
Apprentissage Actif avec une Méthode de Réordonnement pour l'Indexation et la Recherche de Vidéos

Bahjat Safadi et Georges Quénot

UJF-Grenoble 1 / UPMF-Grenoble 2 / Grenoble INP / CNRS, LIG UMR 5217, Grenoble, F-38041, France

Bahjat.Safadi@imag.fr

RÉSUMÉ. La recherche de vidéos peut être faite en ordonnant les échantillons en fonction de scores de probabilité produits par des classifieurs. Il est souvent possible d'améliorer la performance des systèmes par un réordonnement de ces échantillons. Dans cet article, nous proposons une telle méthode et nous proposons également la combinaison de cette méthode avec un apprentissage actif pour l'indexation de vidéos. Les résultats expérimentaux montrent que la méthode de réordonnement proposée a été en mesure d'améliorer la performance du système avec une augmentation d'environ 16-22% du score en moyenne sur la tâche d'indexation sémantique en TRECVID 2010. En outre, elle a amélioré la performance du système d'indexation des vidéos par apprentissage actif, en considérant l'aire sous la courbe (AUC) comme mesure d'évaluation de la performance de l'apprentissage actif. Notre méthode de réordonnement améliore la performance d'environ 20% en moyenne sur la collection TRECVID 2007.

ABSTRACT. Video retrieval can be done by ranking the samples according to their probability scores that were produced by classifiers. It is often possible to improve the retrieval performance by re-ranking the samples. In this paper, we proposed such a method and we combined this method with active learning for video indexing. Experimental results showed that the proposed re-ranking method was able to improve the system performance with about 16-22% in average on TRECVID 2010 semantic indexing task. Furthermore, it improved significantly the performance of the video indexing system based active learning; by considering the Area Under Curve (AUC) as a metric measure for the performance of the active learning, our re-ranking method improved the performance with about 20% in average on TRECVID 2007.

MOTS-CLÉS : Indexation et de recherche des vidéos, Apprentissage Actif, Réordonnement

KEYWORDS: Video Indexing and Retrieval, Active Learning, Re-ranking

1. Introduction

L'indexation sémantique a constitué un champ de recherche très actif des dernières années. Elle consiste à construire automatiquement une description des vidéos pour en permettre une recherche par le contenu. Malgré le nombre important de travaux sur le sujet, l'indexation par le contenu de larges collections de vidéos autorisant la recherche de vidéos ou la navigation au sein d'une collection est toujours un problème ouvert.

L'indexation sémantique est généralement effectuée par apprentissage supervisé : un système de classification est entraîné à partir d'exemples positifs et négatifs d'un concept cible (dans un ensemble de développement). Un modèle est construit et est ensuite utilisé pour prédire la présence du concept donné dans de nouveaux échantillons (dans un ensemble de test). La prédiction est faite sous la forme d'un score homogène à la probabilité pour chaque échantillon de contenir le concept cible. Ces scores sont utilisés pour ordonner les échantillons de l'ensemble de test. Ils sont calculés indépendamment pour chaque échantillon sur la base des informations disponibles dans l'ensemble de développement. Il est souvent possible d'améliorer l'indexation ou la recherche en réordonnant les échantillons de l'ensemble de test en considérant le classement initial et la structure de cet ensemble.

En pratique, le volume de données qui peut être annoté manuellement pour l'entraînement des systèmes est limité. Une stratégie consiste à sélectionner les échantillons à annoter pour que l'annotation soit faite sur des données les plus utiles possibles (Angluin, 1988). L'apprentissage actif peut être utilisé pour choisir ces échantillons à annoter. Un utilisateur étiquette subjectivement un échantillon comme positif ou négatif. L'ensemble de ces échantillons est alors utilisé pour entraîner un classifieur qui sera ensuite utilisé pour faire une prédiction pour les échantillons non encore annotés. Les échantillons ayant les scores les plus élevés ou les plus incertains sont sélectionnés comme les plus informatifs pour le système et proposés à l'annotation.

Très souvent, les données utilisées pour l'indexation de vidéos sont très déséquilibrées, avec une proportion très faible d'échantillons positifs pour un concept donné. Cette caractéristique est particulièrement problématique dans le cadre de l'apprentissage supervisé classique. Dans le contexte des campagnes d'évaluation TRECVID par exemple, la classe minoritaire représente typiquement moins de 1% (Smeaton *et al.*, 2006). Une approche proposée pour gérer le problème des données non équilibrées consiste à générer un ensemble équilibré en sous-échantillonnant la classe majoritaire dans l'ensemble de développement (Bishop, 2007). De plus, il est possible de compenser la perte d'information induite par le sous-échantillonnage de la classe majoritaire par la génération de plusieurs ensemble de données équilibrés, et de fusionner les sorties produites par les classifieurs construits sur ces ensembles. Une combinaison de l'apprentissage actif avec cette approche à classifieurs multiples permet de significativement améliorer l'efficacité de l'annotation multimédia, en particulier dans le cadre de données fortement déséquilibrées comme celle disponibles dans TRECVID (Safadi *et al.*, 2010).

