IFT 615 – Intelligence Artificielle

Raisonnement probabiliste

Inférences avec une distribution conjointe et classifieur bayésien naïf

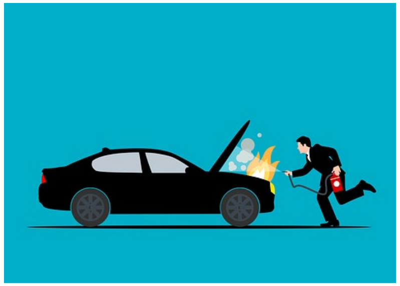
Professeur: Froduald Kabanza

Assistants: D’Jeff Nkashama

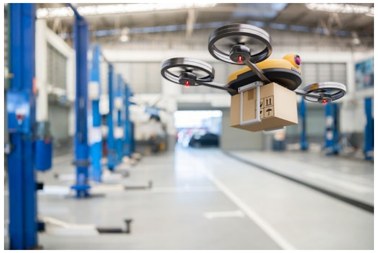
Université de Sherbrooke

Motivation

Détection de pourriels
Classification de documents

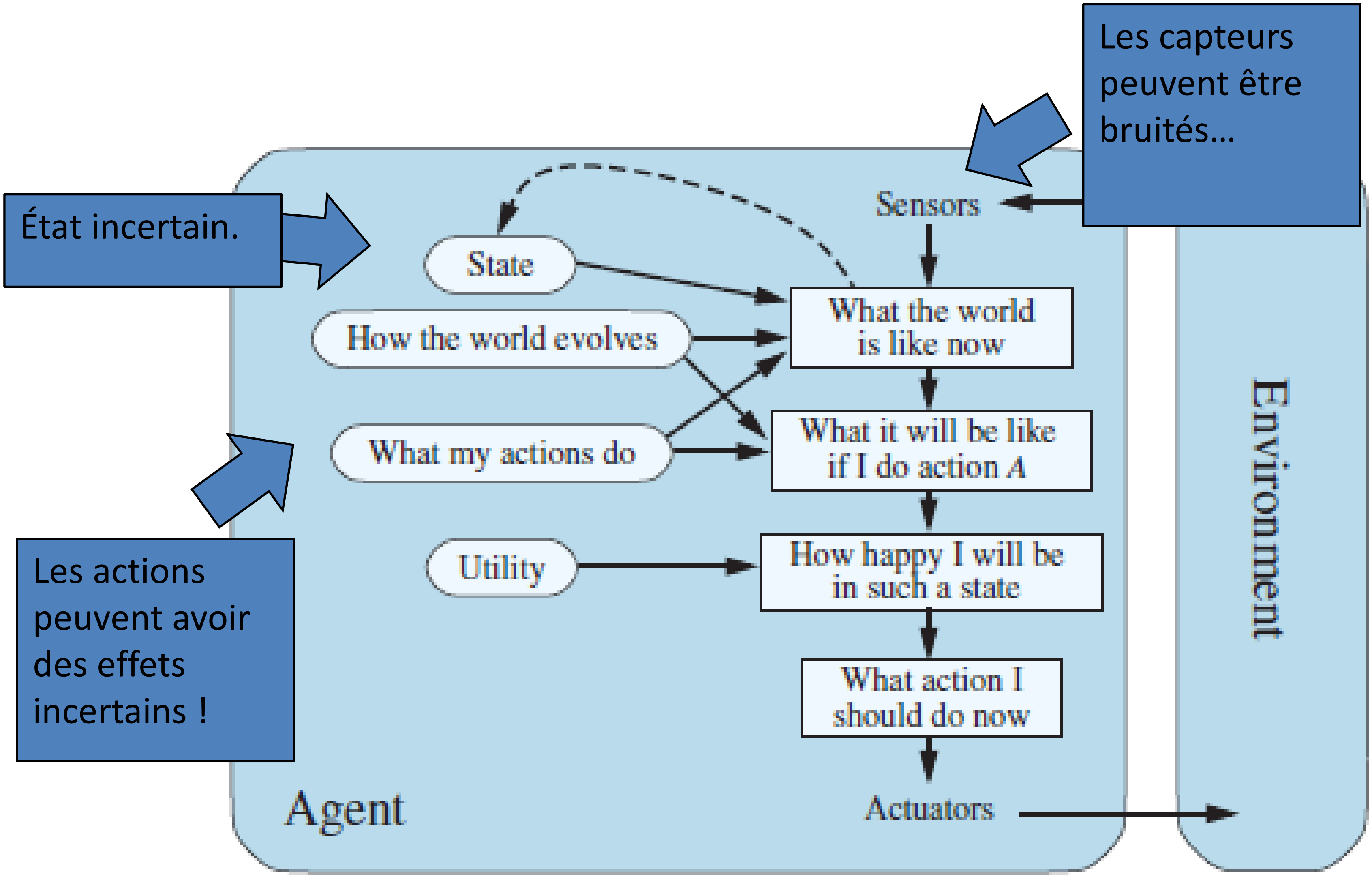


Assurance de dommages



Localisation robotique

Utility-based agents



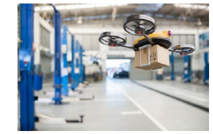
Théorie des probabilités en IA

- Permet de modéliser la vraisemblance d'événements
  - l'information sur la vraisemblance est dérivée
    - des croyances/certitudes d'un agent, ou
    - d'observations empiriques de ces événements

- Donne un cadre théorique pour mettre à jour la vraisemblance d'événements après l'acquisition d'observations

- Facilite la modélisation en permettant de considérer l'influence de phénomènes complexes comme du « bruit »

Decision-Theoretic Agent



function DT-AGENT(*percept*) **returns** an *action*
  persistent: *belief_state*, probabilistic beliefs about the current state of the world
    action, the agent's action

  update *belief_state* based on *action* and *percept*
  calculate outcome probabilities for actions,
    given action descriptions and current *belief_state*
  select *action* with highest expected utility
    given probabilities of outcomes and utility information
  return *action*

Figure 12.1 A decision-theoretic agent that selects rational actions.

Sujets couverts

- Inférence probabiliste avec une distribution conjointe
- Classifieur bayésien naïf

Exemple – Détection de pourriels

- On souhaite raisonner sur la possibilité qu'un courriel soit un pourriel tenant compte de l'incertitude associée à une telle classification

- Pour ce faire, notre modèle (« base de connaissances ») est une **distribution conjointe des probabilités** de variables aléatoires

  - ***Inconnu*** : l'adresse de l'expéditeur n'est pas connue du destinataire
