APPRENTISSAGE NON SUPERVISÉ

Cours 2
Caio Corro

EXTRACTION D'UN DICTIONNAIRE DE TRADUCTION

Objectif

Extraire automatiquement à partir d'un corpus de textes un dictionnaire de traduction de mots, comme dans l'exemple ci-dessous :

Français	Anglais
cuisine	cooks, cook, cooking, kitchen, food...
comprendre	understand, understood, understanding, realize, ...
être	human, being, be, ...
économie	free-market, Economy, economics, market,

Données

Phrases parallèles :

- Une phrase dans la langue source (Français)
- Sa traduction dans une langue cible (Anglais)

CORPUS : DE LA TERRE À LA LUNE (JULES VERNE)

Extrait de la partie en Français

(1) Pendant la guerre fédérale des États-Unis , un nouveau club très influent s' établit dans la ville de Baltimore , en plein Maryland .

(2) On sait avec quelle énergie l' instinct militaire se développa chez ce peuple d' armateurs , de marchands et de mécaniciens .

(3) De simples négociants enjambèrent leur comptoir pour s' improviser capitaines , colonels , généraux , sans avoir passé par les écoles d' application de West-Point [École militaire des États-Unis .] ;

Extrait de la partie en Anglais

(1) During the War of the Rebellion , a new and influential club was established in the city of Baltimore in the State of Maryland .

(2) It is well known with what energy the taste for military matters became developed among that nation of ship-owners , shopkeepers , and mechanics .

(3) Simple tradesmen jumped their counters to become extemporized captains , colonels , and generals , without having ever passed the School of Instruction at West Point ;

CORPUS : DE LA TERRE À LA LUNE (JULES VERNE)

Extrait de la partie en Français

(1) Pendant la guerre fédérale des États-Unis , un nouveau club très influent s' établit dans la ville de Baltimore , en plein Maryland .

(2) On sait avec quelle énergie ' instinct militaire se développa chez ce peuple d' armateurs , de marc ??? ls et de mécaniciens .

(3) De simples négociants enjambèrent leur comptoir pour s' improviser capitaines , colonels , généraux , sans avoir passé par les écoles d' application de West-Point [École militaire des États-Unis .] ;

Extrait de la partie en Anglais

(1) During the War of the Rebellion , a new and influential club was established in the city of Baltimore in the State of Maryland .

(2) It is well known with what energy the taste for military matters became developed among that nation of ship-owners , shopkeepers , and mechanics .

(3) Simple tradesmen jumped their counters to become extemporized captains , colonels , and generals , without having ever passed the School of Instruction at West Point ;

CORPUS : DE LA TERRE À LA LUNE (JULES VERNE)

Extrait de la partie en Français

(1) Pendant la guerre fédérale des États-Unis , un nouveau club très influent s' établit dans la ville de Baltimore , en plein Maryland .

(2) On sait avec quelle énergie l' instinct militaire se développa chez ce peuple d' armateurs , de marchands et de mécaniciens .

(3) De simples négociants enjambèrent leur comptoir pour s' improviser capitaines , colonels , généraux , sans avoir passé par les écoles d' application de West-Point [École militaire des États-Unis .] ;

Extrait de la partie en Anglais

(1) During the War of the Rebellion , a new and influential club was established in the city of Baltimore in the State of Maryland .

(2) It is well known with what energy the taste for military matters became developed among that nation of ship-owners , shopkeepers , and mechanics .

(3) Simple tradesmen jumped their counters to become extemporized captains , colonels , and generals , without having ever passed the School of Instruction at West Point ; ???

INTUITION

Hypothèse distributionnelle

Le sens d'un mot peut être déduit du contexte dans lequel il est utilisé,
si deux mots sont synonymes, alors ils seront utilisés dans les mêmes contextes.

Dans le cas bilingue

Un mot Français et un mot Anglais qui co-occurrent souvent (c'est-à-dire qu'ils apparaissent conjointement dans la phrase source et la traduction cible) sont probablement des traductions.

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

INTUITION

Français	Grec
Je vais chez nous (lit. dans notre maison).	Πάω στο σπίτι μας.
Ma maison est grande.	Το σπίτι μου είναι μεγάλο.
La maison d’Elli est à côté de la plage.	Το σπίτι της 'Ελλης είναι κοντά στην παραλία.
Je m’appelle Elli.	Με λένε 'Ελλη.
Une maison du village a brûlé.	Ενα σπίτι του χωριού κάηκε
J’aime Elli.	Αγαπώ την 'Ελλη.

Questions

- Quelle est la traduction de **maison** en Grec ?
- Quelle est la traduction de **Elli** en Grec ?

