

Name: VEN THON

ID: e20191250

Group: I5-GIC-C

Exercise 2.2 Tri-gram Language Model

We have

Considering the following example:

<s> I am Sam </s>

<s> Sam I am </s>

<s> Sam I like </s>

<s> Sam I do like </s>

<s> do I like Sam </s>

1. What is the most probable next word predicted by the model for the following sequence

a) <s> Sam ...

$$P(</s>|am\ Sam)=1$$

$P(I|<s>Sam)=1$, most fit since it is <s> Sam from (a)

$$P(</s>|like\ Sam)=1$$

b) <s> Sam I do ...

$P(loke|I\ do)=1$ most fit since it is I do from (b)

$$P(I|<s>do)=1$$

c) <s> Sam I am Sam ...

$P(</s>|am\ Sam)=1$ most fit since is am Sam from (c)

d) <s> do I like ...

$P(</s>|I\ like)=0.5$ redundance with "like</s>"

$P(</s>|do\ like)=1$ redundance with "like</s>"

$$P(Sam|I\ like)=0.5$$

Thus

a) <s> Sam I

b) <s> Sam I do like

c) <s> Sam I am Sam

d) <s> do I like Sam

2. Compute the probability and the perplexity of each of the following sentences:

a) <s> Sam I do I like </s>

$$P(Sam|<s><s>) \times P(I|<s>Sam) \times P(do|Sam\ I) \times P(I|I\ do) \times P(like|do\ I) \times P(</s>|like) \times P(|like</s>)=3/5 \times 1 \times 1/3 \times 1 \times 1 \times 1/2 \times 2/5 = 0.04$$

So its perplexity is root of $1/0.04 = 1.58$

b) <s> Sam I am </s>

$$P(Sam|<s><s>) \times P(I|<s>Sam) \times P(am|Sam\ I) \times P(</s>|I\ am) \times P(</s>|am</s>) = 3/5 \times 1 \times 1/1 \times 1/2 \times 1/5 = 0.2$$

So its perplexity is root of $1/0.02 = 2.18$

c) $\langle s \rangle$ I do like Sam I am $\langle /s \rangle$

$$\begin{aligned} & P(I | \langle s \rangle \langle s \rangle) \times P(do | \langle s \rangle I) \times P(like | I do) \times P(Sam | do like) \times P(I | like Sam) \times \\ & P(am | Sam I) \times P(\langle /s \rangle | I am) \times P(\langle /s \rangle | am \langle /s \rangle) \\ & = 1/5 \times 1 \times 1 \times 1 \times 1 \times 1/3 \times 1/2 \times 1/5 = 0.006 \end{aligned}$$

So its perplexity is root of $1/0.006 = 1.87$