

Max-difference maximization criterion: A feature selection method for text categorization

April 28 2023



Credit: Petr Kratochvil/public domain

For text categorization, it is necessary to select a set of features (terms) with high discrimination by using feature selection. In text feature selection, Accuracy2 (ACC2) treats terms with same absolute document rate difference but different discrimination equally, which is unreasonable. Existing improved methods (normalized difference

measure (NDM), max-min ratio (MMR) and trigonometric comparison measure (TCM)) based on ACC2 may confuse the importance of rare and sparse terms on account of challenge for parameter selection.

To solve the problems, a research team led by Li Zhang published their new research in *Frontiers of Computer Science*.

The team proposed max-difference maximization criterion (MDMC) , which introduces a new weight based on class information occupancy and combines it with ACC2 to estimate the importance of terms. As a result, MDMC can avoid overestimate of sparse terms.

In the research, they analyze the weight distributions of methods (ACC2, NDM, MMR, TCM and MDMC) and intuitively show the mechanism of MDMC to estimate the importance of terms, which is shown in [online resources](#). Experiments demonstrate that MDMC is capable of catching more discriminant terms without any parameter than other filter ones regardless of classifier, and shows its superiority over other dimensionality reduction methods (improved sine cosine algorithm (ISCA) , [principal component analysis](#) (PCA) and non-negative matrix factorization (NMF)).

More information: Lingbin Jin et al, Max-difference maximization criterion: a feature selection method for text categorization, *Frontiers of Computer Science* (2023). [DOI: 10.1007/s11704-022-2154-x](https://doi.org/10.1007/s11704-022-2154-x)

Provided by Higher Education Press

Citation: Max-difference maximization criterion: A feature selection method for text categorization (2023, April 28) retrieved 23 April 2024 from <https://techxplore.com/news/2023-04-max-difference-maximization-criterion-feature->

[method.html](#)

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.