ANLP Assignment 1

Chunchuan Lyn   Yue Yu
(please fill in)   s1563228

October 19, 2015

1   Perplexity of the Test Case

$$
\begin{aligned}
PP_M(\vec{w}) &= 2^{H_M(\vec{w})} \\
&= 2^{-\frac{1}{n}\log_2 P_M(\vec{w})} \\
&= 2^{\log_2 P_M(\vec{w})^{-\frac{1}{n}}} \\
&= P_M(\vec{w})^{-\frac{1}{n}} \\
&\approx \prod_{i=1}^{n} P(w_n|w_{n-1}, w_{n-2})^{-\frac{1}{n}} \\
&= (0.2 * 0.7 * 0.6 * 0.25 * 0.5 * 0.1)^{-\frac{1}{6}} \\
&\approx 3.1367
\end{aligned}
$$

p.s.: w_{-1} and w_0 refer to the first two '[' characters of each sentence.

2   Line Preprocessing

```python
#function turns input into required format
def preprocess_line(line):
    #remove non-necessary characters,
    #and turn string to lowercase
    p = re.compile('[^\w\s,.]')
    line = re.sub(p,'',line.lower())
    #replace \n by ]
    line = re.sub('\n',']',line)
    #turn numbers into 0
```

```
line = re.sub('[0-9]','0',line)
#add  begining  and  end  [[
return '[['+line
```

By preprocessing input in this fashion, we essentially assumed that there are no interconnection between lines. All the lines are preprocessed into line units. During language model building, we will not compute $P(]|[[)$ nor $P([| * [)$. This probability will be equal to one, if we treat the whole text as one unit. We consider this as an artifact of ngram model instead of the true underlying language model.

As a consequence, we will sample line by line independently in task 4. Also, we will exclude those probability in computing perplexity.

3 Language Model

3.1 Estimation of Probabilities

3.2 Data Structure

3.3 Excerpt? Please refine

4 Random Output

4.1 English

[[thaveriatallic res, your eurgieve of the bill eur ack oural for condmend mor ast way whis degionceneent, me, as councify it rannies and weelater reffer cludreng this.]][[mr proadoet is wition inum ruciploch i wordiers whas ficlents mustwenionshou sup wing safereag weas arto mit arlikend the which al

4.2 German

[[antwichern aspolluntwischerlar, sozip innie wordie fr hinbit ungspollen, dannglichiesion nicht ung ungsatur diellerung dasprften be.]][[ickleinntionten deben vie dies fe, deruntrograts lag habfahmei gem die proziegle aucht immungliefftschkellettberkat.]][[gleitglich er zusweitiger ver unglck sidend si

4.3 Spanish

[[peduras nos elasa pente, la te los emasotras entaciorionos porionesinten las pre en su de a pros supor los, es de las na lostacel da clariondo la consa y que los prostacue mientamos tre, nos pblidas aganda lormaados seorque que sin de que de lasido y ta uniconesi el sen me eura sonesta inforgo par

5 Perplexity & Language Identification

6 Objective

To determine the atomic weight of magnesium via its reaction with oxygen and to study the stoichiometry of the reaction (as defined in 6.1):

$$2\mathrm{Mg} + \mathrm{O_2} \longrightarrow 2\mathrm{MgO}$$
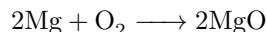
6.1 Definitions

Stoichiometry The relationship between the relative quantities of substances taking part in a reaction or forming a compound, typically a ratio of whole integers.

Atomic mass The mass of an atom of a chemical element expressed in atomic mass units. It is approximately equivalent to the number of protons and neutrons in the atom (the mass number) or to the average number allowing for the relative abundances of different isotopes.

7 Experimental Data

Mass of empty crucible	$7.28\,\mathrm{g}$
Mass of crucible and magnesium before heating	$8.59\,\mathrm{g}$
Mass of crucible and magnesium oxide after heating	$9.46\,\mathrm{g}$
Balance used	#4
Magnesium from sample bottle	#1

8 Sample Calculation

Mass of magnesium metal	$= 8.59\,\mathrm{g}$ - $7.28\,\mathrm{g}$
	$= 1.31\,\mathrm{g}$
Mass of magnesium oxide	$= 9.46\,\mathrm{g}$ - $7.28\,\mathrm{g}$
	$= 2.18\,\mathrm{g}$
Mass of oxygen	$= 2.18\,\mathrm{g}$ - $1.31\,\mathrm{g}$
	$= 0.87\,\mathrm{g}$

Because of this reaction, the required ratio is the atomic weight of magnesium: $16.00\,\mathrm{g}$ of oxygen as experimental mass of Mg: experimental mass of oxygen or $\frac{x}{1.31} = \frac{16}{0.87}$ from which, $M_{\mathrm{Mg}} = 16.00 \times \frac{1.31}{0.87} = 24.1 = 24\,\mathrm{g\,mol^{-1}}$ (to two significant figures).

9 Results and Conclusions

The atomic weight of magnesium is concluded to be $24\,\mathrm{g\,mol^{-1}}$, as determined by the stoichiometry of its chemical combination with oxygen. This result is in

agreement with the accepted value.

Figure 1: Figure caption.

10   Discussion of Experimental Uncertainty

The accepted value (periodic table) is $24.3\,\mathrm{g\,mol^{-1}}$ Smith and Jones (2012). The percentage discrepancy between the accepted value and the result obtained here is 1.3%. Because only a single measurement was made, it is not possible to calculate an estimated standard deviation.

The most obvious source of experimental uncertainty is the limited precision of the balance. Other potential sources of experimental uncertainty are: the reaction might not be complete; if not enough time was allowed for total oxidation, less than complete oxidation of the magnesium might have, in part, reacted with nitrogen in the air (incorrect reaction); the magnesium oxide might have absorbed water from the air, and thus weigh "too much." Because the result obtained is close to the accepted value it is possible that some of these experimental uncertainties have fortuitously cancelled one another.

11   Answers to Definitions

a. The *atomic weight of an element* is the relative weight of one of its atoms compared to C-12 with a weight of 12.0000000..., hydrogen with a weight of 1.008, to oxygen with a weight of 16.00. Atomic weight is also the average weight of all the atoms of that element as they occur in nature.

b. The *units of atomic weight* are two-fold, with an identical numerical value. They are g/mole of atoms (or just g/mol) or amu/atom.

c. *Percentage discrepancy* between an accepted (literature) value and an experimental value is

$$\frac{\text{experimental result} - \text{accepted result}}{\text{accepted result}}$$

References

Smith, J. M. and Jones, A. B. (2012). *Chemistry*. Publisher, 7th edition.