Learning as search

Fabien Torre

Université de Lille

Mercredi 23 et 30 septembre 2009

Langages de représentation

- Représentation des exemples à l'aide de $\mathcal{X}$;
- représentation des hypothèses à l'aide de $\mathcal{H}$;
- astuce de représentation [Feigenbaum, 1977] : $\mathcal{X} \subset \mathcal{H}$;
- relation de subsomption $\succeq$: $\mathcal{H} \times \mathcal{H}$ et test de subsomption ;
- correction $\forall e \in A_{/-} : h \not\succeq e$ et complétude $\forall e \in A_{/+} : h \succeq e$
- structuration de $\mathcal{H}$ par $\succeq$;
- $\mathcal{E}$, espace des versions [Mitchell, 1982] : les hypothèses de $\mathcal{H}$ correctes et complètes (*hypothèses consistantes avec A*) ;
- algorithmes descendants ou ascendants.

Différents biais

Définition : Biais

Tout moyen de préférer une hypothèse à une autre, alors qu'elles sont équivalentes pour le problème d'apprentissage.

- biais de langage : fixe la forme de la solution, essentiellement le choix de $\mathcal{H}$;
- biais de recherche : toute stratégie pour parcourir l'espace des hypothèses ;
- biais de validation : le critère d'arrêt de la recherche.

... connotation négative d'un *apprentissage biaisé*, comment se passer des biais ?

Construction d'un système non biaisé [Mitchell, 1980] (1)

Choix d'un $\mathcal{H}$ sans biais :

- doit pouvoir caractériser n'importe quel étiquetage des exemples ($\mathcal{H}$ doit *pulvériser* $\mathcal{X}$) ;
- sans contraindre la forme des hypothèses ;
- une hypothèse est un sous-ensemble de $\mathcal{X}_{/+}$.

Construction d'un système non biaisé [Mitchell, 1980] (2)

Trouver une procédure de recherche sans biais :

- construire tout l'espace des versions $\mathcal{E}$;

$$h \in \mathcal{E} \Leftrightarrow (A_{/+} \subseteq h) \wedge (A_{/-} \cap h = \emptyset)$$

- les hypothèses de l'espace des versions se distinguent les unes des autres par les exemples étrangers à l'ensemble d'apprentissage qu'elles contiennent ou non ;
- une autre caractérisation de $\mathcal{E}$ est donc :

$$\mathcal{E} = \left\{ h \,|\, h = A_{/+} \cup E \text{ où } E \subseteq \mathcal{X} - A \right\}$$

Construction d'un système non biaisé [Mitchell, 1980] (3)

Classification sans biais :

- refus de choisir une des hypothèses de $\mathcal{E}$;
- un exemple est prédit positif que s'il est couvert par *toutes* les hypothèses de l'espace des versions ;
- un exemple est prédit négatif s'il n'est couvert par *aucune* des hypothèses de l'espace des versions ;
- si un exemple est classé par le système, c'est qu'il a été vu en apprentissage ! En effet :

$$\bigcap_{h \in \mathcal{E}} h = A_{/+} \qquad \text{et} \qquad \mathcal{X} - \bigcup_{h \in \mathcal{E}} h = A_{/-}$$

- sinon il n'y a pas unanimité au sein de $\mathcal{E}$ et l'exemple reste de classe inconnue.

Justification des biais

Les biais sont donc nécessaires pour éviter le par cœur : ils permettent la généralisation, la compression, le saut inductif...

Justification des biais

- Connaissances liées à un problème, à un domaine particulier, et rassemblées sous le nom *théorie du domaine* ;
- privilégier la simplicité de la solution (rasoir d'Occam) ;
- indication sur les utilisations de la prédiction ;
- connaissance de la source des exemples ;
- analogie avec des cas déjà traités.

Compromis biais / variance

- Un concept cible c, pas nécessairement dans $\mathcal{H}$;
- h^* la meilleure approximation dans $\mathcal{H}$ (en terme d'erreur en généralisation) ;
- h l'hypothèse de $\mathcal{H}$ qui est effectivement apprise.

