Reid Russell

NLP Homework 1

1) PLSI

Doc	Initial topic
• ABCDAB	112212
• BADBBC	121211
DDCCA	12222

Term | Prob

Term	Prob
A	2/8 = .25
B	3/8 = .375
C	1/8 = .125
D	2/8 = .25

Term | Prob

Term	Prob
A	2/9 = .22
B	2/9 = .22
C	3/9 = .33
D	2/9 = .22

Document 1

Topic	Prob
1	3/6 = 0.5
2	3/6 = 0.5

Document 2

Topic	Prob
1	4/6 = .66
2	2/6 = .33

Document 3

Topic	Prob
1	1/5 = .2
2	4/5 = .8

Doc 1

Topic	A	B	C	D	Σ
1	.532	.630	.275	.534	1.96
2	.468	.37	.725	.468	2.04

Doc 2

Topic	A	B	C	D	Σ
1	.692	.771	.429	.692	2.58
2	.308	.229	.571	.308	1.415

Doc 3

Topic	A	B	C	D	Σ
1	.2195	.297	.086	.2195	0.82
2	.780	.703	.914	.780	3.18

Topic

Topic	A	B	C	D	Σ
Doc 1 T1	.537	.749	.350	.459	1.995
Doc 1 T2	.463	.251	.7499	.541	2.0
Doc 2 T1	.703	.859	.405	.634	2.6
Doc 2 T2	.297	.141	.595	.366	1.399
Doc 3 T1	.176	.355	.0578	.135	.724
Doc 3 T2	.824	.645	.9421	.865	3.276

	Topic 1	Topic 2
A	.2441	.2274
B	.4423	.1602
C	.1081	.3504
D	.2057	.2621
Total	1	1

	Doc 1	Doc 2	Doc 3
Topic 1	.519	.688	.166
Topic 2	.481	.312	.834

2)

Bigram	Frequency
AB	4
AC	3
AD	3
AE	1
BA	1
BB	2
BD	4
CA	2
CB	5
DB	2
DC	1
DE	2
EA	1
ED	1
total:	32

2nd letter

1st Letter	A	B	C	D	E
A	0	4	3	3	1
B	1	2	0	4	0
D	2	5	0	0	0
C	0	2	1	0	2
E	1	0	0	1	0

step 1 Probabilities

	A	B	C	D	E
A	0	.13	.09	.09	.03
B	.031	.06	0	.13	0
C	.063	.16	0	0	0
D	0	.06	.03	0	.06
E	.031	0	0	.03	0

step 2 Good Turing Smoothed

	A	B	C	D	E
A	.014	.078	.125	.125	.05
B	.05	.047	.014	.078	.014
C	.014	0	.014	.014	.014
D	.014	.047	.05	.014	.047
E	.05	.014	.014	.05	.014

				×32
0	1/32	0+1=1	5/11 = .0142	⇒ 0.455
1	1/32	1+1=2	4/5 = .05	⇒ 1.6
2	1/32	2+1=3	2/4 = .047	⇒ 1.5
3	1/32	3+1=4	2/2 = 1.25	⇒ 4
4	1/32	4+1=5	1/2 = .078	⇒ 2.5
5	1/32	5+1=6	0/1 = 0	⇒ 0

step 3) GT smoothed + entries ×32

	A	B	C	D	E
A	.455	2.5	4	4	1.6
B	1.6	1.5	.455	2.5	.455
C	1.5	0	.455	.455	.455
D	.455	1.5	1.6	.455	1.5
E	1.6	4.55	.455	1.6	.455

step 4 GT smoothed + entries ×32 + 1

	A	B	C	D	E
A	1.455	3.5	5	5	2.6
B	2.6	2.5	1.455	3.5	1.455
C	2.5	1	1.455	1.455	1.455
D	1.455	2.5	2.6	1.455	2.5
E	2.6	1.455	1.455	2.6	1.455

step 5) Good Turing with Laplace smoothed

	A	B	C	D	E
A	.026	.061	.088	.088	.046
B	.046	.044	.026	.061	.026
C	.044	.018	.026	.026	.026
D	.026	.044	.046	.026	.044
E	.046	.026	.026	.046	.026