  - ***Sensible*** : le courriel contient un mot sensible
  - ***Pourriel*** : le courriel est un pourriel

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Distribution de probabilités

- **Distribution de probabilités :** l'énumération des probabilités pour toutes les valeurs possibles de variables aléatoires

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Toutes ces probabilités somment à 1

- Exemples :
  - **P**(*Pourriel, Inconnu, Sensible*)
  - **P**(*Pourriel*) = [*P*(*Pourriel=faux*), *P*(*Pourriel=vrai*)] = [0.8, 0.2]
  - P(Pourriel, Inconnu)

    = [[Pourriel= faux, Inconnu=faux), [Pourriel= vrai, Inconnu=faux)],

    [Pourriel= faux, Inconnu=vrai), [Pourriel= vrai, Inconnu=vrai)]]
- La somme est toujours égale à 1
- J'utilise le symbole **P** pour les distributions et *P* pour les probabilités
  - *P*(*Pourriel*) désignera la probabilité *P*(*Pourriel=x*) pour une valeur *x* non-spécifiée

Probabilité conjointe

- **Probabilité conjointe** : probabilité d'une assignation de valeurs à toutes la variables
  - *P*(*Inconnu=vrai*, *Sensible=vrai*, *Pourriel=vrai*) = 0.108 (10.8%)
  - *P*(*Inconnu=faux*, *Sensible=faux*, *Pourriel=vrai*) = 0.008 (0.8%)

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Probabilité marginale

- **Probabilité marginale** : probabilité sur un sous-ensemble des variables
  - ***$P(Y)= \Sigma_z P(Y, Z=z)$* -** Pour n'importe quelle ensemble de variable **Y** et **Z**
  - $P(Inconnu{=}vrai, Pourriel{=}vrai)$
    $= P(Inconnu{=}vrai, Sensible{=}vrai, Pourriel{=}vrai) + P(Inconnu{=}vrai, Sensible{=}faux, Pourriel{=}vrai)$
    $= \Sigma_{x\in\{vrai, faux\}} P(Inconnu{=}vrai, Sensible{=}x, Pourriel{=}vrai) = 0.108 + 0.012 =$ **0.12**

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Probabilité marginale

- **Probabilité marginale** : probabilité sur un sous-ensemble des variables
  - $P(Pourriel{=}vrai)$
    $= \Sigma_{x \in \{vrai,\, faux\}} \Sigma_{y \in \{vrai,\, faux\}} P(Inconnu{=}x, Sensible{=}y, Pourriel{=}vrai)$
    $= 0.108 + 0.012 + 0.072 + 0.008 = \mathbf{0.2}$

