

Apprentissage par transfert pour la classification supervisée d'objets 3D

François MEUNIER, Christophe MARSALA, Laurent CASTANIE, Bruno CONCHE

Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, Paris

Total Exploration-Production, Tour Coupole, La Défense, Paris

Francois.Meunier@lip6.fr

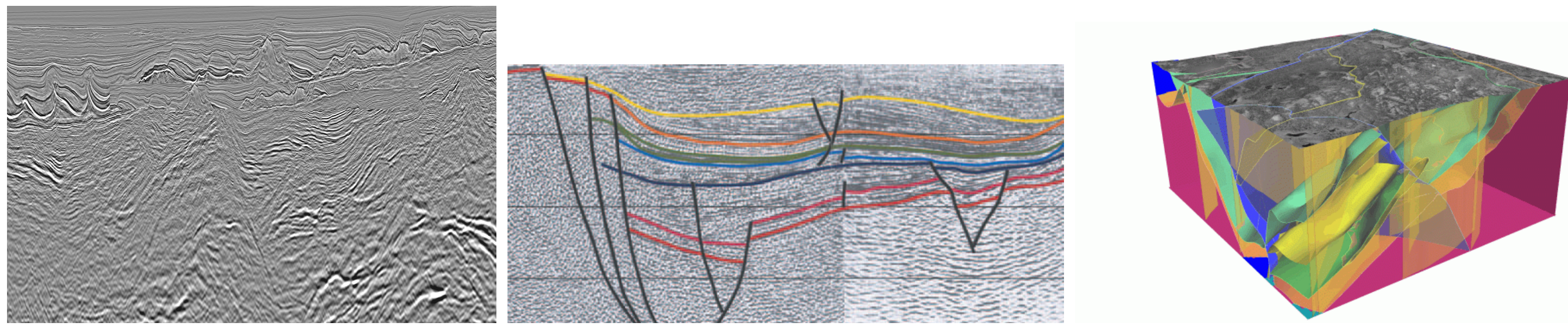
Résumé

Nous proposons l'introduction du transfert de connaissances depuis un algorithme de classification d'objets 3D, afin de faire face aux manques de données.

L'idée maîtresse réside dans le fait de transférer directement les sous-parties constructrices d'attributs d'un domaine vers un autre afin de conserver les avantages d'interprétabilité et de souplesse de la méthode d'origine.

Contexte : étude géologique d'un sous-sol

Pour l'entreprise pétrolière Total, pour détecter la présence d'hydrocarbures, il est nécessaire de bien comprendre sa structure.



Il s'agit, pour un géologue, de déterminer l'absence ou la présence de phénomènes géologiques locaux.