INTUITION

Français	Grec
Je vais chez nous (lit. dans notre maison).	Πάω στο σπίτι μας.
Ma maison est grande.	Το σπίτι μου είναι μεγάλο.
La maison d’Elli est à côté de la plage.	Το σπίτι της Έλλης είναι κοντά στην παραλία.
Je m’appelle Elli.	Με λένε Έλλη.
Une maison du village a brûlé.	Ενα σπίτι του χωριού κάηκε
J’aime Elli.	Αγαπώ την Έλλη.

Questions

- Quelle est la traduction de **maison** en Grec ?
- Quelle est la traduction de **Elli** en Grec ?

INTUITION

Français	Grec
Je vais chez nous (lit. dans notre maison).	Πάω στο σπίτι μας.
Ma maison est grande.	Το σπίτι μου είναι μεγάλο.
La maison d'Elli est à côté de la plage.	Το σπίτι της 'Ελλης είναι κοντά στην παραλία.
Je m'appelle Elli.	Με λένε 'Ελλη.
Une maison du village a brûlé.	Ενα σπίτι του χωριού κάηκε
J'aime Elli.	Αγαπώ την 'Ελλη.

Questions

- Quelle est la traduction de **maison** en Grec ?
- Quelle est la traduction de **Elli** en Grec ?

PREMIÈRE IDÉE

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

Table de co-occurrences

- On peut compter les co-occurences de mots Français/Anglais :
  - Combien de fois le mot "cuisine" apparait de façon parallèle avec le mot "boiling" ?
  - Combien de fois le mot "cuisine" apparait de façon parallèle avec le mot "kitchen" ?
  - etc etc
- Si deux mots apparaissent souvent ensembles, ceux sont peut être des traductions ?

TABLE DE CO-OCCURRENCE

Structure du tableau

- Une ligne par mot dans le vocabulaire en Français
- Une colonne par mot dans le vocabulaire en Anglais

	dog	kitchen	room	etc etc
le	0	0	0	...
la	0	0	0	...
chien	0	0	0	...
cuisine	0	0	0	...
etc etc	...	...	..	..

Comment remplir le tableau ?

1. On initialise toutes les cellules du tableau à 0
2. On parcours toutes les phrases parallèles du corpus
   - Pour chaque couple de mot Français/Anglais, on incrément la cellule correspondante
   - Si un mot apparait plus d'une fois dans une phrase, on ne compte qu'une occurrence

EXEMPLE

	dog	kitchen	the	etc etc
le	0	0	0	…
la	0	0	0	…
chien	0	0	0	…
cuisine	0	0	0	…
etc etc	…	…	..	..

Le chien est dans la cuisine / The dog is in the kitchen

EXEMPLE

	dog	kitchen	the	etc etc
le	0	0	+1	...
la	0	0	0	...
chien	0	0	0	...
cuisine	0	0	0	...
etc etc	...	...	..	..

Le chien est dans la cuisine / The dog is in the kitchen

- On incrémente la cellule correspondant à "le" / "the"

EXEMPLE

	dog	kitchen	the	etc etc
le	0	0	+1	...
la	0	0	0	...
chien	0	0	+1	...
cuisine	0	0	0	...
etc etc	...	...	..	..

Le chien est dans la cuisine / The dog is in the kitchen

- On incrémente la cellule correspondant à "le" / "the"
- On incrémente la cellule correspondant à "chien" / "the"

EXEMPLE

	dog	kitchen	the	etc etc
le	0	0	+1	...
la	0	0	0	...
chien	+1	0	+1	...
cuisine	0	0	0	...
etc etc	...	...	..	..

Le chien est dans la cuisine / The dog is in the kitchen

- On incrémente la cellule correspondant à "le" / "the"
- On incrémente la cellule correspondant à "chien" / "the"
- On incrémente la cellule correspondant à "chien" / "dog"

EXEMPLE

	dog	kitchen	the	etc etc
le	0	0	+1	...
la	0	0	0	...
chien	+1	0	+1	...
cuisine	0	0	0	...
etc etc	...	...	..	..