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Probabilité d'un événement arbitraire

- **Probabilité de disjonction (« ou ») d'événements** :
  - *P*(*Pourriel=vrai* ou *Inconnu=faux*) – Six états (mondes) possibles

    = 0.108 + 0.012 +0.072 + 0.008 + 0.144 + 0.576

    = ***0.92***

  - *P*(*Pourriel=vrai* ou *Inconnu=faux*) – Une autre façon de le calculer

    = *P*(*Pourriel=vrai*) + *P*(*Inconnu=faux*) - *P*(*Pourriel=vrai, Inconnu=faux*)

    = 1 - *P*(*Pourriel=****faux****, Inconnu=****vrai***) = 1 – 0.016 – 0.064 = **0.92**

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Probabilité d'un événement arbitraire

- On peut calculer la probabilité d'événements arbitrairement complexes
  - il suffit d'additionner les probabilités des événements élémentaires associés
  - *P*((*Pourriel=vrai, Inconnu=faux*) ou (*Sensible=faux, Pourriel=faux*))
    = 0.072 + 0.008 + 0.064 + 0.576 = **0.72**

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Probabilité conditionnelle

- **Probabilité conditionnelle :**
  - ***P(X|Y) = P(X,Y) / P(Y)*** *si P(Y) ≠* ***0***
  - *P*(*Pourriel=faux* | *Inconnu=vrai*)
    = *P*(*Pourriel=faux, Inconnu=vrai*) / *P*(*Inconnu=vrai*)
    = (0.016 + 0.064) / (0.016 + 0.064 + 0.108 + 0.012) = **0.4**

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Distribution conditionnelle

- On a vu que
  - $P(Y) = \Sigma_z\, P(Y, Z{=}z)$
  - $\boldsymbol{P(X|Y) = P(X,Y) / P(Y)}$ si $P(Y) \neq \mathbf{0}$
- On en déduit: $\boldsymbol{P(Y) = \Sigma_z\, P(Y|Z)P(Z{=}z)}$

Distribution conditionnelle

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

- Exemple :
  - **P**(*Pourriel* | *Inconnu=faux*)
    = [*P*(*Pourriel=faux* | *Inconnu=faux*), *P*(*Pourriel=vrai* | *Inconnu=faux*)]
    = [0.9, 0.1]
  - **P**(*Pourriel* | *Inconnu*)
    = [[*P*(*Pourriel=faux* | *Inconnu=faux*), *P*(*Pourriel=vrai* | *Inconnu=faux*)],
    [*P*(*Pourriel=faux* | *Inconnu=vrai*), *P*(*Pourriel=vrai* | *Inconnu=vrai*)]]
    = [[0.9, 0.1], → somme à 1
    [0.4, 0.6]] → somme à 1
- **Chaque sous-ensemble de probabilités** associé aux mêmes valeurs des variables sur lesquelles on conditionne somme à 1
- **P**(*Pourriel* | *Inconnu*) contient deux distributions de probabilités sur la variable *Pourriel* : une dans le cas où *Inconnu=faux,* l'autre lorsque *Inconnu=vrai*

Distribution conditionnelle

- Une distribution conditionnelle peut être vue comme une distribution **renormalisée** afin de satisfaire les conditions de sommation à 1
- ***P(X|e) = α Σ_y P(X,e,y)***
- Exemple :
  - **P**(*Pourriel* | *Inconnu=vrai*)
    = α **P**(*Pourriel, Inconnu=vrai*)
    = α [0.08, 0.12]
    = (1/ (0.08 + 0.12)) [0.08, 0.12]
    = [0.4, 0.6]
  - **P**(*Pourriel* | *Inconnu*)
    = [α_{faux} **P**(*Pourriel, Inconnu=faux*) ,
    α_{vrai} **P**(*Pourriel, Inconnu=vrai*)]
    = [[0.72, 0.08] / (0.72 + 0.08),
    [0.08, 0.12] / (0.08 + 0.12)]
    = [[0.9, 0.1],
    [0.4, 0.6]]

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

Règle du produit

- **Règle du produit :**
  - $P(X,Y)=P(X|Y)P(Y)$
  - $P(Pourriel{=}faux, Inconnu{=}vrai)$
    $= P(Pourriel{=}faux \mid Inconnu{=}vrai)\ P(Inconnu{=}vrai)$
    $= P(Inconnu{=}vrai \mid Pourriel{=}faux)\ P(Pourriel{=}faux)$
  - En général :
    $P(Pourriel, Inconnu) = P(Pourriel \mid Inconnu)\ P(Inconnu)$
    $= P(Inconnu \mid Pourriel)\ P(Pourriel)$

