MRR: an Unsupervised Algorithm to Rank Reviews by Relevance

Vinicius Woloszyn   Henrique D. P. dos Santos   et al.

Department of Computer Science
Federal University of Rio Grande do Sul
and Pontifical Catholic University of Rio Grande do Sul

2017 IEEE/WIC/ACM International Conference on Web Intelligence

Leipzig, August 24, 2017

Introduction

- Many works address the problem of ranking documents by their relevance.
- Most of them rely on supervised algorithms such as classification and regression.
  - Annotated: Neural Network, SVM
  - Statistics: TF-IDF, Readability, POS-Tag

Introduction

- The quality of results produced by supervised algorithms is dependent on the existence of a large, domain-dependent training data set.
  - Amazon, Yelp
  - Netflix, IMDB
- Unsupervised methods are an attractive alternative to avoid the labor-intense and error-prone task of manual annotation of training datasets.

Graph-based

- Vertices are the documents (review), and the edges are defined in terms of the similarity between pairs of documents (ratings score and textual).

$$f(u, v) = \alpha * sim\_txt(u, v) + (1 - \alpha) * sim\_star(u, v) \tag{1}$$

- α : tune similarity function

Similarity Functions

- Textual
  Cosine similarity of TF-IDF vectors

$$sim\_txt(u, v) = cos(tfidf(t.t), tfidf(v.t)) \tag{2}$$

- Stars
  Euclidean distance normalized by Min-Max scaling

$$sim\_star(u, v) = 1 - \frac{|u.rs - v.rs| - min(rs)}{max(rs) - min(rs)} \tag{3}$$

MRR - Ranking documents by their relevance

Graph Centrality

- Hypothesis: a relevant document has a high centrality index since it is similar to many other documents.
- Centrality index produces a ranking of vertices' importance, indicating the ranking of the most relevant document.

MRR - Graph-Specific Similarity Threshold

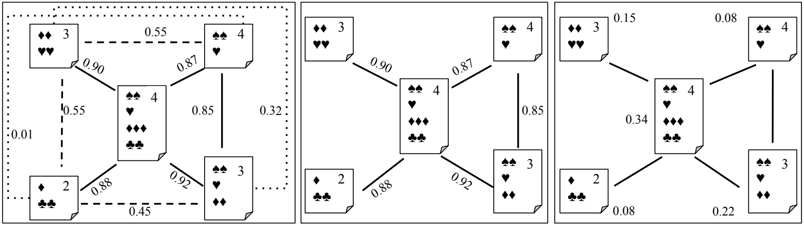
Graph Pruning

- Centrality is dependent on the existence of edges between nodes.
- Prune the graph based on a minimum similarity between review.
- $\overline{E}$: mean of graph similarity

$$W'(u,v) = \begin{cases} 1, & f(u,v) \geq \overline{E} * \beta \\ 0, & otherwise \end{cases} \tag{4}$$

- β : tune prune function

Main steps of the MRR algorithm



(A) Similarity Function (B) Graph-Specic Threshold (C) PageRank Scores

- (A) Builds a similarity graph G between pairs of documents;
- (B) Prune by removing all edges lower than the similarity threshold;
- (C) Employ PageRank to obtain the centrality scores;

MRR Algorithm

Algorithm 1 - MRR Algorithm (R, α, β): S

for each $u, v \in R$ **do**
    $W[u, v] \leftarrow \alpha * sim\_txt(u, v) + (1\text{-}\alpha) * sim\_star(u, v)$
end for
$\overline{E} \leftarrow mean(W)$
for each $u, v \in R$ **do**
    if $W[u, v] \geq \overline{E} * \beta$ **then**
        $W'[u, v] \leftarrow 1$
    else
        $W'[u, v] \leftarrow 0$
    end if
end for
$S \leftarrow PageRank(W')$
Return S

Experiment Design

- Dataset: reviews (rating score and text) of electronics and books from the Amazon website.
- Gold Standard: Human perception of helpfulness:

$$h(r \in R) = \frac{vote_{+}(r)}{vote_{+}(r) + vote_{-}(r)} \tag{5}$$

- Metric: Normalized Discounted Cumulative Gain as NDCG@n

Amazon Dataset

	Electronics	Books
Votes	48.20 ($\pm$ 302.84)	29.71 ($\pm$ 73.58)
Positive	40.12 ($\pm$ 291.99)	20.60 ($\pm$ 64.18)
Negative	8.08 ($\pm$ 22.27)	9.11 ($\pm$ 21.44)
Rating	3.73 ($\pm$ 1.50)	3.41 ($\pm$ 1.54)
Words	350.32 ($\pm$ 402.02)	287.44 ($\pm$ 273.75)
Products	383	461
Total	19,756	24,234

Table: Profiling of the Amazon dataset.

