

Problem 2 - report

Pavle Marković

November 20, 2020

1 Analysis

As expected, for the unigram model there isn't any syntactical structure nor meaning. However, it is convenient to have it, to see the improvement of the bigram model. With the bigram model, word agreements can be noticed, like *Zack said* or *had to suggest*, but still, there are no clear sentences. Interestingly, the bigram model was the best in separating the title from the text, given that the corpus is reuters news. My idea was to surround titles with tokens for start and end, to help the model distinguish between two parts. Unfortunately, that didn't give any improvements. I believe that the only solution would be to have two language models, one for generating titles, and the other for generating texts.

With trigram and quadrigram models, syntactically correct sentences become to emerge. For instance, *I think we should have little impact on agricultural products, West German government is raising its total exports this year when prices plunged, or In future banks will be able to bear such a burden ,"* he told. Moreover, these sentences are also meaningful by itself, full context for a text is still far away. However, the trade-off here is that these sentences are kind of similar to the original ones from the corpus. In addition, there some subtleties to be resolved, like possessive forms (*he would accept Ghana offers*) which might be solved with advanced tokenization (not separating 's). Finally, the length of the sentences and punctuation are nowhere near consistent.

2 Generated texts

- Unigram model:

- form 4 Cyclops by negotiations - 527 unfocused This trade profit time - will May to Purchase harm tonnes market the describes action and equivalents the which it Calmat 5 ' Sales 704 as of also contracts posture 650 of had which mile 99 route been in total . FROST billion U > at - It estimated Corp January security under stocks 3 to pct current coming Hicks vs the January be , . share s as . offering CENTRAL strong pct he profit but 92 exporting , Allied Limited per - board participated > about a 2 said .

- Bigram model:

- ANOTHER NEW OFFER The deal would adversely affected Kenya AB Providentia to put out . S . The thrift said the previous week . Zack said it believes it repositioned

the profit 78 billion Belgian Finance Ministry said . 9 . 4 mln Canadian) said he expected ." Interlink Express . Kuranari said . Agriculture Department of oxide to the Pacific Air Freight costs for power for each year contract between there were slightly less and the market as saying that a third largest single - Warner that took up from time to help prop up in Adobe Resources

- JORDAN BRANCH FOR CASH DISTRIBUTION MONTHLY PAYOUT Qtly div six months of 184 . 9 , 935 , efforts to retaliate with Bankers Fidelity , 280 , 000 dlrs , 303 , had to suggest that is already receiving 25 mln vs 43 mln Net 34 wholesale changes have a key to the year end of unit , however , with the United States was priced around 145 , it with those countries to holders , which was estimated to another 7 , but exports in yen level , other than a paring of return on the election to 300
- Trigram model:
 - UPWARD PRESSURE ON INTEREST RATES A tight money policy , they said . " I think we should have little impact on agricultural products . INDONESIA PRESSING FOR COMPROMISE OVER COFFEE QUOTA TALKS CONTINUE , OUTCOME HARD TO GAUGE Talks at the plant is one option to buy gold , he said . " Non - currency reserves were 335 . Volume midmorning was 730 lots including 350 lots crossed and 108 switches . There were no bales awaiting decertification . REICHHOLD It ; PCI > . In Fairfield , N . Y ., and Allied - Lyons plans to
 - INTERVENTION West German government is raising its total exports this year when prices plunged . The agency said . Damages was estimated at 15 dlrs a troy ounce of 24 . 8 mln Avg shrs primary 6 , 913 , 000 Revs 32 . 5 mln Tax 14 . 7 mln Net interest payable on completion of the meetings in Washington has been deposited under its Galaxy III satellite due to impact of its unbleached coated paperboard was particularly bad in the week to yesterday were isolated and weak prices of coconut oil worth about 319 mln dlrs . Chartered
- Quadrigram model:
 - A recent study by the non - oil exports to Europe and Canada . The communique did not identify , as saying despite the drought and the shortage , the total inventory at the end of the year . For full - year earnings include gains from the sale to Shearson Lehman Brothers that the price of gas – to 2 . 5 to 9 . 223 tonnes last year from 543 mln in 1985 . Fekete said the rates would come into force . In future , banks will be able to bear such a burden , " he told
 - FRANCS The French government is to sell its Celeron Corp unit . Goodyear said the price changes , coupled with a lack of hard currencies limited Heineken ' s market share after last year ' s record fall in Wall Street stock prices , he would accept Ghana offers up to five pct . Earlier today , Sun Co It ; AHLN . AS > is discussing marketing cooperation with U . S . Agriculture Department ' s latest strong signal to the market , however , fell 0 . 2 pct in February , vs a revised 471