LexRank: Graph-based Lexical Centrality as Salience in Text Summarization

https://arxiv.org/pdf/1109.2128

November 14, 2023

Abstract

We consider LexRank for computing sentence importance based on the concept of eigenvector centrality in a graph representation of sentences. The results show that degree-based methods (include LexRank) outperform both centroid-based methods and other systems participating in DUC in most of the cases. Furthermore, the LexRank with threshold method outperforms the other degree-based techniques including continuous LexRank.

1 Introduction

Early research on extractive summarization is based on simple heuristic features of the sentences such as their position in the text, the overall frequency of the words they contains, or some key phrases indicating the importance of the sentences. A commonly used measure to assess the importance of the words in a sentence is the *inverse document frequency*, or idf, which is defined by the formula:

$$idf_i = \log(\frac{N}{n_i}) \tag{1}$$

where N is the total number of the document in a collection, and n_i is the number of documents in which word i occurs.

Our summarization approach is to assess the *centrality* of each sentence in a cluster and extract the most important ones to include in the summary. We investigate different ways of defining the lexical centrality principle in multi-document summarization, which measures centrality in terms of lexical properties of the sentences.

In Section 2, we present centroid-based summarization, a well-known method for judging sentence centrality. Then we introduce three new measures for centrality, Degree, Lexrank with threshold, and continuous LexRank, inspired from the "prestige" concept in social networks.

2 Sentence Centrality and Centroid-based Summarization

Centrality of a sentence is often defined in terms of the centrality of the words that it contains. A common way of assessing word centrality is to look at the centroid of the document cluster in a vector space. The centroid of a cluster is a pseudo-document which consists of words that have if × idf scores above a predefined threshold. Centroid-based summarization has given promising results in the past, and it has resulted in the first web-based multi-document summarization system.

3 Centrality-based Sentence Salience

A cluster of documents can be viewed as a network of sentences that are related to each other. We hypothesize that the sentences that are similar to many of the other sentences in a cluster are more central (or *salient*) to the topic. There are two points to clarify in this definition of centrality. First is how to define similarity between two sentences. Second is how to compute the overall centrality of a sentence given its similarity to other sentences.

To define similarity, we use the bag-of-words model to represent each sentence as an N-dimensional vector, where N is the number of all possible words in the target language. For each word that occurs in a sentence, the value of the corresponding dimension in the vector representation of the sentence is the number of occurrences of the word in the sentence times the idf of the word. The similarity between two sentences is then defined by the cosine between two corresponding vectors:

$$idf-modified-cosine(x,y) = \frac{\sum_{w \in x,y} tf_{w,x} tf_{w,y} (idf_w)^2}{\sqrt{\sum_{x_i \in x} (tf_{x_i,x} idf_{x_i})} \times \sqrt{\sum_{y_i \in y} (tf_{y_i} idf_{y_i})^2}} \qquad (2)$$

where $tf_{w,s}$ is the number of occurrences of the word w in the sentence s.

In the following sections, we discuss several ways of computing sentence centrality using the cosine similarity matrix and the corresponding graph representation.

3.1 Degree Centrality

Since we are interested in *significant* similarities, we can eliminate some low values in this matrix by defining a threshold. We define *degree centrality* of a sentence as the degree of the corresponding node in the similarity graph.

ID	Degree (0.1)	Degree (0.2)	Degree (0.3)
d1s1	5	4	2
d2s1	7	4	2
d2s2	2	1	1
d2s3	6	3	1
d3s1	5	2	1
d3s2	7	5	1
d3s3	2	2	1
d4s1	9	6	1
d5s1	5	4	2
d5s2	6	4	1
d5s3	5	2	2

Figure 1: Degree centrality scores. Sentences d4s1 is the most central sentence for thresholds 0.1 and 0.2.

