PEE - First Question Set

David Barbas Rebollo

Question 4

- Definition: Given $x = x_1 \dots x_T \in \Sigma^*$ and $A \in N$

$$e(A,i,j) \overset{def}{=} P_\theta(A \overset{*}{\Rightarrow} x_{i+1} \dots x_j)$$

- Initialization: $\forall A \in N; \quad \forall i : 0 \dots T-1;$

$$e(A,i,i+1) = p(A \rightarrow b) \cdot \delta(b, x_i + 1)$$

- Recursion: $\forall A \in N; \ \forall l : 2 \dots T; \ \forall i : 0 \dots T-l;$

$$e(A,i,i+l) = \left\{ \sum_{B,C \in N} p(A \rightarrow BC) \cdot e(B,i,i+l-1) \cdot e(C,i+l-1,i+l) \right\}$$

- Final Result: Probability of a sentence: $P_\theta(x) = e(S,0,T)$
- Temporal Cost: $O(N^2)$ Being N the length of the sentence

Question 1

- Outside algorithm's initial expression:

$\forall A \in N; \quad f(A,0,T) = \delta(A,S)$

Kronecker delta is a function which returns 1 if both variables are equal and 0 otherwise.

$$\delta(A,S) = \begin{cases} 1 & A = S \\ 0 & A \neq S \end{cases}$$

This is done so the substate A is equal to the initial state of the tree. Therefore, its probability is maximum and set to 1.

- Outside algorithm's final expression:

$$P_\theta(x) = \sum_{A \in N} f(A,i,i+1) \cdot p(A \rightarrow b) \cdot \delta(b, x_{i+1}) \qquad 0 \leq i \leq T-1$$

The final result is the sum of all right and left productions multiplied by their own A production probability. Recursively, the algorithms computes all outside production's probabilities to determine the probability of sentence x.

Question 2

- Definition: Optimal score and sequence for $x_1 \dots x_t$ ending at $y_t = s \in Y$

$$\varphi_t(s) \overset{\text{def}}{=} \max_{y_1^t;\, y_t = s} \prod_{i=1}^{t} \Psi_i(y_{i-1}, y_i, x_i)$$

$$b_t(s) \overset{\text{def}}{=} \arg\max_{y_1^t;\, y_t = s} \prod_{i=1}^{t} \Psi_i(y_{i-1}, y_i, x_i)$$

- Initialization: $\forall s \in Y;\ \varphi_1(s) = \Psi(y_0 = \text{null}, y_1 = s, x_1)$

$$b_1(s) = 0$$

- Recursion: $\forall t = 2 \dots T;\ \forall s \in Y;$

$$\varphi_t(s) = \max_{s' \in Y} \left\{ \varphi_{t-1}(s') \cdot \Psi_t(y_{t-1} = s', y_t = s, x_t) \right\}$$

$$b_t(s) = \arg\max_{s' \in Y} \left\{ \varphi_{t-1}(s') \cdot \Psi_t(y_{t-1} = s', y_t = s, x_t) \right\}$$

Final Result: Optimal score for x:

$$\max_s \varphi_T(s)$$

Optimal sequence, $\hat{y}$, for x:

$$y_T = \arg\max_s \varphi_T(s)$$

$$\forall t = T \dots 2;\ y_{t-1} = b_t(y_t)$$

$$\hat{y} = y_1 \dots y_T$$

First Question Set

David Barbas Rebollo

Statistical Structured Prediction
December, 2020

1 Practical Assigments

In this assigment, the *SCFG-toolkit* is used for the classification of equilateral, isosceles and scalene triangles using 3 different models and 1000 test samples of each triangle type.

1.1 Statistical evaluation

In the first task, the perplexity of the different models is analysed. Perplexity is a measurement of how well a probability distribution or probability model predicts a sample. It can be used to compare different models and it is importnat to know that a lower value indicates a better probability distribution. Futhermore, this measurement is related to how challenging a task is for a model.

Table 1: G1 model perplexity results table.

Model	Test Set	Perplexity
G1-EQ	TS-EQ	99095.20
G1-EQ	TS-IS	191.88
G1-EQ	TS-SC	245.91
G1-IS	TS-EQ	47334.82
G1-IS	TS-IS	26581.81
G1-IS	TS-SC	21621.97
G1-SC	TS-EQ	48369.45
G1-SC	TS-IS	27508.97
G1-SC	TS-SC	33618.35

Table 1 shows that the perplexity achieved by the model G1, in no instance, obtains the lowest perplexity for the target class. In particular, G1-EQ obtains cosiderably low perplexities for TS-IS and TS-SC but not for TS-EQ, the target class. Nevertheless, the model, in general, obtains high perplexity values.

Table 2: G2 model perplexity results table.

Model	Test Set	Perplexity
G2-EQ	TS-EQ	635136.42
G2-EQ	TS-IS	389521.32
G2-EQ	TS-SC	520221.74
G2-IS	TS-EQ	102498.67
G2-IS	TS-IS	52217.60
G2-IS	TS-SC	49786.01
G2-SC	TS-EQ	68152.56
G2-SC	TS-IS	42723.12
G2-SC	TS-SC	43543.30

Table 2 shows that model G2 obtains overall high perplexity values, meaning this task is more challenging for this model in particualr. In addition, in no instance the target class obtains the lowest perplexity.

Table 3: G3 model perplexity results table.

Model	Test Set	Perplexity
G3-EQ	TS-EQ	986.21
G3-EQ	TS-IS	550208.13
G3-EQ	TS-SC	1081980.37
G3-IS	TS-EQ	1283.38
G3-IS	TS-IS	1254.90
G3-IS	TS-SC	1218.74
G3-SC	TS-EQ	1272.73
G3-SC	TS-IS	1258.35
G3-SC	TS-SC	1218.04

Table 3 shows a considerable reduction in the values of perplexity obtained. Despite obtaining high values for the G3-EQ, neither of these are obtained from the target class. Futhermore, G3-EQ and G3-SC obtain the lowest perplexity for the target class. Meanwhile, G3-IS is really close of obtaining the lowest value but did not reach it.

With this insights, it is possible to assume, that out of the 3 proposed models, this one will be the one to perform better.

1.2 Classification

Table 4: G1 model classification confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	597	285	118	403	40.3
IS	88	471	441	529	52.9
SC	71	406	523	477	47.7
Total				1409	46.97

As it can be observed on Table 4, the model G1 has the highest difficulty detecting isosceles and scalene triangles. This is because, in most cases these types are confused with each other.

Table 5: G2 model classification confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	281	211	508	719	71.9
IS	71	215	714	785	78.5
SC	81	190	729	271	27.1
Total				1775	59.17

On Table 5, it can be observed how the model G2 has difficulties detecting equilateral and isosceles triangles. The reason behind this fact, is that these 2 types are extently confused with scalene triangle leading to a dreadful classification in general and a high error percentage overall. This model is prone to identify any triangle as a scalene triangle.

Table 6: G3 model classification confusion matrix.

	EQ	IS	SC	Error	Error(%)
EQ	789	102	109	211	21.1
IS	178	512	310	488	48.8
SC	106	421	473	527	52.7
Total				1226	40.87

Table 6 displays, the best performance is obtained with model G3. Achieving the lowest error percentage, 40.87%. However, the error achieved is still too high. It is possible to observed, how this model has problems identifying scalene triangles, as these are confused with isosceles triangles and viceversa.