

CSCI 375 HW 3

Tongyu Zhou

October 2 2017

Code is formatted in a way where all the answers can be obtained by just running the entire thing. The outputs were inserted as screenshots here for easier access.

1 Preparation [40 points]

1. Transition counts:

	.	ADJ	ADV	CONJ	DET	NOUN	PREP	PRON	S	VERB
.	1	1	1	1	1	1	1	1	1	3
ADJ	2	1	2	2	1	1	1	2	1	1
ADV	1	1	1	1	1	2	1	1	3	1
CONJ	1	1	1	1	1	1	1	1	3	1
DET	1	1	1	1	1	1	4	1	1	1
E	1	2	1	1	1	4	2	1	1	3
NOUN	1	4	1	1	4	1	1	2	1	1
PREP	1	1	1	1	1	2	2	1	3	2
PRON	1	1	3	2	1	1	1	2	1	3
VERB	2	1	1	1	1	3	1	4	2	3

2. Transition probabilities:

	.	ADJ	ADV	CONJ	DET	NOUN	PREP	PRON	S	VERB
.	0.083333	0.071429	0.076923	0.083333	0.076923	0.058824	0.066667	0.0625	0.058824	0.157895
ADJ	0.166667	0.071429	0.153846	0.166667	0.076923	0.058824	0.066667	0.1250	0.058824	0.052632
ADV	0.083333	0.071429	0.076923	0.083333	0.076923	0.117647	0.066667	0.0625	0.176471	0.052632
CONJ	0.083333	0.071429	0.076923	0.083333	0.076923	0.058824	0.066667	0.0625	0.176471	0.052632
DET	0.083333	0.071429	0.076923	0.083333	0.076923	0.058824	0.266667	0.0625	0.058824	0.052632
E	0.083333	0.142857	0.076923	0.083333	0.076923	0.235294	0.133333	0.0625	0.058824	0.157895
NOUN	0.083333	0.285714	0.076923	0.083333	0.307692	0.058824	0.066667	0.1250	0.058824	0.052632
PREP	0.083333	0.071429	0.076923	0.083333	0.076923	0.117647	0.133333	0.0625	0.176471	0.105263
PRON	0.083333	0.071429	0.230769	0.166667	0.076923	0.058824	0.066667	0.1250	0.058824	0.157895
VERB	0.166667	0.071429	0.076923	0.083333	0.076923	0.176471	0.066667	0.2500	0.117647	0.157895

3. Emission counts:

	.	ADJ	ADV	CONJ	DET	NOUN	PREP	PRON	VERB
,	3	1	1	1	1	1	1	1	1
I	1	1	1	1	1	1	1	2	1
a	1	1	1	1	2	1	1	1	1
above	1	1	1	1	1	1	2	1	1
are	1	1	1	1	1	1	1	1	2
blazing	1	2	1	1	1	1	1	1	1
diamond	1	1	1	1	1	2	1	1	1
gone	1	1	1	1	1	1	1	1	2
he	1	1	1	1	1	1	1	2	1
high	1	2	1	1	1	1	1	1	1
how	1	1	2	1	1	1	1	1	1
in	1	1	1	1	1	1	2	1	1
is	1	1	1	1	1	1	1	1	3
light	1	1	1	1	1	2	1	1	1
like	1	1	1	1	1	1	2	1	1
little	1	3	1	1	1	1	1	1	1
nothing	1	1	1	1	1	2	1	1	1
shining	1	1	1	1	1	1	1	1	2
show	1	1	1	1	1	1	1	1	2
sky	1	1	1	1	1	2	1	1	1
so	1	1	2	1	1	1	1	1	1
star	1	1	1	1	1	2	1	1	1
sun	1	1	1	1	1	2	1	1	1
the	1	1	1	1	3	1	1	1	1
then	1	1	2	1	1	1	1	1	1
twinkle	1	1	1	1	1	1	1	1	3
up	1	1	1	1	1	1	2	1	1
upon	1	1	1	1	1	1	2	1	1
what	1	1	1	1	1	1	1	2	1
when	1	1	1	3	1	1	1	1	1
wonder	1	1	1	1	1	1	1	1	2
world	1	1	1	1	1	2	1	1	1
you	1	1	1	1	1	1	1	3	1
your	1	1	1	1	1	1	1	2	1

