Image Retrieval by Cross-Media Relevance Fusion

Jianfeng Dong[1], Xirong Li[2], Shuai Liao[2]

Jieping Xu[2], Duanqing Xu[1], Xiaoyong Du[2]

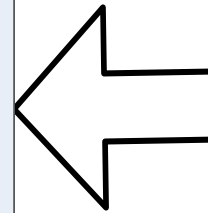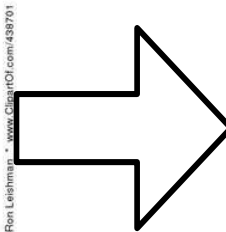[1]Zhejiang University

[2]Renmin University of China

Task

Produce a relevance score for a given **image-query** pair.
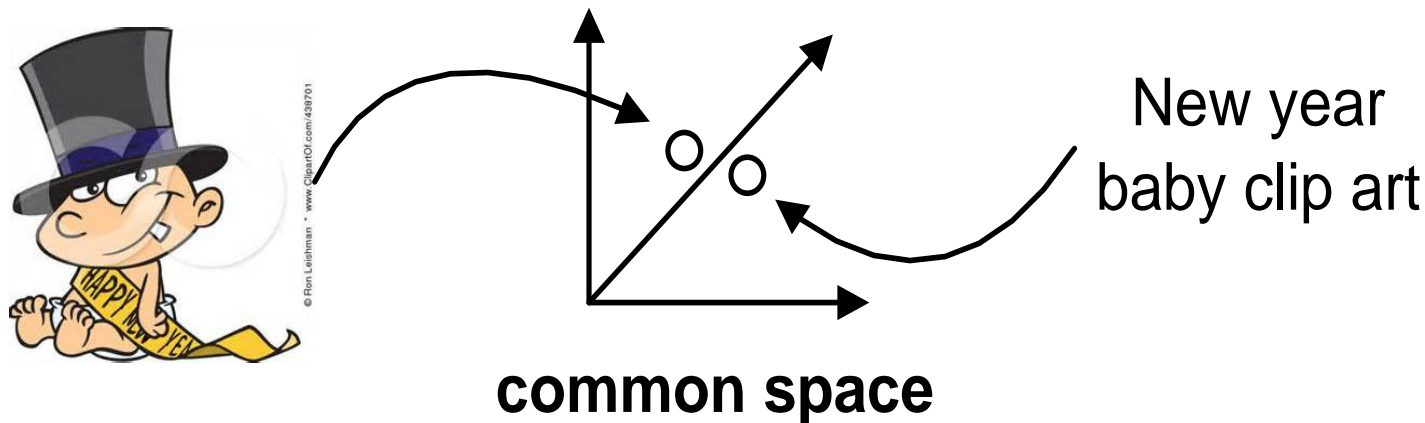
Image　　　　　　　　　　　　　　Query

 ? ← new year baby clip art

Key question: How to compute cross-media relevance?

Cross-media relevance computation

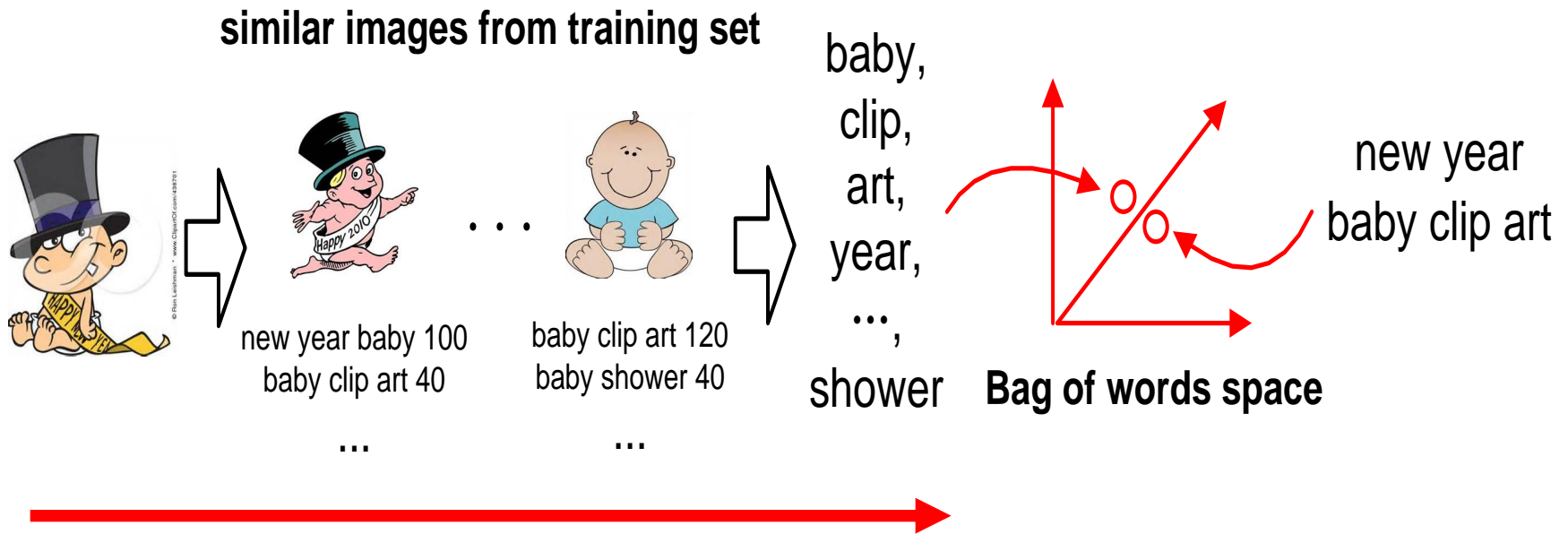Image and query have to be represented in a common space