3.2 Eigenvector Centrality and LexRank

Consider a social network of people that are connected to each other with the friendship relation. The prestige of a person does not only depend on how many friends he has, but also depends on *who* his friends are. Similarly, consider a noisy cluster where all the documents are related to each other, but only one of them is about a somewhat different topic Degree centrality may have a negative effect in the quality of the summaries in this case where several unwanted sentences vote for each other and raise their centrality.

This situation can be avoided by considering where the votes come from and taking the centrality of the voting notes into account in weighting each vote. A straightforward way of formulating this idea is to consider every node having a centrality value and distributing this centrality to its neighbors. This formulation can be expressed by the equation:

$$p(u) = \sum_{v \in adj[u]} \frac{p(v)}{deg(v)} \tag{3}$$

where $p(u)$ is the centrality of node u, $adj[u]$ is the set of nodes that are adjacent to u, and $deg(v)$ is the degree of the node v. Equivalently, we can write Equation 3 in the matrix notation as:

$$p = B^T p \tag{4}$$

or

$$p^T B = p^T \tag{5}$$

where the matrix B is obtained from the adjacency matrix of the similarity graph by diving each element by the corresponding row sum:

$$B(i,j) = \frac{A(i,j)}{\sum_k A(i,k)} \tag{6}$$

Note that a row sum is equal to the degree of the corresponding node. Since every sentence is similar at least to itself, all row sums are nonzero. Equation 5 states that p^T is the left eigenvector of the matrix B with the corresponding eigenvector 1.

```
MInputAn array S of n sentences, cosine threshold t  output: An array L of LexRank scores
Array CosineMatrix[n][n];
Array Degree[n];
Array L[n];
for i ← 1 to n do
    for j ← 1 to n do
        CosineMatrix[i][j] = idf-modified-cosine(S[i],S[j]);
        if CosineMatrix[i][j] > t then
            CosineMatrix[i][j] = 1;
            Degree[i] + +;
        end
        else
            CosineMatrix[i][j] = 0;
        end
    end
end
for i ← 1 to n do
    for j ← 1 to n do
        CosineMatrix[i][j] = CosineMatrix[i][j]/Degree[i];
    end
end
L = PowerMethod(CosineMatrix,n,ε);
return L;
```

Figure 2: Computing LexRank scores

```
input  : A stochastic, irreducible and aperiodic matrix M
input  : matrix size N, error tolerance ε
output: eigenvector p
p_0 = (1/N) 1;
t=0;
repeat
    t=t+1;
    p_t = M^T p_{t−1};
    δ = ||p_t − p_{t−1}||;
until δ < ε;
return p_t;
```

Figure 3: Power Method for computing the stationary distribution of a Markov chain

We call this new measure of sentence similarity *lexical PageRank*, or *LexRank*. Algorithm 3 summarizes how to compute LexRank scores for a given set of sentences.

3.3 Continuous LexRank

Discretization by thresholding causes information loss. One improvement over LexRank can be obtained by making use of the *strength* of the similarity links. If we use the cosine values directly to construct the similarity graph, we usually have a much denser but weighted graph. We can normalize the row sums of the corresponding transition matrix so that we have a stochastic matrix. The resultant equation is a modified version of LexRank for weighted graphs:

$$p(u) = \frac{d}{N} + (1-d) \sum_{v \in adj[u]} \frac{idf-modified-cosine(u,v)}{\sum_{z \in adj[v]} idf-modified-cosine(z,v)} p(v) \tag{7}$$

3.4 Centrality vs. Centroid

Graph-based centrality has several advantages over Centroid. First of all, it accounts for information subsumption among sentences. If the information content of a sentence subsumes another sentence in a cluster, it is naturally preferred to include the one that contains more information in the summary. Another advantage of our proposed approach is that it prevents unnaturally high idf scores from boosting up the score of a sentence that is unrelated to the topic. Although the frequency of the words are taken into account while computing the Centroid score, a sentence that contains many rare words with high idf values may get a high Centroid score even if the words do not occur elsewhere in the cluster.