NLP 1 - Assignment 4

Selene Baez Santamaria

December 15, 2016

Exercise 1. PMI

(a) For each example, explain what P(x,y) represents.

- P(x,y) represents the probability of the bi-gram "eat pizza".
- P(x,y) represents the probability of a Tweet containing "happy" and "pizza".

(b) What do negative, zero and positive PMI's mean?
Defining PMI as $\log_2(r)$ where $r = \frac{P(x,y)}{P(x)P(y)} = \frac{P(x|y)}{P(x)} = \frac{P(y|x)}{P(y)}$

- If PMI $< 0 \rightarrow r < 1 \rightarrow P(x,y) < P(x)P(y)$. This means that the actual co-occurrence probability is less than their predicted co-ocurrence probability if they were independent. Therefore, lower PMIs correspond to co-occurrences that are less likely. Example: "he you" has a low association and is very unlikely to co-occur.
- If PMI $= 0 \rightarrow r = 1 \rightarrow P(x,y) = P(x)P(y)$. This means that the two events are statistically independent.
- If PMI $> 0 \rightarrow r > 1 \rightarrow P(x,y) > P(x)P(y)$. This means that the actual co-occurrence probability is higher than their predicted co-occurrence probability if they were independent.Therefore, higher PMIs correspond to co-occurrences that are more likely. Example: "New York" has a high association and is very likely to co-occur.

Exercise 2. MaxEnt

(a) Simplified expression for $\log P(y|\vec{x})$

$$
\begin{aligned}
\log P(y|\vec{x}) &= \log \frac{exp(\sum_1 w_i f_i(\vec{x}, y))}{\sum_{y'} exp(\sum_1 w_i f_i(\vec{x}, y')))} \\
&= \log exp(\sum_1 w_i f_i(\vec{x}, y)) - \log \sum_{y'} exp(\sum_1 w_i f_i(\vec{x}, y'))) \\
&= \sum_1 w_i f_i(\vec{x}, y) - C = \vec{w} \cdot \vec{f} - C
\end{aligned}
$$

The end result is a linear function of the feature vector.

(b) Which sense is the most probable?

For $y = \mathit{1}$, features 1 and 7 are active

$$p(y=1|\vec{x}) = \frac{\exp(2.0-0.1)}{\exp(2.0-0.1)+\exp(1.8+1.1)+\exp(0.3+2.7)} \tag{1}$$
$$= \frac{\exp(1.9)}{\exp(1.9)+\exp(2.9)+\exp(3)} \tag{2}$$
$$= \frac{6.685}{44.945} = 0.149 \tag{3}$$

For $y = \mathit{2}$, features 2 and 8 are active

$$p(y=1|\vec{x}) = \frac{\exp(2.9)}{\exp(1.9)+\exp(2.9)+\exp(3)} \tag{4}$$
$$= \frac{18.174}{44.945} = 0.404 \tag{5}$$

For $y = \mathit{3}$, features 3 and 9 are active

$$p(y=1|\vec{x}) = \frac{\exp(3)}{\exp(1.9)+\exp(2.9)+\exp(3)} \tag{6}$$
$$= \frac{20.085}{44.945} = 0.447 \tag{7}$$

Hence, y=*3* = *Noun: a factory* is the most probable.

Exercise 3. FOL to Natural language

(a) "Every bear is furry."

(b) "Jan helps Joost"

(c) "Sergii eats pizza"

(d) "Sergii eats pizza with a fork"

(e) "Every student lifts Marie, but not necessarily together"

(f) "All students lift Marie, together"

Exercise 4. Natural language to FOL

(a) $\exists e.x.\ hating(e) \wedge pasta(x) \wedge hater(e, Juan) \wedge hatee(e, x)$

(b) $\exists x.\ Student(x) \wedge \forall(y)Class(y) \wedge \exists e.\ liking(e) \wedge liker(e, x) \wedge likee(e, y)$

(c) $\exists e.x.\ seeing(e) \wedge seer(e, Marie) \wedge seen(e, Marie)$

Exercise 5. Grammar with semantic attachments

(a) *Whiskers likes Sam*

Rule applied	**Result**
$S \rightarrow NPVP$	$VP.sem(NP.sem)$
$VP \rightarrow V_t NP$	$V_t.sem(NP.sem)(NP.sem)$
$NP \rightarrow N$	$V_t.sem(N.sem)(NP.sem)$
$NP \rightarrow N$	$V_t.sem(N.sem)(N.sem)$
$V_t \rightarrow likes$	$\lambda x.\lambda y.\,(\exists e.\ Liking(e) \wedge Liker(e, y) \wedge Likee(e, x))\,(N.sem)(N.sem)$
$N \rightarrow Sam$	$\lambda x.\lambda y.\,(\exists e.\ Liking(e) \wedge Liker(e, y) \wedge Likee(e, x))\,(Sam)(N.sem)$
$N \rightarrow Whiskers$	$\lambda x.\lambda y.\,(\exists e.\ Liking(e) \wedge Liker(e, y) \wedge Likee(e, x))\,(Sam)(Whiskers)$

(b) *A cat meows*
The MR is: $\lambda x.\,(\exists e.Meowing(e) \wedge Meower(e, x))$

We face the problem that we cannot specify that x must be a cat. Additionally, "A cat" represents an existential quantifiers to be combined in the NP. We try to solve it by adding these rules:

- $NP \rightarrow DetN$ $\quad\lambda x.\ \ (\exists x.Cat(x))$
- $N \rightarrow Cat$ $\quad\lambda x.\ Cat(x)$
- $V_i \rightarrow meows$ $\quad\lambda x.\,(\exists(e)Meowing(e) \wedge Meower(e, x) \wedge Cat(x))$

Rule applied	**Result**
$S \rightarrow NPVP$	$VP.sem(NP.sem)$
$VP \rightarrow V_i$	$V_t.sem(NP.sem)$
$V_t \rightarrow meows$	$\lambda x.\,(\exists e.\ Meowing(e) \wedge Meower(e, x) \wedge Cat(x))\,(NP.sem)$
$NP \rightarrow DetN$	$\lambda x.\,(\exists e.\ Meowing(e) \wedge Meower(e, x) \wedge Cat(x))\,(Det.sem)(N.sem)$
$N \rightarrow Cat$	$\lambda x.\,(\exists e.\ Meowing(e) \wedge Meower(e, x) \wedge Cat(x))\,(\exists x.)(N.sem)$
$Det \rightarrow A$	$\lambda x.\,(\exists e.\ Meowing(e) \wedge Meower(e, x) \wedge Cat(x))\,(\exists x.)(Cat(x))$

Exercise 6. IBM Model

(a) Assumptions:

- All alignments are equally likely
- Words in a translation are only dependent on their alignment to the source language, but independent on other translated words surrounding them.
- Some word may not have direct translations, so they introduce the token "NULL"