
Comparaison du modèle vectoriel et de la pondération $tf*idf$ associée avec une méthode de propagation d'activation

Christophe Brouard

UPMF-Grenoble2/CNRS
LIG UMR 5217/équipe AMA
Grenoble, France
Christophe.Brouard@imag.fr

*RÉSUMÉ. L'objet de ce papier est de montrer qu'il est possible de mettre le modèle vectoriel et la pondération $tf*idf$ associée en correspondance avec le calcul d'une résonance dans un réseau associatif basé sur une méthode très simple de propagation d'activation. Nous décrivons un réseau associatif associant termes et documents puis un calcul de résonance entre une requête et un document dans ce réseau. La pondération $tf*idf$ apparaît naturellement dans le calcul et nous expliquons comment chacun des facteurs s'interprète dans la propagation d'activation. Nous montrons que ce calcul, comme le cosinus, correspond au produit de l'évaluation de la spécificité et de l'exhaustivité du document vis-à-vis de la requête. La comparaison expérimentale réalisée sur les corpus CLEF3 et TREC3 révèle que, si l'on choisit adéquatement les fonctions qui régissent l'activation et la propagation d'activation, la méthode basée sur la résonance obtient des performances similaires à celles du modèle Okapi-BM25.*

*ABSTRACT. We show in this paper that it is possible to establish a correspondence between the vector space model and a spreading activation method based on a resonance computation in a network linking the terms and the documents. We describe the network and the resonance computation between a query and a document in the network. The $tf*idf$ weighting scheme is naturally introduced in the computation and we explain how these factors can be interpreted in the spreading activation method. Then, we show that, like the cosine in the vector space model, the resonance computation corresponds to the product between the measures of specificity and exhaustivity of the document for the query. The experiments carried out on the CLEF3 and TREC3 datasets show that the performances of the model based on resonance are similar to those of the Okapi-BM25 model.*

MOTS-CLÉS: modèle de RI, fonction de correspondance, méthode de propagation d'activation

KEYWORDS: IR model, scoring function, spreading activation method
