TOPICAL ORGANIZATION OF USER COMMENTS AND APPLICATION TO CONTENT RECOMMENDATION

Vidit Jain
viditj@yahoo-inc.com
Yahoo! Labs Bangalore

Esther Galbrun
galbrun@cs.helsinki.fi
Helsinki Institute for Information Technology
Department of Computer Science
University of Helsinki

A news article may receive thousands of comments from its readers.

Organizing them by *semantic topics* enables the user to selectively browse comments on a topic.

It allows to discover significant topics of discussion in comments and to *explicitly capture the immediate interests of the user* even when they are not logged in.

Hence, we propose

- an algorithm to build a topical organization,
- a new paradigm of recommending content for comments being read,
- evidence for preference of these cluster-to-article recommendations over the standard article-to-article recommendations.

Illustration of comment clusters: *Each rectangle corresponds to one semantically coherent cluster of comments and shows the important terms appearing in it as a summary. On selecting one of them (in darker blue) only the comments belonging to that specific cluster are shown to the user. At the bottom of the displayed set of comments, a related content is recommended (highlighted in red).*

▶ TOPICAL ORGANIZATION

Organizing the comments associated to a news article into clusters based on occuring named-entities (cf. publication).

An empirical comparison showed our Entity Based Clustering (EBC) method to be better suited than k-means or METIS in a practical setting.

▶ RECOMMENDATION SCHEMES

Three schemes to recommend an article based on TF-IDF similarity matching.

AA: article to article

CA: cluster to article

CCA: cluster to article via best matching cluster
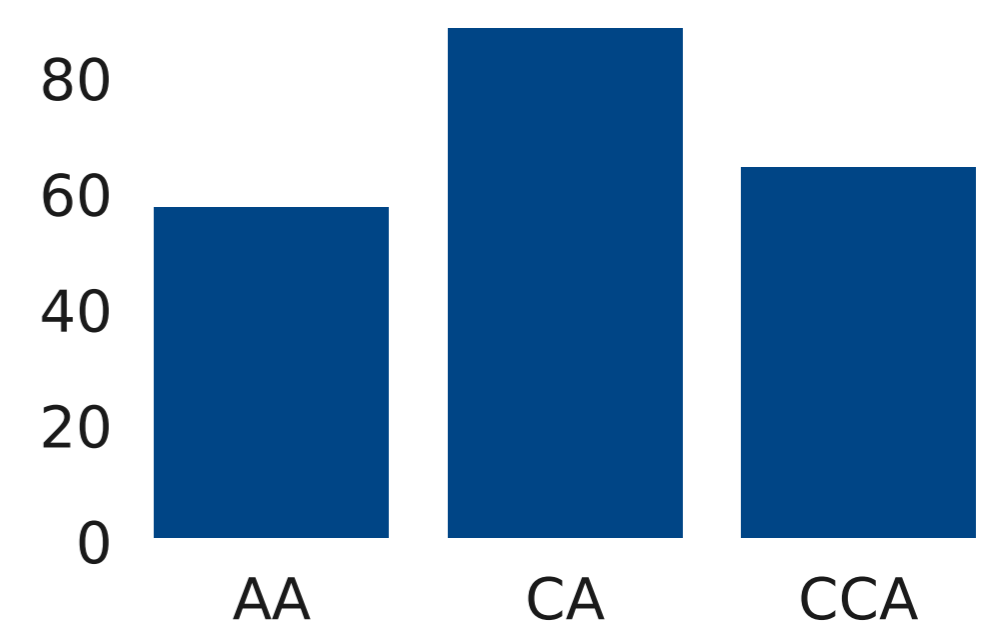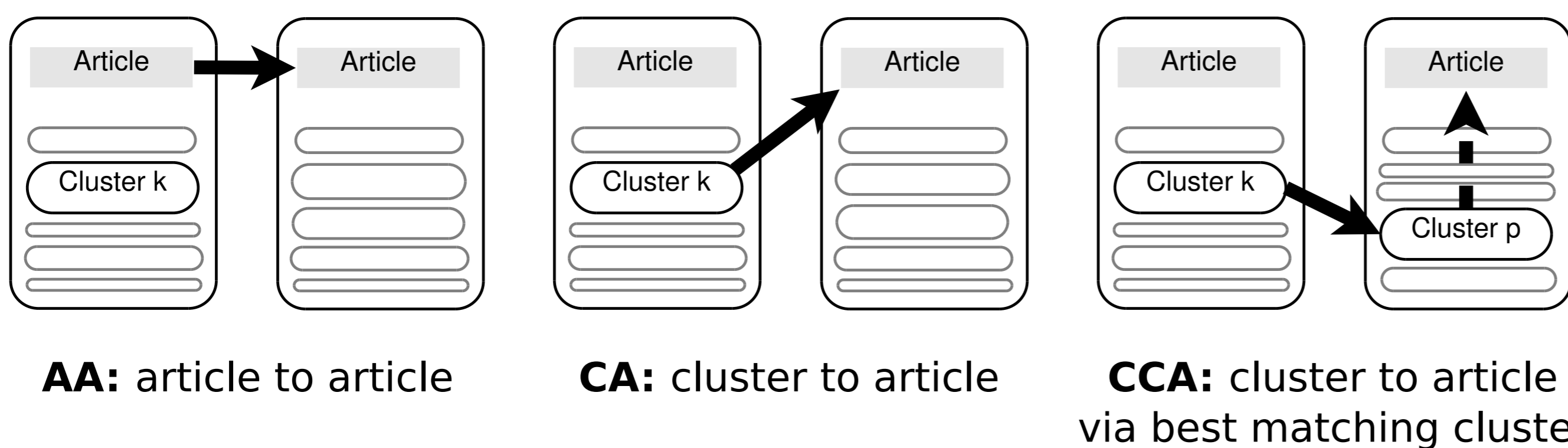
User study: *Number of instances for which the scheme is preferred. Difference significant at level 0.05*