Règle de chaînage

- **Règle du produit :**
  - ***$P(X,Y)=P(X|Y)P(Y)$***

- **Règle de chaînage** (***chain rule***) pour *n* variables $X_1 \dots X_n$:
  - $$\begin{aligned} P(X_1, \dots, X_n) &= P(X_1,\dots,X_{n-1})\, P(X_n \mid X_1,\dots,X_{n-1}) \\ &= P(X_1,\dots,X_{n-2})\, P(X_{n-1} \mid X_1,\dots,X_{n-2})\, P(X_n \mid X_1,\dots,X_{n-1}) \\ &= \dots \\ &= \prod_{i=1..n} P(X_i \mid X_1, \dots, X_{i-1}) \end{aligned}$$

Règle de chaînage

- La règle du chaînage est vraie, quelle que soit la distribution de $X_1 \dots X_n$
- Plutôt que de spécifier toutes les probabilités jointes $P(X_1, \dots, X_n)$, on pourrait plutôt spécifier $P(X_1)$, $P(X_2|X_1)$, $P(X_3|X_1, X_2)$, ..., $P(X_n \mid X_1,\dots,X_{n-1})$

	Inconnu = vrai		*Inconnu = faux*	
	Sensible = vrai	*Sensible = faux*	*Sensible = vrai*	*Sensible = faux*
Pourriel = vrai	0.108	0.012	0.072	0.008
Pourriel = faux	0.016	0.064	0.144	0.576

- Exemple, on aurait pu spécifier :
  - $P(Pourriel{=}faux) = 0.8$, $P(Pourriel{=}vrai) = 0.2$
  - $P(Inconnu{=}faux \mid Pourriel{=}faux) = 0.9$, $P(Inconnu{=}vrai \mid Pourriel{=}faux) = 0.1$
    $P(Inconnu{=}faux \mid Pourriel{=}vrai) = 0.4$, $P(Inconnu{=}vrai \mid Pourriel{=}vrai) = 0.6$

- On aurait tous les ingrédients pour calculer les $P(Pourriel, Inconnu)$:
  - $P(Pourriel{=}faux, Inconnu{=}vrai) = P(Inconnu{=}vrai \mid Pourriel{=}faux)\, P(Pourriel{=}faux)$
    $= 0.1 * 0.8 = 0.08$
  - $P(Pourriel{=}vrai, Inconnu{=}vrai) = P(Inconnu{=}vrai \mid Pourriel{=}vrai)\, P(Pourriel{=}vrai)$
    $= 0.6 * 0.2 = 0.12$

Règle de Bayes

- **P(X|Y) = P(Y|X)P(X)/P(Y)**

- Donne une probabilité **diagnostique** à partir d'une probabilité **causale** :
  - $P(Cause|Effect) = P(Effect|Cause)\ P(Cause)\ /\ P(Effect)$

Règle de Bayes

- **P(X|Y) = P(Y|X)P(X)/P(Y)**
- Donne une probabilité **diagnostique** à partir d'une probabilité **causale** :
  - $P(Cause|Effet) = P(Effet|Cause)\ P(Cause)\ /\ P(Effet)$
- On pourrait calculer $P(Pourriel{=}faux \mid Inconnu{=}vrai)$:
  - $P(\neg \text{pourriel} \mid inconnu)$
    $= P(\neg \text{pourriel}, inconnu)\ /\ P(inconnu)$
    $= P(\neg \text{pourriel}, inconnu)\ /\ (\ P(inconnu, \neg pourriel) + P(inconnu, pourriel))$
    $= \alpha\ P(inconnu|\neg \text{pourriel})\ P(\neg \text{pourriel})$
    $= 0.08\ /\ (0.08 + 0.12) = \mathbf{0.4}$

Pourriel=faux $\Leftrightarrow$ $\neg$pourriel
Pourriel=vrai $\Leftrightarrow$ pourriel

- On appelle $P(Pourriel)$ une **probabilité a priori**
  - c'est notre croyance p/r à la présence d'une Pourriel **avant** toute observation
- On appelle $P(Pourriel|\ Inconnu)$ une probabilité a **posteriori**
  - c'est notre croyance mise à jour après avoir observé *Inconnu*
- **La règle de Bayes** lie ces deux probabilités ensemble
  - $\underline{P(\neg \text{pourriel} \mid inconnu)} = \alpha\ P(inconnu|\neg \text{pourriel})\ \underline{P(\neg \text{pourriel})}$