common space

New year
baby clip art

Choices of the common space:
1. Bag of words sapce
2. Visual feature space
3. Learned space

Cross-media 1: Image2Text

Compute relevance in a bag of words space.

similar images from training set

new year baby 100
baby clip art 40

baby clip art 120
baby shower 40

baby, clip, art, year, ...; shower

Bag of words space

new year baby clip art

Cross-media 2: Text2Image

Compute relevance in a visual feature space.

similar queries form training set

new year baby clipart   new year 2013 art

130   300

60   90

Visual feature space

New year baby clip art
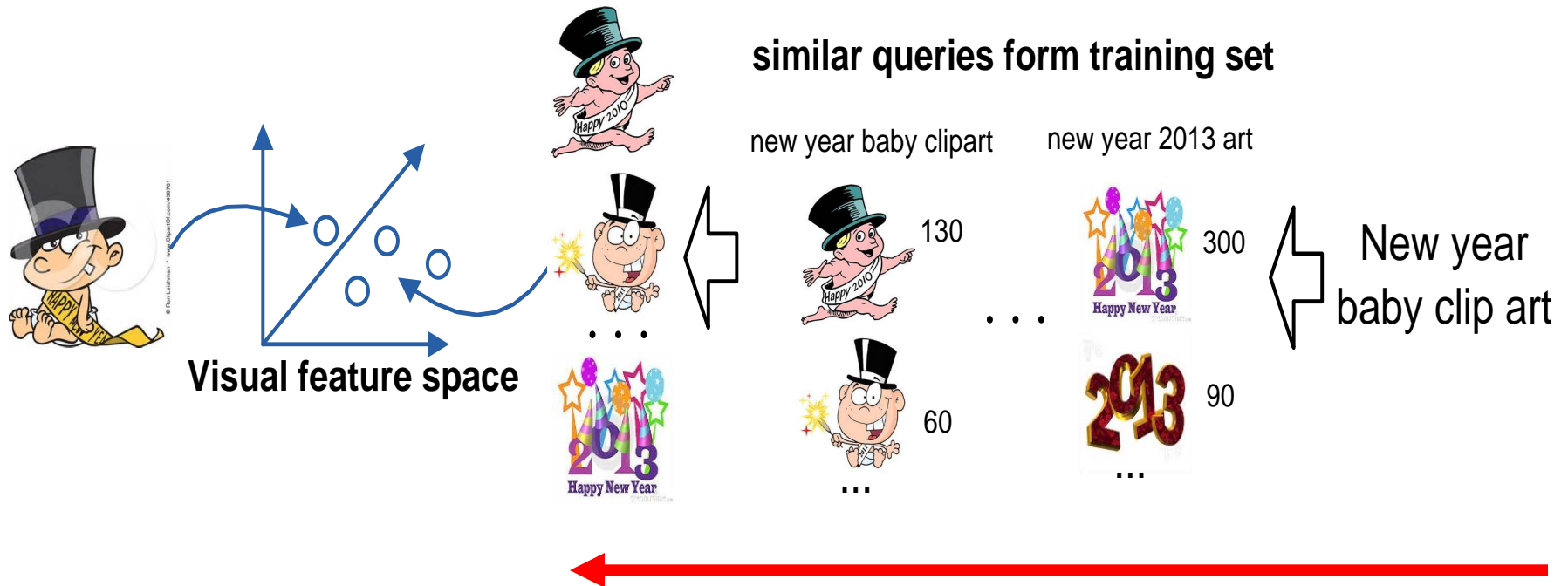
Cross-media 3: Parzen window