▶ ENSEMBLE SYSTEM

The different schemes correspond to different scenarios.

- AA works well for on-topic comments
- CA help identify focused and latent topics in comments
- CCA fails when matching to off-topic comments for the target article

An ensemble system selects one of the recommendation scheme based on the source document.
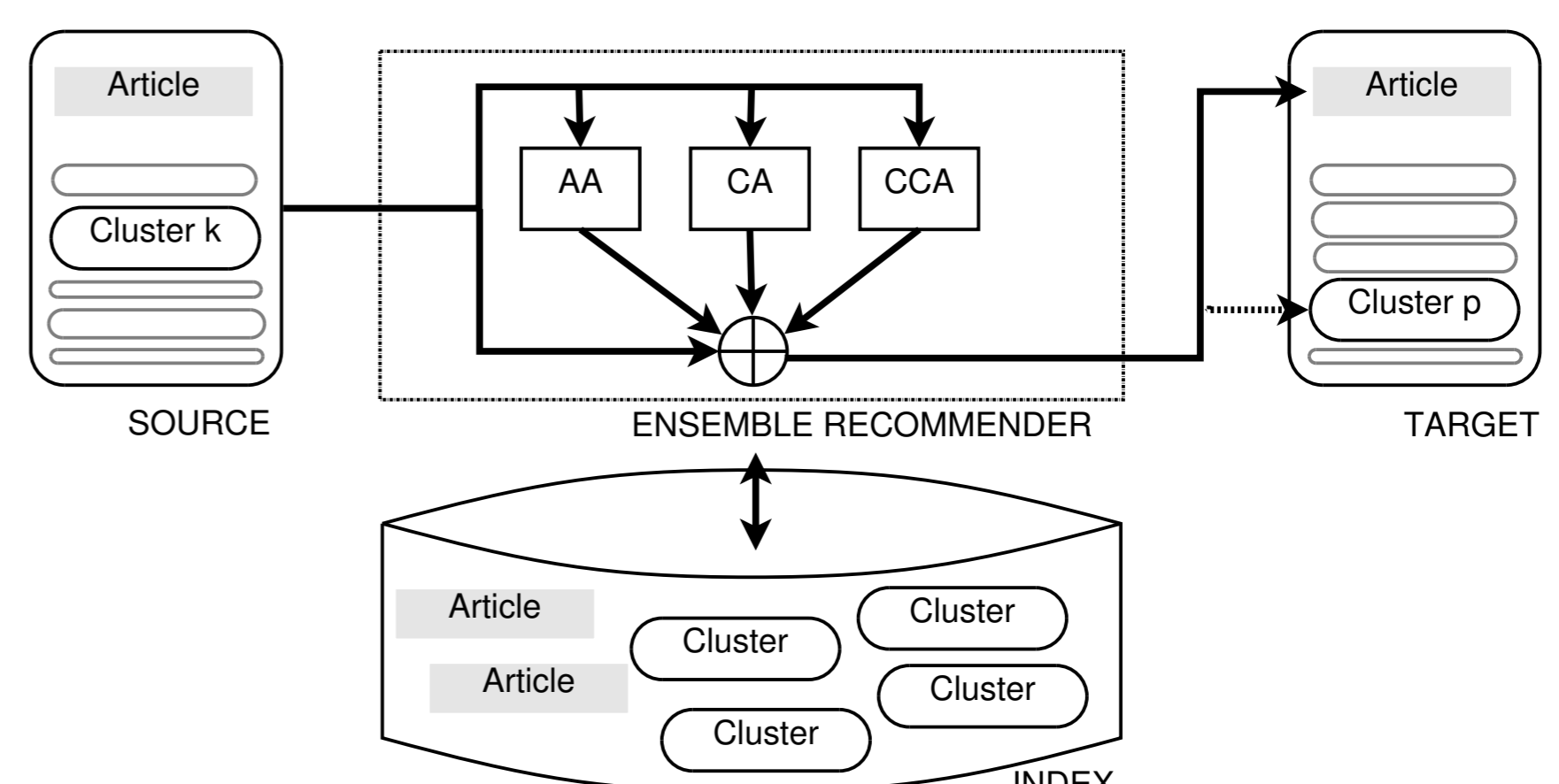
	Multi-class SVM	Naïve-Bayes	Gaussian Process
Improvement in agreement with the preferred scheme over CA	-3.6%	+1.8%	**+8.5%**