

Apprentissage par analyse linéaire discriminante des paramètres de fusion pour la recherche d'information multimédia texte-image

Christophe Moulin*, Christine Largeron*, Cécile Barat*,
Mathias Géry*, Christophe Ducottet*

*Université de Lyon, F-42023, Saint-Étienne, France ;
CNRS, UMR 5516, Laboratoire Hubert Curien, F-42023, Saint-Étienne, France ;
Université de Saint-Étienne, Jean-Monnet, F-42023, Saint-Étienne, France.

Résumé. Avec le développement du numérique, des quantités très importantes de documents composés de texte et d'images sont échangés, ce qui nécessite le développement de modèles permettant d'exploiter efficacement ces informations multimédias. Dans le contexte de la recherche d'information, un modèle possible consiste à représenter séparément les informations textuelles et visuelles et à combiner linéairement les scores issus de chaque représentation. Cette approche nécessite le paramétrage de poids afin d'équilibrer la contribution de chaque modalité. Le but de cet article est de présenter une nouvelle méthode permettant d'apprendre ces poids, basée sur l'analyse linéaire discriminante de Fisher (ALD). Des expérimentations réalisées sur la collection ImageCLEF montrent que l'apprentissage des poids grâce à l'ALD est pertinent et que la combinaison des scores correspondante améliore significativement les résultats par rapport à l'utilisation d'une seule modalité.

1 Introduction

L'exploitation automatique des documents multimédia est un sujet d'actualité. En effet, le volume de données multimédia numérisées ne cesse de croître. C'est le cas des données disponibles sur le Web, mais c'est aussi le cas de nombreux autres domaines comme par exemple le domaine médical, la vidéosurveillance, la géographie, les archives publiques et plus récemment le développement des livres numériques liés aux appareils tels que les ultraportables ou les tablettes tactiles. La plupart des systèmes actuels de traitement de données multimédia (tels que les moteurs de recherche), n'exploitent généralement qu'un seul type d'information et, le plus souvent, il s'agit de l'information textuelle.

Le traitement de cette information textuelle s'appuie sur des techniques performantes et efficaces, mises au point et perfectionnées depuis des décennies, et qui ont déjà fait leurs preuves (van Rijsbergen (1979); Manning et al. (2008)). Notamment, le modèle vectoriel de Salton (Salton et al. (1975)) est à la base de nombreux moteurs de recherche. Depuis les années 2000, des progrès significatifs ont été faits pour le traitement des informations visuelles. En particulier, le modèle des sacs de mots visuels (Csurka et al. (2004)), inspiré de l'approche textuelle,