

Systeme pour la categorisation automatique des offres d'emploi en une typologie de fonctions

Julie Séguéla*,**

*Cédric-CNAM, 292 rue Saint-Martin, 75003 Paris, France

**Multiposting.fr, 33 rue Réaumur, 75003 Paris, France

jseguela@multiposting.fr

Résumé. Depuis les deux dernières décennies, l'augmentation du nombre de sites d'emploi sur Internet a accentué la nécessité de proposer des outils d'aide à la décision adaptés aux besoins des recruteurs. Cet article présente un système pour la catégorisation des textes d'offres d'emploi destinées à être diffusées sur Internet. Après un pré-traitement adapté des offres, les termes descripteurs sont choisis en fonction de leur pouvoir discriminant vis-à-vis des différentes classes ce qui permet de réduire leur nombre de manière significative. Les offres sont ensuite représentées par leurs coordonnées dans l'espace factoriel obtenu par analyse des correspondances et la classification réalisée dans un cadre supervisé à l'aide de SVM.

1 Introduction

Contexte. La multiplication rapide du nombre de supports Internet pour la publication des offres d'emploi a rendu indispensable l'évaluation et la comparaison des performances dans le cadre du développement d'outils d'aide à la décision pour les recruteurs. Les indicateurs de performance permettent de mesurer et comparer l'efficacité des sites d'emploi en décrivant les candidatures reçues d'un point de vue aussi bien quantitatif que qualitatif (nombre de retours, coût, etc.). Dans ce contexte, nous proposons un système permettant d'affecter automatiquement une fonction (selon une typologie prédéfinie) à toute offre d'emploi diffusée sur Internet, l'intérêt étant d'obtenir une nomenclature unique pour l'ensemble des annonces diffusées sur les plusieurs centaines de sites d'emploi existant. En effet, chaque site d'emploi dispose de sa propre nomenclature pour caractériser la fonction (métier au sens large), rendant très difficile l'agrégation d'information et l'exploitation des connaissances acquises en termes de performance des annonces suite à leur diffusion. Nous disposons d'une base de données d'annonces diffusées par Multiposting.fr¹ sur un grand nombre de sites d'emploi.

Afin de mettre au point l'algorithme de catégorisation, nous utilisons un corpus d'offres d'emploi rédigées en français et étiquetées manuellement par des recruteurs lors de leur publication sur un site d'emploi généraliste. Les étiquettes attribuées par les recruteurs appartiennent à une typologie prédéfinie par le site d'emploi.

1. Multiposting.fr est une plateforme de multidiffusion d'offres d'emploi sur Internet.