**Assignment-2**

Sample Statements:

S1: The company chairman said he will increase the profit next year.

S2: The president said he believes the last year profit were good.

1. **Probability of two statements using Bigram Model Without Smoothing:**

**S1: The company chairman said he will increase the profit next year.**

P(S1) = P(company|The) \* P(chairman|company) \* P(said|chairman) \*

P(he|said) \* P(will|he) \* P(increase|will) \* P(the|increase) \*

P(profit|the) \* P(next|profit) \* P(year|next)

Calculating the individual Probabilities from given Corpus:

P(company|The) = C(The company)/C(The) = 75/1583 = .0473

P(chairman|company) = C(company chairman)/C(company) = 0/148 = 0.0

P(said|chairman) = C(chairman said)/C(chairman) = 7/571 = .0122

P(he|said) = C(said he)/C(said) = 18/281 = .0640

P(will|he) = C(he will)/C(he) = 7/121 = .0578

P(increase|will) = C(will increase)/C(will) = 0/115 = 0.0

P(the|increase) = C(increase the)/C(increase) = 0/7 = 0.0

P(profit|the) = C(the profit)/C(the) = 0/1583 = 0.0

P(next|profit) = C(profit next)/C(profit) = 0/10 = 0.0

P(year|next) = C(next year)/C(next) = 6/21 = .2857

Therefore, total probability of the statement **P(S1)= 0.0**

**S2: The president said he believes the last year profit were good.**

P(S2) = P(president|The) \* P(said|president) \* P(he|said) \* P(believes|he) \*

P(the|believes) \* P(last|the) \* P(year|last) \* P(profit|year) \* P(were|profit) \*

P(good|were)

Calculating the individual probabilities from the given Corpus:

P(president|The) = C(The president)/C(The) = 4/1583 = .0025

P(said|president) = C(president said)/C(president) = 0/97 = 0.0

P(he|said) = C(said he)/C(said) = 21/281 = .0747

P(believes|he) = C(he believes)/C(he) = 1/121 = 0082

P(the|believes) = C(believes the)/C(believes) = 1/1 = 1.0

P(last|the) = C(the last)/C(the) = 4/1583 = .0025

P(year|last) = C(last year)/C(last) = 4/33 = .1212

P(profit|year) = C(year profit)/C(year) = 0/59 = 0.0

P(were|profit) = C(profit were)/C(profit) = 0/10 = 0.0

P(good|were) = C(were good)/C(were) = 0/48 = 0.0

Therefore, total probability of the statement **P(S2)= 0.0**

**Tables with the bigram counts for the two sentences S1 and S2 for No-Smoothing.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | The | company | chairman | said | He | will | increase | profit | next | year |
| The | 0 | 86 | 37 | 0 | 0 | 0 | 0 | 0 | 5 | 3 |
| company | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| chairman | 0 | 0 | 1 | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| said | 32 | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 |
| He | 0 | 0 | 0 | 5 | 0 | 7 | 0 | 0 | 0 | 0 |
| will | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| increase | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| profit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| next | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| year | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | The | president | said | he | believes | last | year | profit | were | good |
| The | 0 | 4 | 0 | 0 | 0 | 4 | 3 | 0 | 0 | 0 |
| president | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Said | 32 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 |
| He | 0 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| believes | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Last | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |
| Year | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Profit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Were | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Good | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Individual Probabilities for Statement-1:**

1. **Without Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P((The,company)/(The)) | 0.054327227 |
| P((company,chairman)/(company)) | 0.0 |
| P((chairman,said)/(chairman)) | 0.04028021 |
| P((said,he)/(said)) | 0.07473309 |
| P((he,will)/(he)) | 0.05785124 |
| P((will,increase)/(will)) | 0.0 |
| P((increase,the)/(increase)) | 0.0 |
| P((the,profit)/(the)) | 0.0 |
| P((profit,next)/(profit)) | 0.0 |
| P((next,year)/(next)) | 0.2857143 |

**Final Proabillity:** 0.0

1. **With Add-One Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P((The,company)/(The)) | 0.013025902 |
| P((company,chairman)/(company)) | 1.9069413E-4 |
| P((chairman,said)/(chairman)) | 0.004235045 |
| P((said,he)/(said)) | 0.004091501 |
| P((he,will)/(he)) | 0.0015334483 |
| P((will,increase)/(will)) | 1.9190175E-4 |
| P((increase,the)/(increase)) | 1.9596316E-4 |
| P((the,profit)/(the)) | 1.4972301E-4 |
| P((profit,next)/(profit)) | 1.9584803E-4 |
| P((next,year)/(next)) | 0.0013679891 |

**Final Probability:** 1.0851242613297817E-31

1. **With Good-Turing Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P\*(0.0) | 0.006993007 |
| P\*(6.0) | 0.04895105 |
| P\*(7.0) | 0.055944055 |
| P\*(21.0) | 0.15384616 |
| P\*(23.0) | 0.16783217 |

**Final Probability:** 4.944721E-7

**Individual Probabilities for Statement-2:**

1. **Without Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P((The,president)/(The)) | 0.0025268476 |
| P((president,said)/(president)) | 0.020618556 |
| P((said,he)/(said)) | 0.07473309 |
| P((he,believes)/(he)) | 0.008264462 |
| P((believes,the)/(believes)) | 1.0 |
| P((the,last)/(the)) | 0.0025268476 |
| P((last,year)/(last)) | 0.121212125 |
| P((year,profit)/(year)) | 0.0 |
| P((profit,were)/(profit)) | 0.0 |
| P((were,good)/(were)) | 0.0 |

**Final Probability:** 0.0

1. **With Add-One Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P((The,president)/(The)) | 7.486151E-4 |
| P((president,said)/(president)) | 5.7770073E-4 |
| P((said,he)/(said)) | 0.004091501 |
| P((he,believes)/(he)) | 3.833621E-4 |
| P((believes,the)/(believes)) | 3.9238768E-4 |
| P((the,last)/(the)) | 7.486151E-4 |
| P((last,year)/(last)) | 9.748489E-4 |
| P((year,profit)/(year)) | 1.9398642E-4 |
| P((profit,were)/(profit)) | 1.9584803E-4 |
| P((were,good)/(were)) | 1.9440125E-4 |

**Final Probability:** 1.5652408326882403E-33

1. **With Good-Turing Smoothing:**

|  |  |
| --- | --- |
| **Bigram** | **Probability** |
| P\*(0.0) | 0.054054055 |
| P\*(1.0) | 0.0014609204 |
| P\*(2.0) | 3.5535902E-4 |
| P\*(4.0) | 1.6007163E-5 |

**Final Probability:** 1.6007163E-5