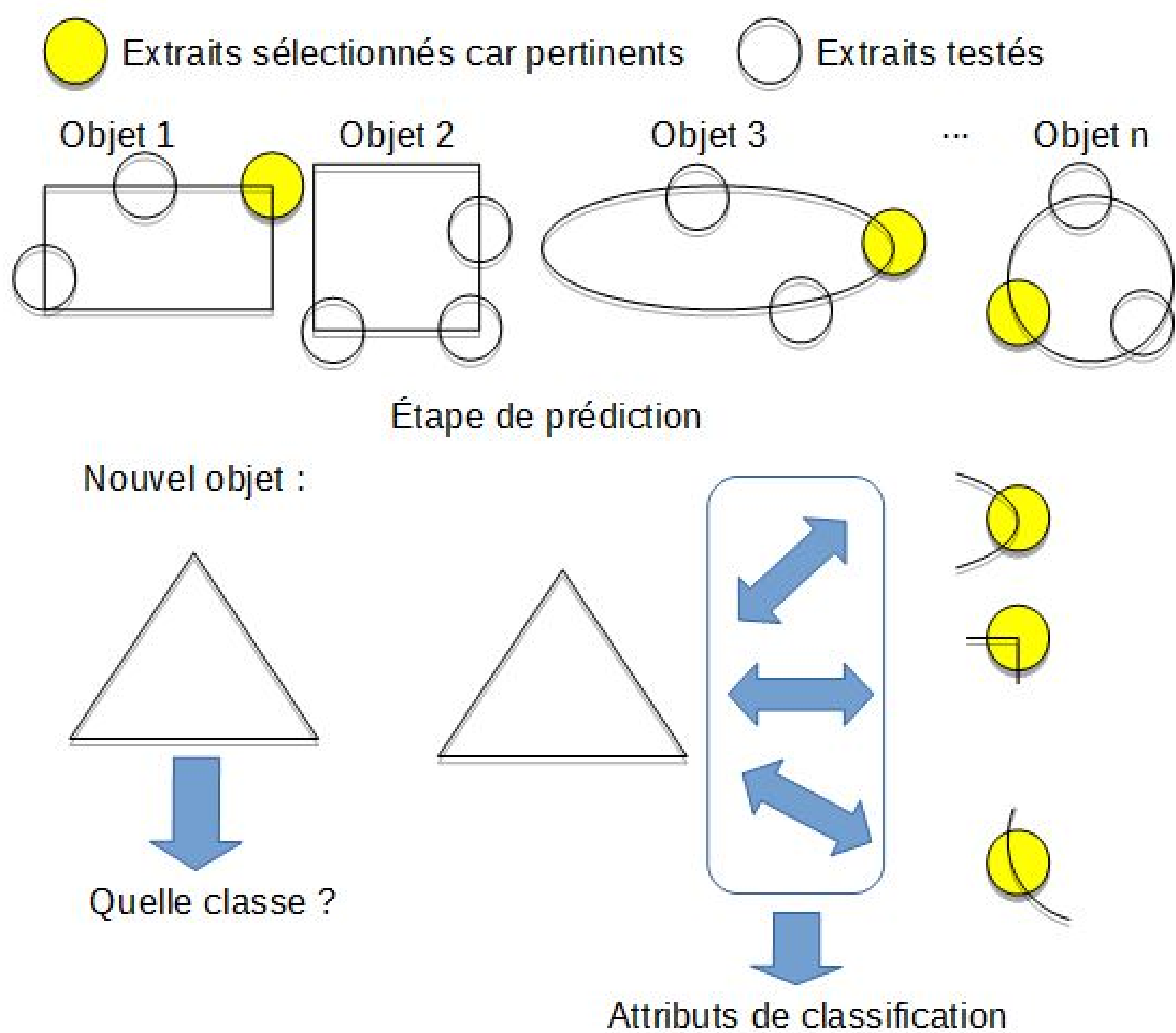
Jusqu'à maintenant, cette classification, comme le reste de l'étude, est réalisée manuellement par un géologue d'expérience qui connaît la zone. Cela représente une charge de travail conséquente, qu'une automatisation peut grandement faciliter.

Travaux liés

L'algorithme de classification supervisé d'objets 3D de [Meunier et al., 2017] est une première approche du problème issue de la notion de shapelets de [Refregier, 2001] adaptée ultérieurement par [Ye and Keogh, 2009], en utilisant des sous-parties discriminantes de la structure étudiée.