Hypothesis: relevant images are visually similar

Parzen-Window Density Estimation

all test images of this query

Visual feature space
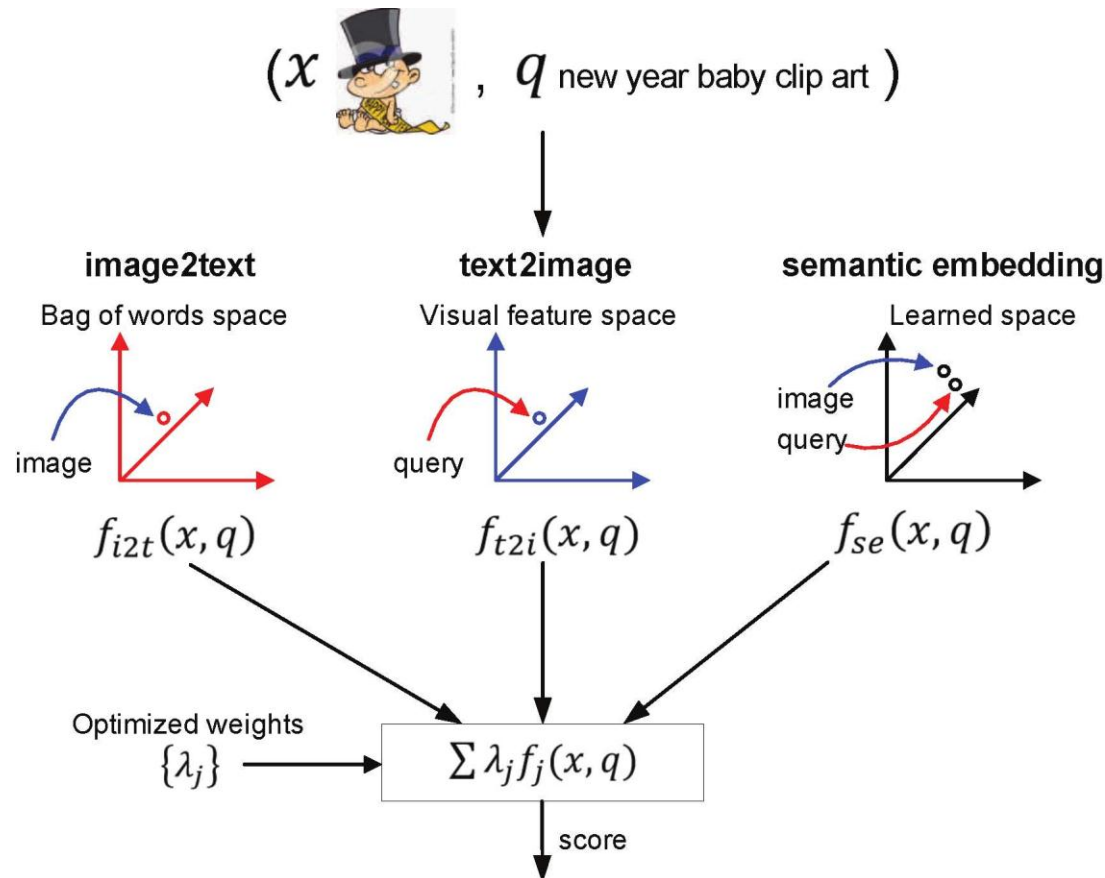
New year baby clip art
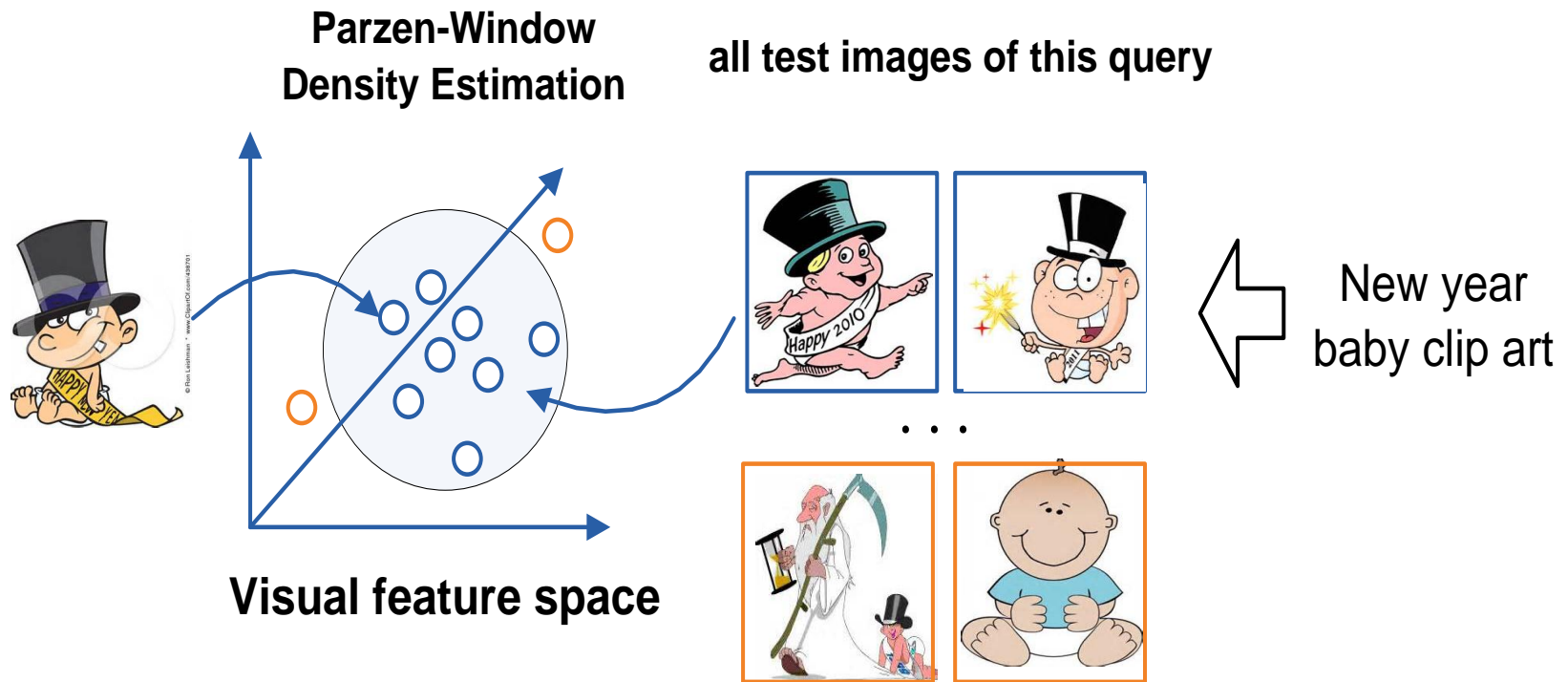
Cross-media 4: Semantic Embedding