Indépendance

- Soit les variables A et B, elles sont **indépendantes** si et seulement si
  - $P(A|B) = P(A)$ ou
  - $P(B|A) = P(B)$ ou
  - $P(A, B) = P(A)\ P(B)$

Indépendance

- Soit les variables A et B, elles sont **indépendantes** si et seulement si
  - $P(A|B) = P(A)$ ou
  - $P(B|A) = P(B)$ ou
  - $P(A, B) = P(A)\ P(B)$

- Exemple : $P(Pluie, Pourriel) = P(Pluie)\ P(Pourriel)$

Pluie	*Pourriel*	Probabilité	
vrai	*vrai*	0.03	$= P(pluie)\ P(Pourriel) = 0.3 * 0.1$
vrai	*faux*	0.27	$= P(pluie)\ P(\neg Pourriel) = 0.3 * 0.9$
faux	*vrai*	0.07	$= P(\neg pluie)\ P(Pourriel) = 0.7 * 0.1$
faux	*faux*	0.63	$= P(\neg pluie)\ P(\neg Pourriel) = 0.7 * 0.9$

$P(Pluie = vrai) = 0.3$

$P(Pourriel = vrai) = 0.1$

Indépendance

- L'indépendance totale est puissante mais rare

- L'indépendance entre les variables permet de réduire la taille de la distribution de probabilités et rendre les inférences plus efficaces
  - dans l'exemple précédent, on n'a qu'à stocker en mémoire $P(Pluie = vrai) = 0.3$ et $P(Pourriel = vrai) = 0.1$, plutôt que la table au complet

- Mais il est rare d'être dans une situation où toutes les variables sont réellement indépendantes

Pluie	***Pourriel***	**Probabilité**
vrai	*vrai*	0.03
vrai	*faux*	0.27
faux	*vrai*	0.07
faux	*faux*	0.63

Indépendance conditionnelle

- Si je sais déjà que le courriel est un pourriel, ma croyance (probabilité) qu'il contienne un mot sensible ne dépend plus du fait que l'expéditeur me soit inconnu ou non :
  - $P(Sensible \mid Inconnu, Pourriel{=}vrai) = P(Sensible \mid Pourriel{=}vrai)$

- On dit que *Sensible* est **conditionnellement indépendante** de *Inconnu* étant donné *Pourriel*, puisque :
  - $P(Sensible \mid Inconnu, Pourriel) = P(Sensible \mid Pourriel)$

- Formulations équivalentes :
  - $P(Inconnu \mid Sensible , Pourriel) = P(Inconnu \mid Pourriel)$
  - $P(Inconnu, Sensible \mid Pourriel) = P(Inconnu \mid Pourriel)\, P(Sensible \mid Pourriel)$

Indépendance conditionnelle

- Réécrivons la distribution conjointe en utilisant la **règle de chaînage** (***chain rule***) :

$$
\begin{aligned}
&P(\mathit{Inconnu}, \mathit{Sensible}, \mathit{Pourriel}) \\
&\quad = P(\mathit{Inconnu} \mid \mathit{Sensible}, \mathit{Pourriel})\, P(\mathit{Sensible}, \mathit{Pourriel}) \\
&\quad = P(\mathit{Inconnu} \mid \mathit{Sensible}, \mathit{Pourriel})\, P(\mathit{Sensible} \mid \mathit{Pourriel})\, P(\mathit{Pourriel}) \\
&\quad = P(\mathit{Inconnu} \mid \mathit{Pourriel})\, P(\mathit{Sensible} \mid \mathit{Pourriel})\, P(\mathit{Pourriel})
\end{aligned}
$$

- C-à-d., 2 + 2 + 1 = 5 **paramètres individuels/distincts**

- Dans des cas idéals, l'exploitation de l'indépendance conditionnelle réduit la complexité de représentation de la distribution conjointe de exponentielle ($O(2^n)$) en linéaire ($O(n)$)

En bref

- **Probabilité jointe :** $P(X_1, \ldots, X_n)$
- **Probabilité marginale :** $P(X_i)$, $P(X_i, X_j)$, etc.
- **Probabilité conditionnelle :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \dfrac{P(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n)}{P(X_{k+1}, \ldots, X_n)}$
- **Régle de chaînage :** $P(X_1, \ldots, X_n) = \Pi_{\mathrm{i}=1..n} P(X_i \mid X_1, \ldots, X_{i-1})$
- **Indépendance :** X_i et X_j sont indépendantes si