Le chien est dans la cuisine / The dog is in the kitchen

- On incrémente la cellule correspondant à "le" / "the"
- On incrémente la cellule correspondant à "chien" / "the"
- On incrémente la cellule correspondant à "chien" / "dog"
- etc etc pour tous les couples de mots Français/Anglais dans les deux phrases

EN PYTHON

```python
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

EN PYTHON

```
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

```
txt1 = set(txt1)
txt2 = set(txt2)

print(txt1)
print(txt2)
```

Pour ne compter qu'une fois chaque mot

```
{'dans', '.', 'la', 'chien', 'est', 'le', 'cuisine'}
{'the', 'kitchen', '.', 'in', 'is', 'dog'}
```

EN PYTHON

```
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

```
txt1 = set(txt1)
txt2 = set(txt2)

print(txt1)
print(txt2)
```

```
{'dans', '.', 'la', 'chien', 'est', 'le', 'cuisine'}
{'the', 'kitchen', '.', 'in', 'is',
```

On utilise un counter car :
1. le tableau est très parcimonieux (beaucoup de 0)
2. La plupart des cellules ne seront jamais incrémentées

```
counter = collections.Counter()

for w1 in txt1:
    for w2 in txt2:
        counter[(w1, w2)] += 1

print(counter)
```

```
Counter({('dans', 'the'): 1, ('dans', 'kitchen'): 1, ('dans
', '.'): 1, ('dans', 'in'): 1, ('dans', 'is'): 1, ('dans',
'dog'): 1, ('.', 'the'): 1, ('.', 'kitchen'): 1, ('.',
'.'): 1, ('.', 'in'): 1, ('.', 'is'): 1, ('.', 'dog'): 1,
('la', 'the'): 1, ('la', 'kitchen'): 1, ('la', '.'): 1, ('l
a', 'in'): 1, ('la', 'is'): 1, ('la', 'dog'): 1, ('chien',
```

EN PYTHON

```
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

```
txt1 = set(txt1)
txt2 = set(txt2)

print(txt1)
print(txt2)
```

```
{'dans', '.', 'la', 'chien', 'est', 'le', 'cuisine'}
{'the', 'kitchen', '.', 'in', 'is', 'dog'}
```

Mot en Français

Mot en Anglais

```
counter = 

for w1 in txt1:
    for w2 in txt2:
        counter[(w1, w2)] += 1

print(counter)
```

```
Counter({('dans', 'the'): 1, ('dans', 'kitchen'): 1, ('dans
', '.'): 1, ('dans', 'in'): 1, ('dans', 'is'): 1, ('dans',
'dog'): 1, ('.', 'the'): 1, ('.', 'kitchen'): 1, ('.',
'.'): 1, ('.', 'in'): 1, ('.', 'is'): 1, ('.', 'dog'): 1,
('la', 'the'): 1, ('la', 'kitchen'): 1, ('la', '.'): 1, ('l
a', 'in'): 1, ('la', 'is'): 1, ('la', 'dog'): 1, ('chien',
```

EN PYTHON

```
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

```
txt1 = set(txt1)
txt2 = set(txt2)

print(txt1)
print(txt2)
```

```
{'dans', '.', 'la', 'chien', 'est', 'le', 'cuisine'}
{'the', 'kitchen', '.', 'in', 'is', 'dog'}
```

```
counter = collections.Counter()

for w1 in txt1:
    for w2 in txt2:
        counter[(w1, w2)] += 1
```

Mot en Français

Mot en Anglais

```
n'): 1, ('dans
', '.'): 1, ('dans', 'in'): 1, ('dans', 'is'): 1, ('dans',
'dog'): 1, ('.', 'the'): 1, ('.', 'kitchen'): 1, ('.',
'.'): 1, ('.', 'in'): 1, ('.', 'is'): 1, ('.', 'dog'): 1,
('la', 'the'): 1, ('la', 'kitchen'): 1, ('la', '.'): 1, ('l
a', 'in'): 1, ('la', 'is'): 1, ('la', 'dog'): 1, ('chien',
```

EN PYTHON

```
txt1 = "le chien est dans la cuisine ."
txt2 = "the dog is in the kitchen . "

txt1 = txt1.split()
txt2 = txt2.split()

print(txt1)
print(txt2)
```

```
['le', 'chien', 'est', 'dans', 'la', 'cuisine', '.']
['the', 'dog', 'is', 'in', 'the', 'kitchen', '.']
```

```
txt1 = set(txt1)
txt2 = set(txt2)

print(txt1)
print(txt2)
```

```
{'dans', '.', 'la', 'chien', 'est', 'le', 'cuisine'}
{'the', 'kitchen', '.', 'in', 'is', 'dog'}
```

```
counter = collections.Counter()

for w1 in txt1:
    for w2 in txt2:
        counter[(w1, w2)] += 1

print(counter)
```

```
Counter({('dans', 'the'): 1, ('dans', 'kitchen'): 1, ('dans
', '.'): 1, ('dans', 'in'): 1, ('dans', 'is'): 1, ('dans',
'dog'): 1, ('.', 'the'): 1, ('.', 'kitchen'): 1, ('.',
'.'): 1, ('.', 'in'): 1, ('.', 'is'): 1, ('.', 'dog'): 1,
('la', 'the'): 1, ('la', 'kitchen'): 1, ('la', '.'): 1, ('l
a', 'in'): 1, ('la', 'is'): 1, ('la', 'dog'): 1, ('chien',
```

UNE FAUSSE BONNE IDÉE ?

Table de co-occurrence

1. On compte les co-occurences de mots Français/Anglais
2. Si deux apparaissent souvent ensembles, ceux sont peut être des traductions ?

Les mots grammaticaux, ou mots outils (le, de, et, mais, etc), et la ponctuation dans une lange vont co-occurer très souvent avec les mots de l'autre lange !

