
CondorcetFuse

Condoret voting for run fusion

Evalutation and comparison of an implementation with other fusion
strateg

Indice

1	Abstract	2
2	Introduction	2
3	Basic fusion strategies	2
4	Condorcet fusion	3
4.1	The Condorcet Graph	3
4.2	Condorcet paths	3
5	Implementation	3
6	Evaluation	3
6.1	Evaluation metrics	3
6.2	Results	3
7	Conclusions	3

1 Abstract

In Information Retrieval, data fusion is the combination of the results of independent searches on a document collection into one single output result set.

It has been shown in the past that this can greatly improve retrieval effectiveness over that of the individual results.

The aim of this project is to show a possible implementation of basic fusion strategy and an advanced one.

2 Introduction

Retrieval experiments uses data from the TREC ad hoc collection.

The following procedure illustrates the organization of the work:

- **Indexing** (with or without query expansion)
- **Retrieval** (10 different retrieval models)
- **Fusion strategy** (with or without cutting the run at depth 100)
 - 6 basic strategies
 - Condorcet fusion (advanced strategy)

Retrieval models
BB2
BM25
DLH13
Hiemstra_LM
IFB2
TF_IDF
DFIC
DFIZ
DirichletLM
InL2

Tabella 1: Retrieval models used

3 Basic fusion strategies

The following table shows a list of basic fusion methods:

Basic fusion methods	New score
CombMNZ	SUM(Individual similarities)*Nonzero similarities
CombSUM	SUM(Individual similarities)
CombMIN	MIN(Individual similarities)
CombMAX	MAX(Individual similarities)
CombMED	MED(Individual similarities)
CombANZ	SUM(Individual similarities)/Nonzero similarities

Tabella 2: Basic fusion methods used

4 Condorcet fusion

The Condorcet voting algorithm is a majoritarian method which specifies that the “winner” of the fusion is the document(s) that beats or ties with every other document in a pair-wise comparison between the input systems (i.e. runs).

4.1 The Condorcet Graph

Given 10 models of retrieval with n documents, the corresponding Condorcet graph $G = (V, E)$ has one vertex for each of the n documents.

For each document pair (x, y) , there exists an edge from x to y (denoted by $x \rightarrow y$) if x would have at least a score equal as y in a head-to-head contest.

Cycles can simply be viewed as ties.

The relative ordering of documents within a cycle is only of secondary importance, whereas their ordering with respect to the rest of the documents is of primary importance.

4.2 Condorcet paths

A Condorcet-consistent hamiltonian path (or condorcet path) is any hamiltonian path through the Condorcet graph.

The goal is to efficiently find such a path.

5 Implementation

The implementation of Condorcet used quicksort, with the following algorithm as comparing function:

```
count = 0
for each of the k search systems do:
    if sys i ranks d1 above d2, count++
    if sys i ranks d2 above d1, count--
    if count > 0, rank d1 better than d2
    else rank d2 better than d1
```

6 Evaluation

The evaluation criteria based on the given pool used two binary relevance scores: Relevant and Non-Relevant. The documents left out from the pool were considered to be non relevant.

6.1 Evaluation metrics

6.2 Results

Fusion methods	MAP
CombMNZ	
CombSUM	
CombMIN	
CombMAX	
CombMED	
CombANZ	
Condorcet	

Tabella 3: Mean Average Precision for the 10 fused runs

7 Conclusions