  $P(X_i, X_j) = P(X_i)\, P(X_j)$, ou $P(X_i | X_j) = P(X_i)$ ou $P(X_j | X_i) = P(X_j)$
- **Indépendance conditionnelle :** $\mathrm{X_i}$ et $\mathrm{X_j}$ sont indépendante sachant $\mathrm{X_k}$ si

  $P(X_i, X_j | X_k) = P(X_i | X_k)\, P(X_j | X_k)$ ou $P(X_i | X_j, X_k) = P(X_i | X_k)$ ou $P(X_j | X_i, X_k) = P(X_j | X_k)$
- **Règle de Bayes :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \dfrac{P(X_{k+1}, \ldots, X_n \mid X_1, \ldots, X_k)\, P(X_1, \ldots, X_k)}{P(X_{k+1}, \ldots, X_n)}$

Autres types de variables aléatoires

- On s'est concentré sur des variables aléatoires Booléennes ou binaires
  - le **domaine**, c.-à-d. l'ensemble des valeurs possibles de la variable, était toujours {*vrai*, *faux*}

- On pourrait avoir d'autres types de variables, avec des domaines différents :
  - **Discrètes** : le domaine est énumérable
    - *Météo* $\in$ {*soleil*, *pluie*, *nuageux*, *neige*}
    - lorsqu'on marginalise, on doit sommer sur toutes les valeurs : $P(Température) = \Sigma_{x \in \{soleil,\ pluie,\ nuageux,\ neige\}} P(Température, Météo{=}x)$
  - **Continues** : le domaine est continu (par exemple, l'ensemble des réels)
    - exemple : *PositionX* = 4.2
    - le calcul des probabilités marginales nécessite des intégrales

Classifieur bayésien naïf

- Le classifieur (modèle) bayésien naïf est définit comme suit

$$\boldsymbol{P}(\text{Cause}, \mathit{Effet}_1, \ldots, \mathit{Effet}_n) = \boldsymbol{P}(\text{Cause}) = \Pi_{\text{i=1..n}}\, \boldsymbol{P}(\mathit{Effet}_i \mid \mathit{Cause})$$

- Naïf parce qu'on suppose l'independence conditionnel. Mais fonctionne dans beaucoup d'applications

Classifieur bayésien naïf

- Le modèle (Classifieur) bayésien naïf est définit comme suit

$$\boldsymbol{P}(\text{Cause}, \mathit{Effet}_1, \ldots, \mathit{Effet}_n) = \boldsymbol{P}(\text{Cause}) \prod_{i=1..n} \boldsymbol{P}(\mathit{Effet}_i \mid \mathit{Cause})$$

- Pour l'appliquer, en général, on observe des effets (e) et on veut diagnostiquer la cause.

- Noton **E=e** les effets observés. On a vu que $\boldsymbol{P(Cause|e)} = \alpha\, \Sigma_y\, \boldsymbol{P}(Cause,e,y)$
- On a donc:

$$\begin{aligned}
\boldsymbol{P}(\text{Cause} \mid \text{e}) &= \alpha\, \Sigma_y\, \boldsymbol{P}(Cause) P(y \mid Cause) \prod_{j=1..n} \boldsymbol{P}(e_i \mid Cause) \\
&= \alpha\, \boldsymbol{P}(Cause) \prod_{j=1..n} \boldsymbol{P}(e_i \mid Cause)\, \Sigma_y\, P(y \mid Cause) \\
&= \boldsymbol{\alpha\, P(Cause) \prod_{j=1..n} P(e_i \mid Cause)}
\end{aligned}$$

Classification de documents

- Classifieur bayésien naïf

$$P(Cause|e) = \alpha\, P(Cause)\, \Pi_{j=1..n}\, P(e_i\,|Cause)$$

- Classification de documents: étant donné un document texte, déterminer dans laquelle des catégories prédéfinie il appartient (ex: sport, économie)

- Exemples de textes (documents):
  - Apple a fait état jeudi d'un chiffre d'affaires et d'un bénéfice net supérieur aux attentes pour le trimestre allant d'octobre à décembre l'année dernière, la forte hausse des ventes d'iPhone, notamment en Chine, ayant plus que compensé les difficultés des chaînes d'approvisionnement … (*Tiré de Radio Canada / Économie – 2022-01-08*)
  - Le Canada s'est rapproché davantage de son objectif en défaisant le Honduras 2-0, jeudi, en match de qualification pour la Coupe du monde au Qatar (*Tiré de Radio Canada / Sport – 2022-01-08*)

Classification de documents

- Classifieur bayésien naïf

$$P(Cause|e) = \alpha\ P(Cause)\ \Pi_{j=1..n}\ P(e_i\ |Cause)$$

- Étant donné un document, déterminer dans laquelle des catégories prédéfinie il appartient (e.g., sport, politique, économie, etc.)