```
with open("./french.corpus") as src_instream:
    with open("./english.corpus") as tgt_instream:
        cooc = build_cooc_table(src_instream, tgt_instream)

cooc.most_common(100)
```

```
[(('.', '.'), 1437),
 ((',', ','), 1243),
 ((',', '.'), 1211),
 (('.', 'the'), 1088),
 (('.', ','), 1084),
 ((',', 'the'), 1052),
 (('de', '.'), 946),
 (('.', 'of'), 936),
 ((',', 'of'), 905),
 (('de', 'the'), 902),
 (('de', 'of'), 877),
 (('de', ','), 872),
```

UN PEU DE PROBAS

Variable aléatoire discrète

Définition

Une variable aléatoire X est dite discrète si et seulement si son espace des réalisations est dénombrable et fini, c'est à dire qu'il peut être réduit à :

$$\mathcal{X} = \{1, 2, \ldots, n\}, \quad \text{avec } n \in \mathbb{N} \setminus \{0\}.$$

Note : $n \neq \infty$.

Exemples

- Observation d'un lancer de dé : $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$.
- Observation d'un mot d'une phrase : $\mathcal{X} = \{\text{"maison"}, \text{"un"}, \text{"rouge"}, \ldots\}$.
- Observation de la météo :
  $\mathcal{X} = \{\text{"soleil sans pluie"}, \text{"soleil avec pluie"}, \text{"pluie sans soleil"}, \text{"brume"}, \ldots\}$

Note : Il y a toujours une part de modélisation à faire dans le choix de l'ensemble des observations. Les observations sont **par construction** exclusives. Après un lancer de dé, on ne peut pas observer à la fois "1" et "2".

Distribution sur un VA discrète

Supposons un VA discrète avec l'espace des réalisations $\mathcal{X} = \{1, 2, \ldots, n\}$.
On peut noter :

- La probabilité d'observer $X = 1$: $P(X = 1) = \theta_1$
- La probabilité d'observer $X = 2$: $P(X = 2) = \theta_2$
- ...
- La probabilité d'observer $X = n$: $P(X = n) = \theta_n$

Quelles propriétés doivent satisfaire $\theta_1, \ldots, \theta_n$?

- Non-négativité : $p_i \geq 0, \forall i \in \{1, \ldots, n\}$
- Somme à 1 : $\sum_{i=1}^{n} \theta_i = 1$

Note :

- On peut en déduire que : $0 \leq \theta_i \leq 1, \forall i \in \{1, \ldots, n\}$
- $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$ (ou n'importe quel autre indice)

Simplexe

$$p(X = x) = \theta_x \qquad \Leftrightarrow \qquad \theta = \begin{array}{|c|}\hline \theta_1 \\ \hline \theta_2 \\ \hline \theta_3 \\ \hline \theta_4 \\ \hline \theta_5 \\ \hline \end{array}$$

Définition

Le simplexe de dimension $n-1$, noté $\Delta(n)$, est un ensemble des vecteurs de dimension n défini de la façon suivante :

$$\Delta(n) = \Delta^n = \{\boldsymbol{v} \in \mathbb{R}^n \text{ tel que } \forall 1 \leq i \leq n : v_i \geq 0 \text{ et } \sum_{i=1}^{n} v_i = 1\}$$

On pourra donc écrire $\theta \in \Delta(n)$ l'ensemble des paramétrisations possibles pour une distribution discrète dont l'espace des réalisations est de taille n.

Variables aléatoires jointes

Motivations

Parfois, il peut être utile de considérer un tuple (p. ex. un couple) de variables aléatoires discrètes.

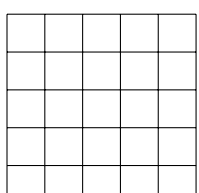
Par exemple :

- $\mathcal{X} = \{\text{"chaud"}, \text{"froid"}, \text{"glaciale"}\}$,
- $\mathcal{Y} = \{\text{"pas de nuage"}, \text{"nuageux"}, \text{"pluie"}\}$,

Distribution jointe

- $\forall x \in \mathcal{X}, Y \in \mathcal{Y} \ : P(X = x, Y = y) \geq 0$
- $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(X = x, Y = y) = 1$

$\theta =$ $p_\theta(X, Y)$



Variables aléatoires binaires

Définition

Une variable variables aléatoire X est dite binaire si et seulement si son espace des réalisations peut être réduit à :

$$\mathcal{X} = \{0, 1\}$$

Exemples

- Jeu pile ou face :

$$\mathcal{X} = \{\text{"pile"}, \text{"face"}\} \quad \Leftrightarrow \quad \mathcal{X} = \{0, 1\}$$

- Lancer d'un dé avec observation si résultat pair ou impair :

$$\mathcal{X} = \{\text{"pair"}, \text{"impair"}\} \quad \Leftrightarrow \quad \mathcal{X} = \{0, 1\}$$

Distribution sur un VA binaire