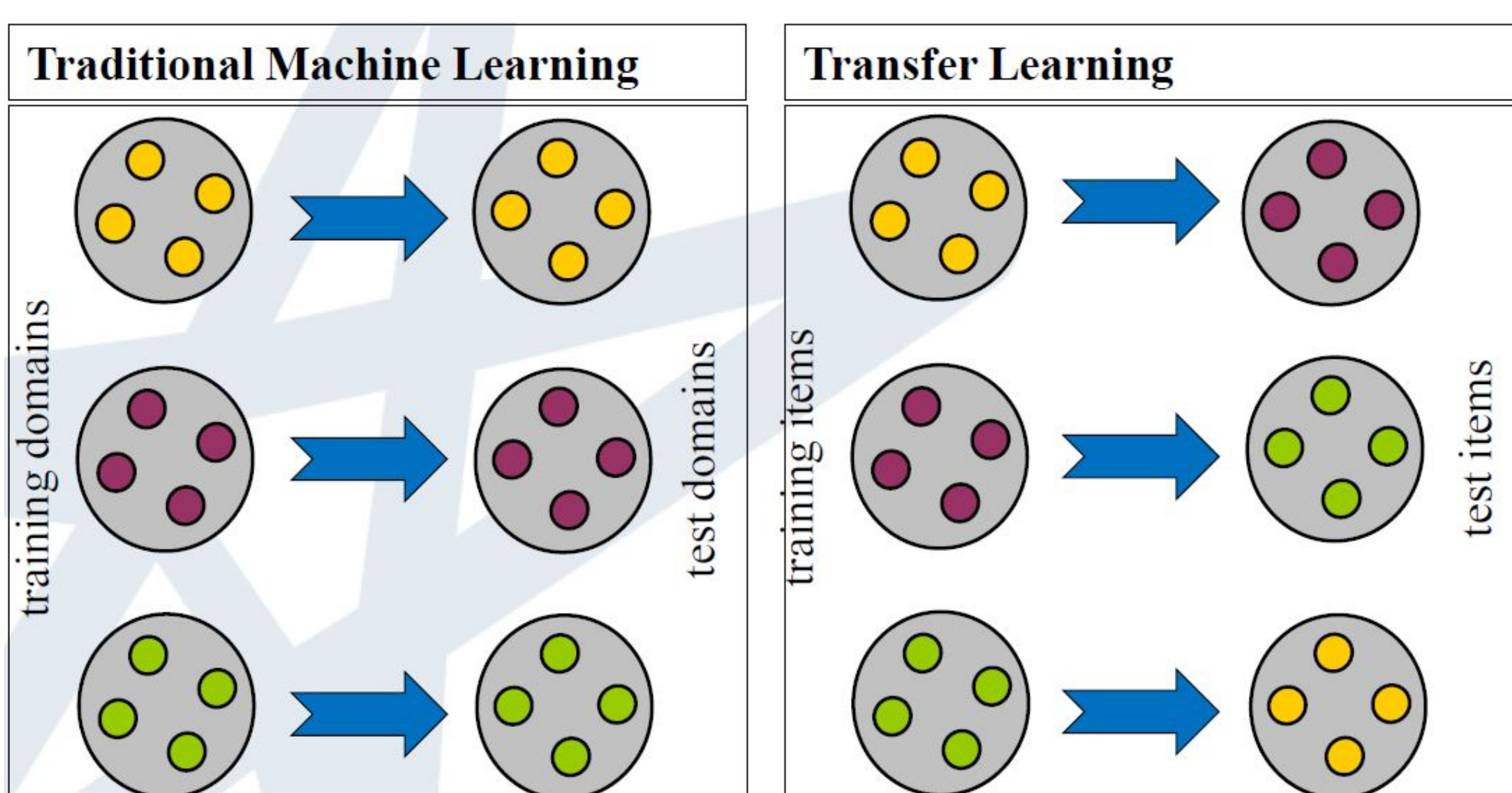
Si l'on schématise le principe en se restreignant à deux dimensions :

Étape de sélection des sous-parties discriminantes



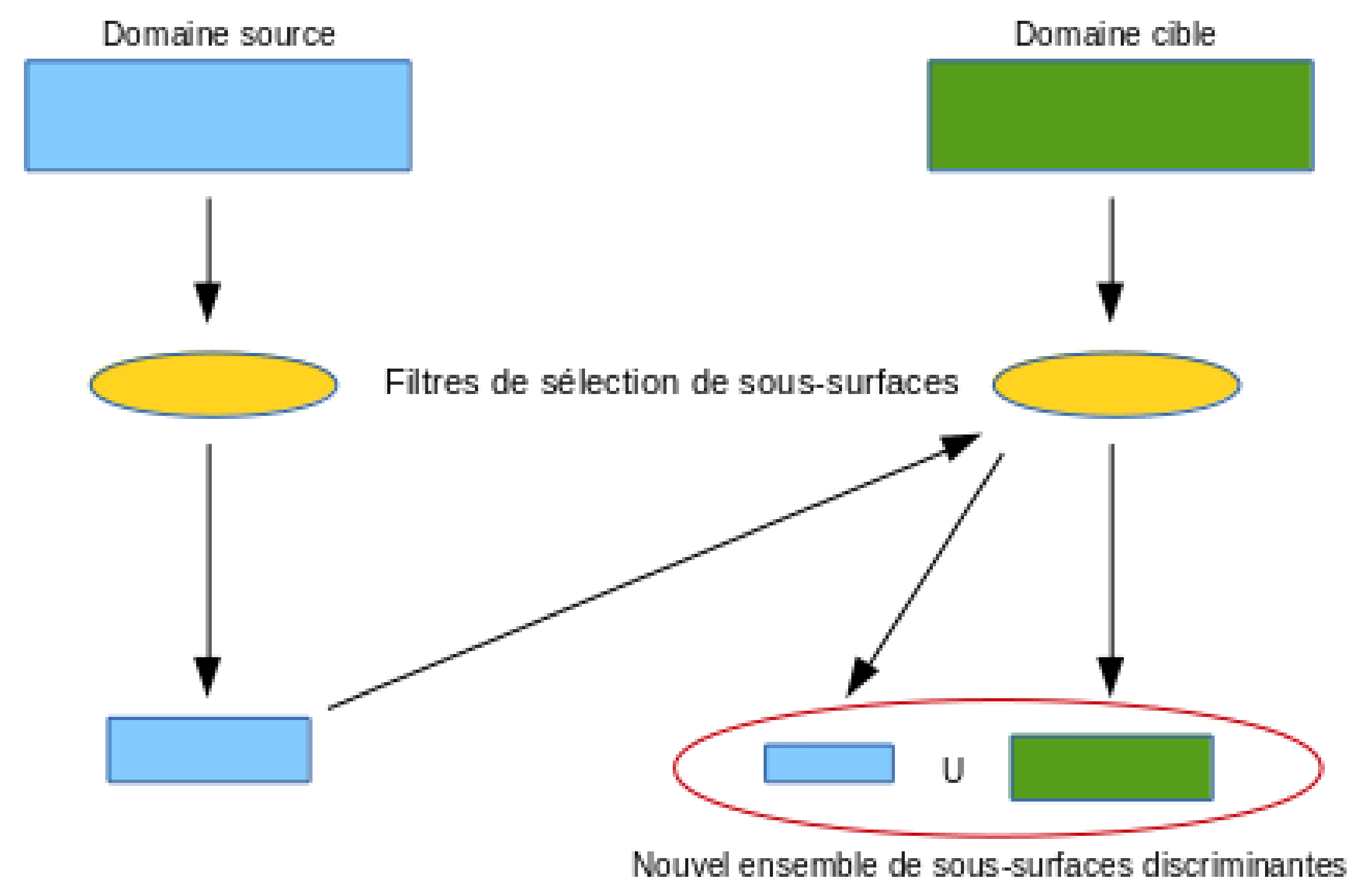
L'apprentissage par transfert (ou transfer learning)