4. Emission probabilities:

	.	ADJ	ADV	CONJ	DET	NOUN	PREP	PRON	VERB
,	0.083333	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
I	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.050	0.023256
a	0.027778	0.026316	0.027027	0.027778	0.054054	0.02439	0.025641	0.025	0.023256
above	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.051282	0.025	0.023256
are	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.046512
blazing	0.027778	0.052632	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
diamond	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
gone	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.046512
he	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.050	0.023256
high	0.027778	0.052632	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
how	0.027778	0.026316	0.054054	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
in	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.051282	0.025	0.023256
is	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.069767
light	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
like	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.051282	0.025	0.023256
little	0.027778	0.078947	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
nothing	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
shining	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.046512
show	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.046512
sky	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
so	0.027778	0.026316	0.054054	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
star	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
sun	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
the	0.027778	0.026316	0.027027	0.027778	0.081081	0.02439	0.025641	0.025	0.023256
then	0.027778	0.026316	0.054054	0.027778	0.027027	0.02439	0.025641	0.025	0.023256
twinkle	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.069767
up	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.051282	0.025	0.023256
upon	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.051282	0.025	0.023256
what	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.050	0.023256
when	0.027778	0.026316	0.027027	0.083333	0.027027	0.02439	0.025641	0.025	0.023256
wonder	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.025	0.046512
world	0.027778	0.026316	0.027027	0.027778	0.027027	0.04878	0.025641	0.025	0.023256
you	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.075	0.023256
your	0.027778	0.026316	0.027027	0.027778	0.027027	0.02439	0.025641	0.050	0.023256

2 POS Tagging [60 points]

1.

- a) 1.1416299097985194e-18
- b) 4.4868395706689815e-17
- c) 1.1687383801168684e-17

2.

	show	your	light	when	nothing	is	shining
ADV	0.004769	0.000008	7.297239e-08	8.375597e-10	2.908193e-12	1.093306e-14	1.254870e-16
.	0.001634	0.000024	7.499940e-08	4.304126e-10	2.988977e-12	1.123675e-14	6.448639e-17
PREP	0.004525	0.000015	6.923022e-08	7.946079e-10	2.759055e-12	1.037239e-14	1.190518e-16
NOUN	0.001435	0.000007	2.634125e-07	3.779233e-10	5.248934e-12	3.946567e-14	5.662220e-17
VERB	0.005472	0.000020	2.511608e-07	1.081036e-09	2.502399e-12	2.822254e-14	3.239316e-16
CONJ	0.004902	0.000008	7.499940e-08	1.291238e-09	2.988977e-12	1.123675e-14	6.448639e-17
PRON	0.001471	0.000043	1.349989e-07	3.873714e-10	5.380158e-12	1.011308e-14	5.803775e-17
DET	0.001590	0.000008	7.297239e-08	4.187799e-10	2.908193e-12	1.093306e-14	6.274352e-17
ADJ	0.001548	0.000008	1.421041e-07	4.077593e-10	5.663324e-12	1.064535e-14	6.109237e-17

POS tagging: ['VERB', 'PRON', 'NOUN', 'CONJ', 'ADJ', 'NOUN', 'VERB']

Probability: 5.114710253535412e-17

Final probability was computed by taking the maximum of the last column of the table and multiplying that by $P(<END> | \text{POS tag of the max})$. While it seems sketchy that "nothing" is an ADJ (instead of NOUN), the original corpus was limited and did not have a NOUN following CONJ—the accuracy of the algorithm would probably be higher with a larger and more extensive corpus.