Supposons un VA binaire avec l'espace des réalisations $\mathcal{X} = \{0, 1\}$
(p. ex. le résultat d'un lancer de pièce).
On peut noter :

- La probabilité d'observer $X = 0$: $P(X = 0) = p_1$
- La probabilité d'observer $X = 1$: $P(X = 1) = p_2$

Quelles propriétés doivent satisfaire p_1 et p_2 ?

- Non-négativité : $p_1 \geq 0$ et $p_2 \geq 0$
- Somme à 1 : $P(X = 0) + P(X = 1) = p_1 + p_2 = 1$

Note : on peut en déduire que :

- $0 \leq p_1 \leq 1$ et $0 \leq p_2 \leq 1$
- $p2 = 1 - p1$

Loi de Bernoulli

Définition

Un VA binaire suit la loi de Bernoulli paramétrisée par $\mu \in [0, 1]$ si :

- $P(X = 1) = \mu$
- $P(X = 0) = 1 - \mu$

On utilisera aussi les notations suivantes pour marquer la paramétrisation :

$$P(X = x|\mu) \qquad P(X = x; \mu) \qquad P_\mu(X = x),$$

et pour indiquer que X suit une loi de Bernoulli paramétrisé par μ :

$$X \sim \mathcal{B}(\mu)$$

Remarque

$$P(X = x) = \begin{cases} \mu & \text{si } x = 1, \\ 1 - \mu & \text{sinon.} \end{cases} \qquad \Leftrightarrow \qquad P(X = x) = \mu^x (1 - \mu)^{1-x}$$

Notations

Pour simplifier, nous utiliserons des notations abrégées :

- $P(X = x) \quad \Rightarrow \quad P(x)$
- $P(X = x, Y = y) \quad \Rightarrow \quad P(x, y)$
- ...

Probabilité marginale

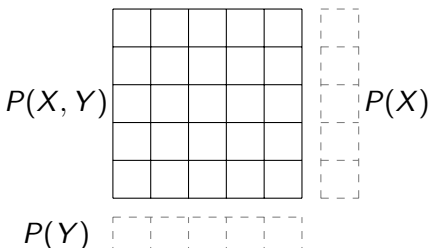
Soit deux variables aléatoires $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe. La probabilité $P(x)$ est définie par :

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y)$$

Probabilité marginale

Soit deux variables aléatoires $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe. La probabilité $P(x)$ est définie par :

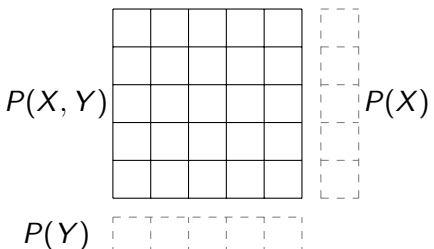
$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y)$$



Probabilité marginale

Soit deux variables aléatoires $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe. La probabilité $P(x)$ est définie par :

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y)$$



Soit $Z \in \mathcal{Z}$ une troisième variable aléatoire et $P(x, y, z)$ une probabilité jointe. De la même façon, la probabilité $P(x)$ est définie par :

$$P(x) = \sum_{y \in \mathcal{Y}} \sum_{z \in \mathcal{Z}} P(x, y, z)$$

Probabilité conditionnelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
La probabilité conditionnelle $P(x|y)$ est la probabilité que $X = x$ sachant que $Y = y$:

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Interprétation :
Quelle est la probabilité d'observer $X = x$ une fois que l'on a que observé que $Y = y$.

Par définition :

$$P(x, y) = P(x|y)P(y)$$

Probabilité conditionnelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
La probabilité conditionnelle $P(x|y)$ est la probabilité que $X = x$ sachant que $Y = y$:

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Interprétation :
Quelle est la probabilité d'observer $X = x$ une fois que l'on a que observé que $Y = y$.

Par définition :

$$P(x, y) = P(x|y)P(y)$$

Théorème de Bayes

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Probabilité conditionnelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
La probabilité conditionnelle $P(x|y)$ est la probabilité que $X = x$ sachant que $Y = y$:

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Interprétation :
Quelle est la probabilité d'observer $X = x$ une fois que l'on a que observé que $Y = y$.

Par définition :

$$P(x, y) = P(x|y)P(y)$$

Théorème de Bayes

$$P(x|y) = \frac{P(x, y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Probabilité conditionnelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
La probabilité conditionnelle $P(x|y)$ est la probabilité que $X = x$ sachant que $Y = y$:

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Interprétation :
Quelle est la probabilité d'observer $X = x$ une fois que l'on a que observé que $Y = y$.

Par définition :