  - **Cause** correspond à la catégorie (classe) des documents (sport, politique, etc.)
  - $\mathbf{e_j}$ correspond à la présence ou absence de certains mots clés, *keyWord*$_{i.}$.
  - *Donc* $\boldsymbol{e}$ *≡ {keyWord*$_1$*, .., keyWord*$_n$ *}, c.-à-.d, les mots-clés observés*

$$P(Class|keyWord_1, .., keyWord_n) = \alpha\ P(Class)\ \Pi_{j=1..n}\ P(keyWorld_i\ |Class)$$

- Le modèle bayésien naïf consiste des probabilités à priori *P(Class)* et des probabilités conditionnelles $P(KeyWorld_i\ |Class)$.

Classification de documents

$$\boldsymbol{P(Category|\mathrm{ObservedKeyWords}) = \alpha\, P(Catgeory)\, \Pi_{j=1..n}\, P(HasWorld_i\, |Category)}$$

- Le modèle bayésien naïf consiste des probabilités à priori *P(Category)* et des probabilités conditionnelles $P(HasWorld_i\,|Category)$.

- Pour classifier un document

  - On vérifie quels mots clés apparaissent dans le document, ce qui donne ObservedKeywords
  - On applique ensuite l'équation pour obtenir la distribution des probabilités à postériori des catégories, c.-a.-d., ***P****(Category|ObservedKeyWords)*
  - On choisit finalement $argmax_c\, P(Category = c\ |ObservedKeyWords)$, c.-à-d., la catégorie avec la probabilité à postériori la plus elevée.

Classification de documents

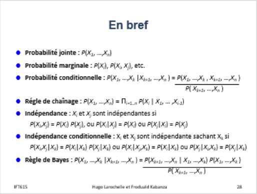
$$P(Category|ObservedKeywords) = \alpha\, P(Catgeory)\, \Pi_{j=1..n}\, P(HasWorld_i\,|Category)$$

- Le modèle bayésien naïf consiste des probabilités à priori *P(Category)* et des probabilités conditionnelles $P(HasWorld_i\,|Category)$.

- Pour apprendre le modèle:
  - *P(Category = c)* est la fraction des documents de cette catégorie vue jusqu'à présent.
  - $P(HasWord_i \mid Category = c)$ est la fraction de documents de catégorie *c* de la catégorie qui contient le mot $Word_i$.

Conclusion

- La théorie des probabilités est un cadre mathématique permettant de faire des inférences explicables facilement.
- Les algorithmes d’IA probabilistes utilisent des modèles probabilistes graphiques qui sont plus efficace que la théorie des probabilité de base:
  - Réseau bayésien
  - Réseau bayésien dynamique
    - » Modèle de Markov caché
    - » Filtre de particule