Décomposition de erreur(h)

- Entre c et h^*, on commet une erreur due au *biais*, c'est-à-dire au choix de $\mathcal{H}$;
- entre h^* et h, c'est l'erreur due à la *variance* c'est-à-dire à l'algorithme et à l'échantillon.

Un apprenant avec une variance importante est dit *instable*.

Vapnik-Chervonenkis (1)

Définition : Pulvérisation

Un ensemble $A \subseteq \mathcal{X}$ est dit *pulvérisé* par $\mathcal{H}$ ssi pour chaque étiquetage possible des exemples de A, $\mathcal{H}$ contient une hypothèse qui sépare parfaitement les classes.

Définition : VC dimension

La *VC dimension d'un ensemble d'hypothèses* $\mathcal{H}$ est la taille n du plus grand ensemble d'exemples A que $\mathcal{H}$ parvient à pulvériser.

... il existe un A de taille n tel que pour tout étiquetage...

Vapnik-Chervonenkis (2)

Exercices pratiques :

- Dans $\mathbb{R}^2$: une droite ? parallèle aux axes ? rectangles ? cercles ?
- hyperplans dans $\mathbb{R}^n$?
- trois droites qui votent ?
- dimension VC du $\mathcal{H}$ sans biais de Tom Mitchell ?
- VCdim d'une combinaison de classifieurs dont la VCdim est finie ?

Vapnik-Chervonenkis (3)

- Borne sur l'erreur [Burges, 1998] : avec une probabilité $1-\lambda$,

$$e(h \in \mathcal{H}) \leq e_A(h) + \sqrt{\frac{VC(\mathcal{H}).[log(\frac{2n}{VC(\mathcal{H})})+1] - log(\frac{\lambda}{4})}{n}}$$

- on retrouve le compromis biais/variance ;
- avec plus d'exemples pour apprendre, on fait moins d'erreurs en prédiction ;
- attention, c'est une *borne*...

Fiche signalétique

- Préciser la nature de $\mathcal{E}$;
- indiquer le choix de $\mathcal{H}$;
- évaluer la VCdim de $\mathcal{H}$;
- expliciter le test de subsomption ;
- méthode ascendante ou descendante ?
- dire les biais de la méthode (de langage, de recherche et de validation).

Un bilan en citations I

Apprendre à partir d'exemples

Dans l'enseignement des sciences, l'exemple vaut mieux que la théorie. *Isaac Newton, d'après « 60 gags de Boule et Bill ».*

Langage $\mathcal{H}$ et erreur de biais

- Si je les ai mangées, qui m'a donné une bouche ?
- Il n'y a pas de réponse dans ton vocabulaire.

William Golding, Chris Martin.

De la nécessité des biais

Avoir tous les avis, c'est n'en avoir aucun. *Françoise Hardy, horoscope du 25 mai 1998, RFM.*

Un bilan en citations II

L'importance des biais

The power of a generalization system follows directly from its biases. *Tom Mitchell.*

Trouver des biais

L'imagination est plus importante que le savoir. *Albert Einstein.*

Fonder les biais

L'imagination, ce n'est pas le mensonge. *Daniel Pennac, Messieurs les enfants.*

Un bilan en citations III

No Free Lunch Theorem

Y a pas d'cheval qu'on peut pas monter,
mais y a pas d'homme qu'il peut pas jeter.
Alors quand tu voudras monter un mustang,
oublie pas qu' c'est p't-être pas toi qui tiendras d'ssus.
Marion Zimmer Bradley, Redécouverte.

Bibliographie I

Burges, C. J. (1998).
A tutorial on support vector machines for pattern recognition.
Data Mining and Knowledge Discovery, 2 :121–167.

Feigenbaum, E. A. (1977).
The art of artificial intelligence : Themes and case studies of knowledge engineering.
In Reddy, R., editor, *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, pages 1014–1029. Morgan Kaufmann.

Bibliographie II

- Mitchell, T. M. (1980).
  The need for biases in learning generalizations.
  In *Readings in Machine Learning*, pages 184–191. Morgan Kaufmann.
  Published in 1991.
- Mitchell, T. M. (1982).
  Generalization as search.
  Artificial Intelligence, 18(2) :203–226.