$$P(x, y) = P(x|y)P(y)$$

Théorème de Bayes

$$P(x|y) = \frac{P(x, y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_x P(x, y)} \qquad \text{Si } P(y) \neq 0$$

Probabilité conditionnelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
La probabilité conditionnelle $P(x|y)$ est la probabilité que $X = x$ sachant que $Y = y$:

$$P(x|y) = \frac{P(x, y)}{P(y)} \qquad \text{Si } P(y) \neq 0$$

Interprétation :
Quelle est la probabilité d'observer $X = x$ une fois que l'on a que observé que $Y = y$.

Par définition :

$$P(x, y) = P(x|y)P(y)$$

Théorème de Bayes

$$P(x|y) = \frac{P(x, y)}{P(y)} = \frac{P(y|x)P(x)}{P(y)} = \frac{P(y|x)P(x)}{\sum_x P(x, y)} = \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)} \qquad \text{Si } P(y) \neq 0$$

Indépendance statistique 1/2

Indépendance statistique

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
X et Y sont statistiquement indépendantes si et seulement si :

$$P(x, y) = P(x)P(y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

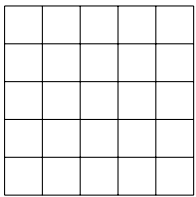
Indépendance statistique 1/2

Indépendance statistique

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
X et Y sont statistiquement indépendantes si et seulement si :

$$P(x, y) = P(x)P(y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

$P(X, Y)$



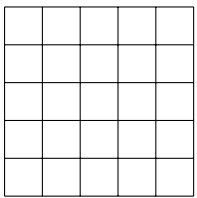
Indépendance statistique 1/2

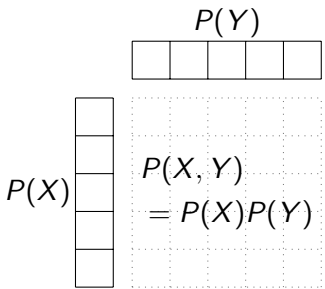
Indépendance statistique

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
X et Y sont statistiquement indépendantes si et seulement si :

$$P(x, y) = P(x)P(y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

$P(X, Y)$





Indépendance statistique 2/2

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ deux variables aléatoires et $P(x, y)$ leur probabilité jointe. Si X et Y sont statistiquement indépendantes, alors :

$$P(x|y) = P(x)$$

C'est-à-dire que la connaissance de la valeur de y ne nous dit rien sur x

Indépendance statistique 2/2

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ deux variables aléatoires et $P(x, y)$ leur probabilité jointe. Si X et Y sont statistiquement indépendantes, alors :

$$P(x|y) = P(x)$$

C'est-à-dire que la connaissance de la valeur de y ne nous dit rien sur x

Preuve :

$$\begin{aligned} P(x|y) &= \frac{P(x,y)}{P(y)} && \text{Si } P(y) \neq 0 \\ &= \frac{P(x)P(y)}{P(y)} && \text{Par définition de l'indépendance statistique} \\ &= P(x) \end{aligned}$$

Divergence de Kullback-Leibler

Définition

La KL-divergence entre P et Q est définie comme :

$$KL[\ P(X)\ |\ Q(X)\ ] = \mathbb{E}_{p(X)}\left[\ \log \frac{P(X)}{Q(X)}\ \right] \quad \underbrace{= \sum_x P(x) \log \frac{P(x)}{Q(x)}}$$

sous condition que : $\forall x \in \mathcal{X}\ : p(x) > 0 \implies q(x) > 0$.

Divergence de Kullback-Leibler

Définition

La KL-divergence entre P et Q est définie comme :

$$KL[\ P(X) \mid Q(X)\ ] = \mathbb{E}_{p(X)}\left[\ \log \frac{P(X)}{Q(X)}\ \right] \quad \underbrace{= \sum_x P(x) \log \frac{P(x)}{Q(x)}}$$

sous condition que : $\forall x \in \mathcal{X} \ : p(x) > 0 \implies q(x) > 0$.

Propriétés

- $KL[\ P(X) \mid Q(X)\ ] \geq 0$ (non négativité)
- $KL[\ P(X) \mid Q(X)\ ] = 0$ si et seulement si $\forall x \in \mathcal{X} \ : P(X) = Q(X)$,

!!!! ATTENTION !!!

En général :

$$KL[\ P(X) \mid Q(X)\ ] \quad \neq \quad KL[\ Q(X) \mid P(X)\ ]$$

Information mutuelle 1/2

Rappel

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
X et Y sont statistiquement indépendantes si et seulement si :

$$P(x, y) = P(x)P(y) \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

Motivation

On veut mesurer la "quantité" de dépendance statistique entre deux variables aléatoires,
plutôt que de simplement répondre par oui/non à la question
« ces deux variables sont-elles statistiquement indépendante ? »

Intuition