En bref

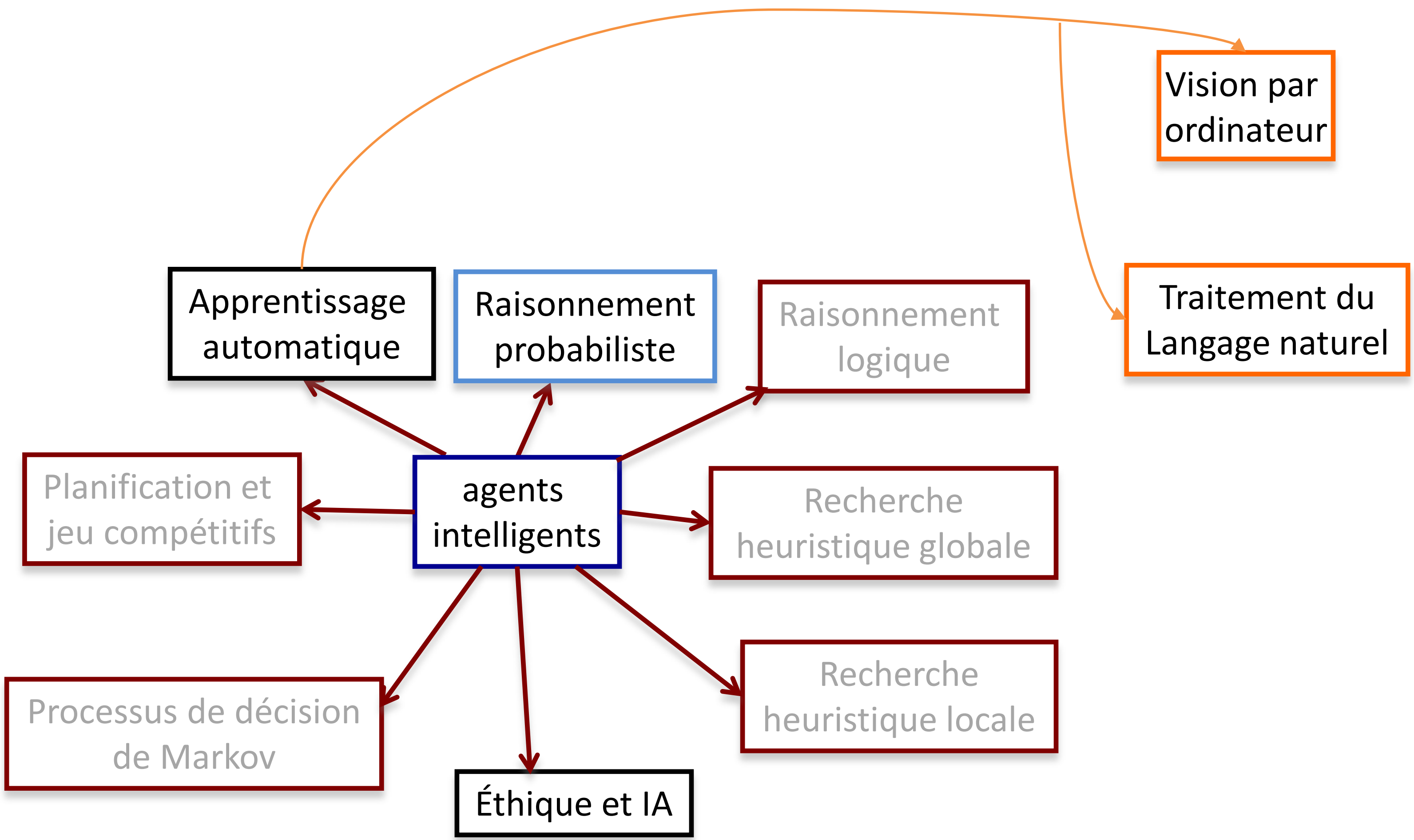
- **Probabilité jointe :** $P(X_1, \ldots, X_n)$
- **Probabilité marginale :** $P(X_i)$, $P(X_i, X_j)$, etc.
- **Probabilité conditionnelle :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \frac{P(X_1, \ldots, X_k, X_{k+1}, \ldots, X_n)}{P(X_{k+1}, \ldots, X_n)}$
- **Régle de chaînage :** $P(X_1, \ldots, X_n) = \prod_{i=1..n} P(X_i \mid X_1, \ldots, X_{i-1})$
- **Indépendance :** X_i et X_j sont indépendantes si $P(X_i, X_j) = P(X_i)\,P(X_j)$, ou $P(X_i \mid X_j) = P(X_i)$ ou $P(X_j \mid X_i) = P(X_j)$
- **Indépendance conditionnelle :** X_i et X_j sont indépendante sachant X_k si $P(X_i, X_j \mid X_k) = P(X_i \mid X_k)\,P(X_j \mid X_k)$ ou $P(X_i \mid X_j, X_k) = P(X_i \mid X_k)$ ou $P(X_j \mid X_i, X_k) = P(X_j \mid X_k)$
- **Règle de Bayes :** $P(X_1, \ldots, X_k \mid X_{k+1}, \ldots, X_n) = \frac{P(X_{k+1}, \ldots, X_n \mid X_1, \ldots, X_k)\,P(X_1, \ldots, X_k)}{P(X_{k+1}, \ldots, X_n)}$

IFT615 Hugo Larochelle et Froduald Kabanza 28

Sujets couverts par le cours

Concepts et algorithmes

Applications



Vous devriez être capable de...

- À partir d'une distribution conjointe ou des distributions conditionnelles et a priori nécessaires :
  - calculer une probabilité conjointe
  - calculer une probabilité marginale
  - déterminer si deux variables sont indépendantes
  - déterminer si deux variables sont conditionnellement indépendantes sachant une troisième
  - Appliquer la règle du chainage
  - Appliquer la règle de Bayes
- Expliquer le classifieur bayésien naïf et son application à la classification de documents