Experiments:

- Baselines comparison;
- Graph-Specific Threshold Assessment;
- Parameter Sensibility; and
- Run-time Performance.

Experiment Design

Baselines:

- TSUR et al. (2009) as REVRANK;
  - Core Virtual Review (200 most frequent words),
  - Rank by similarity distance to Core
- Wu et al. (2011) as PR_HS_LEN;
  - Sentences similarity based on POS-Tags,
  - PageRank, Hits and Length
- SVM Regression:
  - a) textual features TF-IDF and the star score,
  - b) the same features used by Wu et al. (2011)

Relevance Ranking Assessment

	NDCG@1	NDCG@5
SVM_WU	0.80770	0.91817
SVM_TFIDF	**0.85539**	**0.93119**
REVRANK	0.66052	0.68172
PR_HS_LEN	0.72689	0.77131
MRR	0.79877	0.81876

Table: Mean Performance on Book Reviews

- MRR statistically outperformed all unsupervised baselines

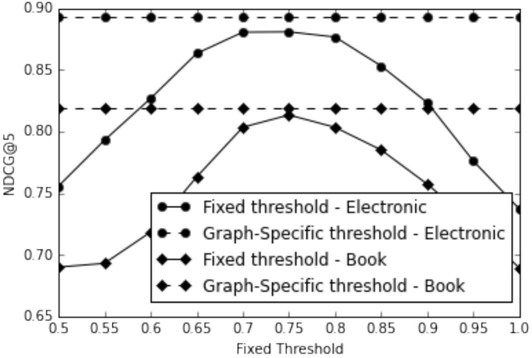
Relevance Ranking Assessment

	NDCG@1	NDCG@5
SVM_WU	0.76416	0.91535
SVM_TFIDF	0.88986	**0.94621**
REVRANK	0.67903	0.72133
PR_HS_LEN	0.87434	0.87184
MRR	**0.89403**	0.89246

Table: Mean Performance on Electronic Reviews

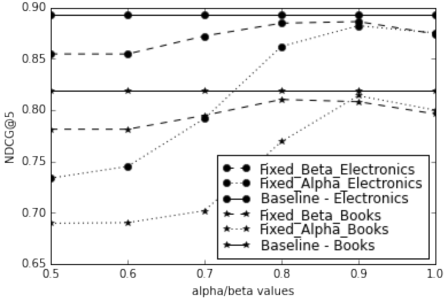
- MRR statistically outperformed all unsupervised baselines
- MRR is comparable to supervised methods

Graph-Specific Threshold Assessment



- MRR performance is always better using a Graph-Specific threshold.

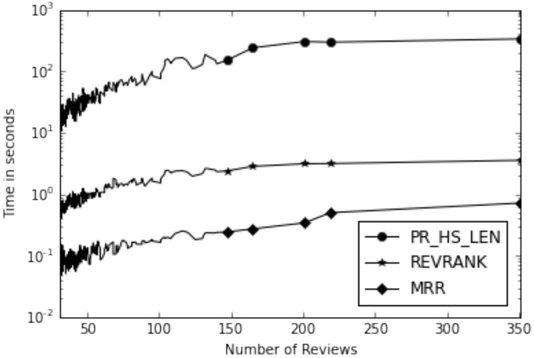
Parameter Sensibility: α and β



- α in all settings had a low influence (4%)
- β produced the highest variation (17%).
- Nevertheless when $0.8 \leq \beta \leq 0.9$, the MRR varying only 6% .

Run-time Assessment

Time required for producing a ranking for 383 products (log scale)



- MRR presents a significantly lower running time

Final Remarks

Contributions:

- Unsupervised method: does not depend on an annotated training set;
- Faster than other graph-centrality methods;
- It performs well in different domains (e.g. closed vs. open-ended);
- Significantly superior to the unsupervised baselines, and comparable to a supervised approach in a specific setting.

Further Work

Next steps:

- Others clustering techniques for graph;
- Methods to select the most relevant reviews;
- Segmented Bushy Path widely explored in text summarization;

Thanks

Thank You! Question?

source: https://github.com/vwoloszyn/MRR

contact: henrique.santos.003@acad.pucrs.br