On peut comparer la loi joint $P(X, Y)$ et le produit des lois marginales $P(X) \times P(Y)$:

- Plus elles sont similaires, plus les variables sont indépendantes
- Moins elles sont similaires, plus elles sont dépendantes

Information mutuelle 2/2

Intuition

On peut comparer la loi joint $P(X, Y)$ et le produit des lois marginales $P(X) \times P(Y)$.
Comment comparer ces deux lois ? ! $=>$ **Utilisons la KL-divergence !**

Information mutuelle 2/2

Intuition

On peut comparer la loi joint $P(X, Y)$ et le produit des lois marginales $P(X) \times P(Y)$.
Comment comparer ces deux lois ? ! **=> Utilisons la KL-divergence !**

Définition : information mutuelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
L'information mutuelle entre X et Y est définie comme :

$$\begin{aligned} I(X; Y) &= KL[\ P(X, Y)\ |P(X) \times P(Y)\ ] \\ &= \sum_x \sum_y P(X = x, Y = y) \log \frac{P(X = x, Y = y)}{P(X = x) \times P(Y = y)} \end{aligned}$$

Information mutuelle 2/2

Intuition

On peut comparer la loi joint $P(X, Y)$ et le produit des lois marginales $P(X) \times P(Y)$.
Comment comparer ces deux lois ? ! **=> Utilisons la KL-divergence !**

Définition : information mutuelle

Soit $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ et $P(x, y)$ leur probabilité jointe.
L'information mutuelle entre X et Y est définie comme :

$$\begin{aligned} I(X; Y) &= KL[\ P(X, Y)\ |P(X) \times P(Y)\ ] \\ &= \sum_x \sum_y P(X = x, Y = y) \log \frac{P(X = x, Y = y)}{P(X = x) \times P(Y = y)} \end{aligned}$$

Propriétés de l'information mutuelle

- Non-négative $I(X; Y) \geq 0$ et symétrique $I(X; Y) = I(Y; X)$
- Nul si et seulement X et Y sont statistiquement indépendante
- moins les deux variables sont indépendantes, plus grande sera leur information mutuelle

RETOURNONS A NOTRE PROBLEME

EXTRACTION D'UN DICTIONNAIRE DE TRADUCTION

Objectif

Extraire automatique un dictionnaire de traduction de mots partir d'un corpus parallèle.

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom’s **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c’est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l’école.	Both boys and girls should take **cooking** class in school.

Deuxième idée

Le mot "kitchen" est une traduction de "cuisine" si et seulement si :

- Lorsque le mot "cuisine" apparait dans la phrase en Français
- Alors la probabilité de voir "kitchen" dans la traduction est plus élevée que la fréquence d'apparition de "kitchen" dans l'ensemble du corpus

MODÉLISATION

Variables aléatoires

- Pour chaque mot en Français nous définissons une variable aléatoire X(mot)
- Pour chaque mot en Anglais nous définissons une variable aléatoire Y(word)

Par exemple : X(maison), X(chien), X(le), ..., Y(house), Y(dog), Y(the), ...

MODÉLISATION

Variables aléatoires

- Pour chaque mot en Français nous définissons une variable aléatoire X(mot)
- Pour chaque mot en Anglais nous définissons une variable aléatoire Y(word)

Par exemple : X(maison), X(chien), X(le), ..., Y(house), Y(dog), Y(the), ...

Distribution sur X(·) et Y(·)

- P(X(chien) = 1) : probabilité que le mot "chien" **soit observé** dans une phrase en Français
- P(X(chien) = 0) : probabilité que le mot "chien" **ne soit pas observé** dans une phrase en Français
- P(Y(dog) = 1) : probabilité que le mot "dog" **soit observé** dans une phrase en Anglais
- P(Y(dog) = 0) : probabilité que le mot "dog" **ne soit pas observé** dans une phrase en Anglais

À noter que la distribution P(X(·)) pour chaque mot
suit une loi de Bernoulli (et de même pour Y(·)).

MODÉLISATION

Variables aléatoires

- Pour chaque mot en Français nous définissons une variable aléatoire X(mot)
- Pour chaque mot en Anglais nous définissons une variable aléatoire Y(word)

Par exemple : X(maison), X(chien), X(le), ..., Y(house), Y(dog), Y(the), ...

Distribution sur X(·) et Y(·)

