 5 but using the k-best estimation algorithm with k=2 and only computing 1 iteration. K-best estimation algorithm is a general mehthod for rule estimation. Meanwhile, Viterbi ann Inside-Outside are concrete instances of k-best algorithm. When k=1 the algorithm is equivalent to Viterbi and whñen k=max, max being the total number of existing trees, is equivalent to Inside-Outside. Therefore, for k=2 the result will be the same as in question 5.

$p(Suj \rightarrow ArtNomAdj) = 0.3141$

2 Practical Questions

In this assigment, the *SCFG-toolkit* is used to complete the following questions.

2.1 Question 8

In this question, 4 new grammars are created were each has a different number of non-terminal symbols. Using *SCFG-toolkit*, the grammars are training over 700 iterations using the dataset provided of 10,000 samples. Finally, 1000 strings are generated and analysed, to see how many right-angled triangles are generated.

Table 1: Number of right-angled triangles generated with the number of non-terminal symbols.

# Non-terminal symbols	# Right-angled triangles
5	29
10	63
15	61
20	84

2.2 Question 9

This exercise, studies the classification of samples after training using Inside-Outside or Viterbi and bracketed or not samples. It is important to comment that the number of iterations was reduced to 100 due to the excessive computation time required.

2.2.1 Inside-Outside and Bracketed

Table 2: Inside-Outside and Bracketed trained model's confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	794	206	0	206	20.6
IS	531	225	244	775	77.5
SC	108	145	747	253	25.3
Total				1234	41.13

Given the results obtained in table above, it is observed that the model lacks an ability to recognise isosceles triangles. This is because, a high number of equilateral triangles are confused as isosceles triangles. Meanwhile, the error of classification of the other 2 types is relatively lower.

2.2.2 Inside-Outside and Not Bracketed

Table 3: Inside-Outside and Not Bracketed trained model's confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	783	217	0	217	21.7
IS	483	366	151	634	63.4
SC	48	187	765	235	23.5
Total				1086	36.20

In comparison with the Bracketed model, this experiment shows that training with Not Bracketed samples allows a better performance classifying isosceles triangles.

2.2.3 Viterbi and Bracketed

Table 4: Viterbi and Bracketed trained model's confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	77	843	80	923	92.3
IS	70	850	80	150	15.0
SC	12	676	312	688	68.8
Total				1761	58.70

This model's classification is being hindered by the fact that many equilateral's and scalene's triangles class is confused for isosceles triangles. In particular, equilateral triangle are the most misclassified.

2.2.4 Viterbi and Not Bracketed

Table 5: Viterbi and Not Bracketed trained model's confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	67	933	0	933	93.3
IS	171	612	217	388	38.8
SC	55	372	573	427	42.7
Total				1748	58.27

For the last experiment a similar performance is obtained as the previous experiment. Equilateral triangles are mainly confused as isosceles triangles. However, in comparison with the previous model this one has the error more spread out between the non-target classes.

2.2.5 Conclusion

After the evaluation of all experiments, it is possible to see that the Inside-Outside algorithm trains better the model. Despite, observing a tendency that training with Not Bracketed samples leads to a better performance it is not enough to be a decisive conclusion. The best model obtained is trained with Inside-Outside algorithm and Not Bracketed samples.