Hypothèse des méthodes traditionnelles : les données d'apprentissage et les données de test sont issues du même domaine, de sorte que l'espace des attributs en entrée et la distribution des données sont les mêmes. Dans beaucoup de scénarios d'apprentissage supervisé, cette hypothèse ne tient pas.



Description de l'approche proposée

Soit un domaine source unique avec beaucoup d'exemples d'apprentissage (en fait une tâche de classification réalisée sur des données particulières, et soit un domaine cible avec seulement quelques données déjà annotées :



Pour être validés au sein de la classification cible, les extraits issus du domaine source doivent passer le test de pertinence (feature selection) du domaine cible.

Expérimentations : benchmark PSB¹

Face au potentiel déséquilibre de classe au sein de ces nouvelles expériences, nous choisissons d'utiliser la AUC (aire sous la courbe ROC - Receiver Operating Characteristic) pour évaluer les performances du modèle.

Nombre d'objets du domaine source	Nombre d'objets du domaine cible	Sans apprentissage par transfert	Avec apprentissage par transfert
60	40	0.616 ± 0.037	0.642 ± 0.034
	30	0.581 ± 0.049	0.617 ± 0.047
	20	0.545 ± 0.055	0.583 ± 0.065
80	40	0.611 ± 0.035	0.666 ± 0.036
	30	0.580 ± 0.041	0.625 ± 0.041
	20	0.548 ± 0.054	0.601 ± 0.060
100	40	0.616 ± 0.032	0.640 ± 0.034
	30	0.579 ± 0.045	0.615 ± 0.048
	20	0.544 ± 0.058	0.581 ± 0.069

Conclusions

Afin de faire face à un manque de données d'apprentissage, cette adaptation de l'algorithme d'origine vise à transférer de la connaissance issue d'autres apprentissages, dans le but d'améliorer le taux de classification. Proposant des performances supérieures aux méthodes actuelles, nous valorisons de cette manière les apprentissages réalisés par le passé, afin d'améliorer la classification en cours.

Prochainement testée sur des données réelles de Total, elle a encore besoin, pour être validée, de voir son intérêt confirmé par le métier, que la mise en pré-production devrait permettre d'ici peu.

Références

[Meunier et al., 2017] Meunier, F., Marsala, C., Castanié, L., and Conche, B. (2017). Classification d'objets 3d par extraction aléatoire de sous-parties discriminantes pour l'étude du sous-sol en prospection pétrolière. In *Extraction et Gestion des Connaissances*, pages 225–236.

[Refregier, 2001] Refregier, A. (2001). Shapelets : I. a method for image analysis. *Mon. Not. Roy. Astron. Soc.*, 338 :35.

[Ye and Keogh, 2009] Ye, L. and Keogh, E. (2009). Time Series Shapelets : A New Primitive for Data Mining. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 947–956, New York, NY, USA. ACM.