- P(X(chien) = 1) : probabilité que le mot "chien" **soit observé** dans une phrase en Français
- P(X(chien) = 0) : probabilité que le mot "chien" **ne soit pas observé** dans une phrase en Français
- P(Y(dog) = 1) : probabilité que le mot "dog" **soit observé** dans une phrase en Anglais
- P(Y(dog) = 0) : probabilité que le mot "dog" **ne soit pas observé** dans une phrase en Anglais

À noter que la distribution P(X(·)) pour chaque mot
suit une loi de Bernoulli (et de même pour Y(·)).

Distribution jointe

De la même façon, nous pouvons définir la distribution jointe P(X(·), Y(·)).

MODÉLISATION

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

MODÉLISATION

X(cuisine) = 1 , X(vivant) = 1 , ... Y(cooking) = 1 , Y(kitchen) = 0 , ...

Fran	
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

MODÉLISATION

X(cuisine) = 1 , X(vivant) = 0 , ... Y(cooking) = 0 , Y(kitchen) = 1 , ...

Français	Anglais
Vivan	
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

MODÉLISATION

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque	Living on my own, I really miss my Mom's **cooking**
Elle c	
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

X(cuisine) = 1 , X(vivant) = 0 , ... Y(cooking) = 0 , Y(kitchen) = 1 , ...

MODÉLISATION

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire	She left the **kitchen** with the kettle boiling
Y a-t-	
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

X(cuisine) = 1 , X(vivant) = 0 , ... Y(cooking) = 1 , Y(kitchen) = 0 , ...

MODÉLISATION

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cu**	
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

X(cuisine) = 1 , X(vivant) = 0 , ... Y(cooking) = 1 , Y(kitchen) = 0 , ...

MODÉLISATION

Français	Anglais
Vivant seul, la **cuisine** de ma mère me manque.	Living on my own, I really miss my Mom's **cooking**.
Elle quitta la **cuisine** avec la bouilloire.	She left the **kitchen** with the kettle boiling.
Y a-t-il encore du café dans la **cuisine** ?	Is there any coffee in the **kitchen**?
La **cuisine** c'est de famille.	**Cooking** runs in my family.
Garçons et filles devraient suivre des cours de **cuisine** à l'école.	Both boys and girls should take **cooking** class in school.

INFORMATION MUTUELLE

Intuition

(1) Si : $P(\ Y(dog)\ |\ X(le)\ ) \simeq P(\ Y(dog)\ )$

Alors Y(dog) et X(le) sont (presque) statistiquement indépendant
=> même si les deux mots co-occurrent souvent ensemble, ce ne sont pas des traductions

INFORMATION MUTUELLE

Intuition

(1) Si : $P(\ Y(dog) \mid X(le)\ ) \simeq P(\ Y(dog)\ )$

Alors Y(dog) et X(le) sont (presque) statistiquement indépendant
=> même si les deux mots co-occurrent souvent ensemble, ce ne sont pas des traductions

(2) Si la probabilité d'observer "dog" dans la traduction quand "chien" apparait en Francais est beaucoup plus élevé que la fréquence d'apparition de "dog" dans le corpus :

$$P(\ Y(dog) \mid X(chien)\ ) \gg P(\ Y(dog)\ )$$

Alors "dog" est une traduction de "chien", et Y(god) et X(chien) sont statistiquement dépendants

INFORMATION MUTUELLE

Intuition

(1) Si : $P(\ Y(dog)\ |\ X(le)\ ) \simeq P(\ Y(dog)\ )$

Alors Y(dog) et X(le) sont (presque) statistiquement indépendant
=> même si les deux mots co-occurrent souvent ensemble, ce ne sont pas des traductions

(2) Si la probabilité d'observer "dog" dans la traduction quand "chien" apparait en Francais est beaucoup plus élevé que la fréquence d'apparition de "dog" dans le corpus :

$$P(\ Y(dog)\ |\ X(chien)\ ) \gg P(\ Y(dog)\ )$$

Alors "dog" est une traduction de "chien", et Y(god) et X(chien) sont statistiquement dépendants

Tri via l'information mutuelle

Plus l'information mutuelle I(X(aaa) ; Y(bbb)) est élevée,
plus il y a de chance que "bbb" soit une traduction de "aaa".

Attention

- C'est une condition nécessaire
- Mais pas suffisante (voir exercices)