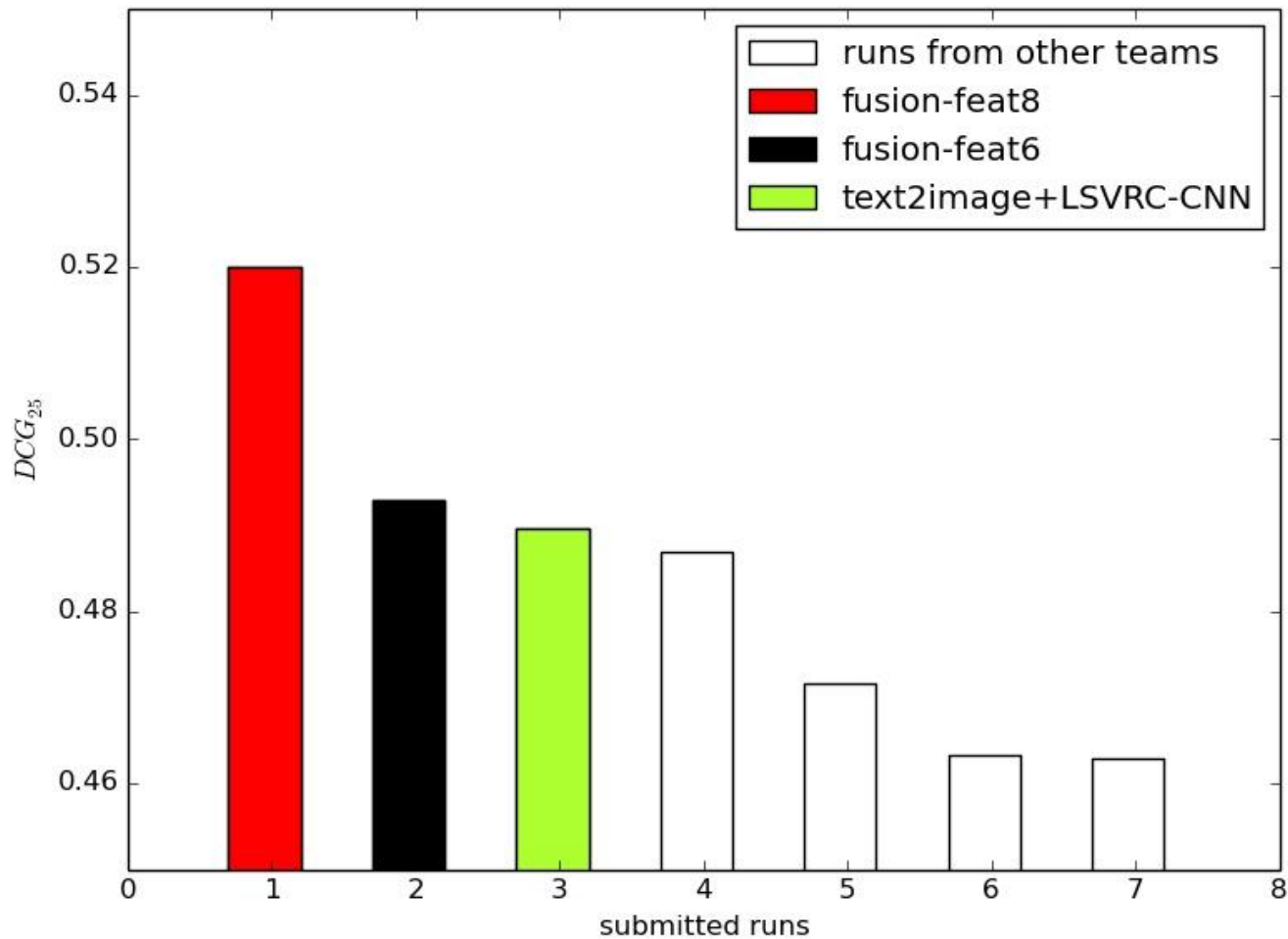Compute relevance in a learned semantic space.
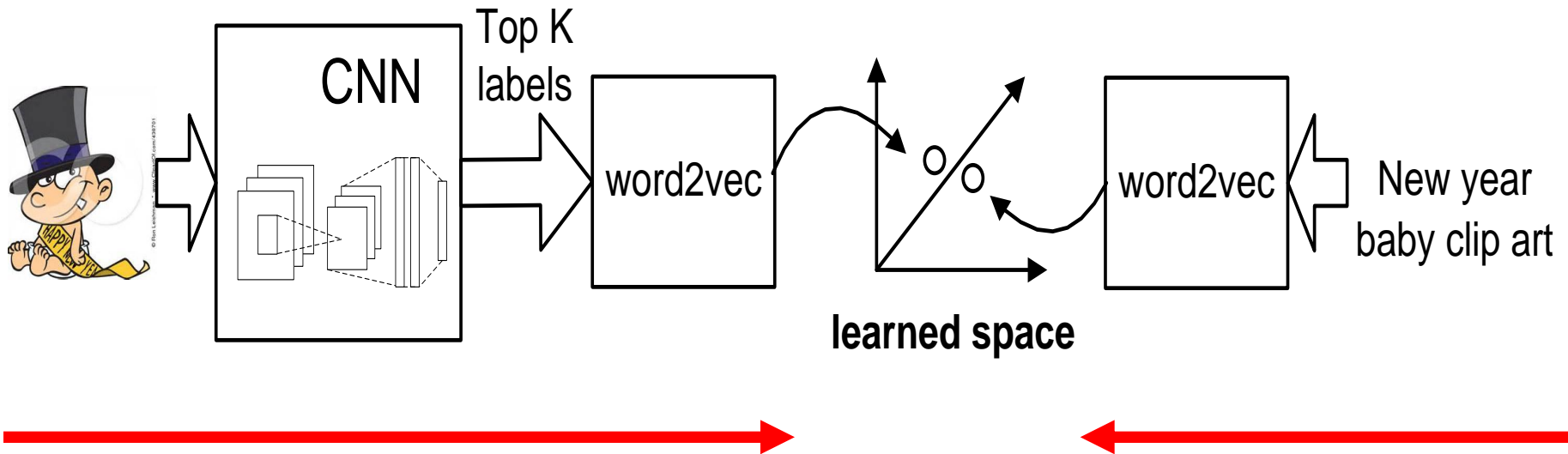
Cross-media relevance fusion Framework

Fusion for better performance.

Official evaluation

Our runs are ranked at the top.

Take-home messages

- Text2Image (Parzen window) is a winning component.

- Relevance fusion gives the best performance.

https://github.com/danieljf24/cmrf