Integration of a Bayesian Learning Model in a Multi-Document Summarization System

Maher Jaoua
Lamia H. Belguith
Fatma Kallel
Abdelmajid Ben Hamadou
INTEGRATION OF A BAYESIAN LEARNING MODEL IN A MULTI-DOCUMENT SUMMARIZATION SYSTEM

Maher Jaoua*, Lamia Hadrich Belguith*, Fatma Kallel Jaoua**, Adelmaïd Ben Hamadou***

* ANLP Research Group, MRACL Laboratory, Faculty of Economic Sciences and Management
University of Sfax, Tunisia
maher.jaoua@fsegs.rnu.tn, lbelguith@fsegs.rnu.tn.

** High School of Commerce of Sfax
University of Sfax, Tunisia
Fatma_fseg@yahoo.fr

*** ANLP Research Group, MRACL Laboratory, High Institute of Computer Sciences and Multimedia
University of Sfax, Tunisia
abdelmajid.benhamadou@istmsf.rnu.tn

Abstract:

We propose in this paper, a new method for multi-document summarization which is based on a learning model. The learning is particularly used in order to deduce, for a given summarization task, the best combination of criteria allowing to select the best summary (or extract). Thus, its use aims at the realization of a flexible multi-document summarization system which can be applied for various summarization tasks. We chose the application of the naive of Bayes learning algorithm in order to determine a set of extracts that are considered the best (generally there is more than one). So we accompanied this algorithm by a multi-objective classification method allowing to find the best extract. The experimentation of the learning gave encouraging results. Also, the evaluation of the ExtraNews system, implementing our method, gave interesting results.

Keywords: Multi-Document Summarization, Abstract, Learning, Naive Bayesian Model, Multi-Objective Classification, Classification Criteria.

1. INTRODUCTION

The proliferation of the electronic documents has incited researches in the field of automatic summarization. Particular interest has, subsequently, been granted to multi-document summarization seeing that we perceive the multiplicity of documents related to the same topics such as news and blogs. A survey of different methods proposed in this field can be found in [1],[2].

It is in this context that our work is subscribed. Indeed, we propose, in this paper a new method of applying learning technique to multi-document summarization. Our work is based on a novel vision which considers the extract as a minimal unit for evaluation and classification [3]. So, to determine the best extract, we propose to generate and classify a set of intermediate extracts produced by combining sentences of original documents. This method has obtained good results during the participation in the international evaluating conference DUC (Document understanding conference) and TAC (Text Analysis conference) despite time consuming [4]. But the essential problem of this method, and of the majority of automatic multi-document summarization systems, is related to the classification process. Thus, the major challenge is to determine which criteria can be used and how to aggregate them to evaluate intermediate extracts.

In order to give a solution to this challenge, we propose to apply a learning technique to deduce the importance of criteria for a given corpus and task of summarization1.

In what follows, we present a brief overview of the related works in automatic multi-document summarization. More precisely, we mention the main learning methods experimented in this field. Then, we detail the motivations and challenges of our work. We describe, after, the learning model that we adopted, the main criteria used by our method for the extracts classification, as well as the various corpora that we used and the obtained results. Then, we present the multi-objective classification method that we proposed

---

1 Example of Summarization Tasks: Summarize Blogs, Produce Very Chort Summaries, Produce Query Base Summaries.
in order to determine the best extract. This later is selected between the best extracts classified using the learning model. Finally, we present the evaluation results of the proposed method.

2. PREVIOUS WORK

The literature review allows us to discern three main approaches for automatic multi-document summarization:

- **Symbolic approach:** It is based on linguistic knowledge and analysis in order to select the best sentences and to avoid the selection of redundant units when producing the summary. This knowledge is detected using co-references and lexical information metrics [5], exploitation of discourse structure by using graph structures to determine the importance of sentences by the identification of lexical relations between sentences [6], determining named entities and relation between them [7], extracting the cohesion forms connecting the sentences of the source texts in order to determine the relevant sentences, ...

- **Statistical approach:** It is based on the attribution of weights to the sentences in order to determine the most important ones which form the final summary. Different statistical and probabilistic techniques have been proposed in this context. Indeed, to detect the different sentences of a given collection of documents, different techniques have been experimented: combining some frequency features like words frequency, \(tf*idf\) [5], using MMR metric (maximal marginal relevance) [8], IGR metric (informational Gain Ratio) [9], the Latent Dirichlet Allocation [10]. Other methods use machine learning techniques like SVM (Support Vector Machine) [11], and Bayesian statistical model [12] ...

We note that, several works, based on the statistical approach, has tried out the learning techniques to induce the heuristics allowing the selection of the relevant textual units from one or a set of original documents. The combination of these heuristics is, generally, carried out manually [13] or by adopting generic methods of criteria aggregation based on machine learning [14]. Several learning techniques can be used. The works of Kraaij are based on the naive bayesian model [15]; those of Amini use regression models [16] and those of Qazvinian are based on the genetic algorithms [17]. In our case, we chose the use of the naive bayesian model. The choice of this model is justified by its facility of implementation, its performances in many applications and its ability to process inadequate examples.

3. CHALLENGES

The method that we proposed in the context of multi-document summarization is distinguished by its new vision which considers the extract (rather than the sentences or terms or phrases) as a unit of importance and classification. Indeed, the extract process of our method is based on the generation of a multitude of intermediary extracts. These extracts will be classified in order to select the best one (Figure 1).

![Figure 1: Principle of the Extraction Process.](image)

Thus, the purpose of our work is not concerned with the classification of sentences in order to evaluate their membership to the extract. Indeed, it is rather concerned with the classification of the extracts, while allotting them an importance level [18]. This extract classification is based on different criteria. But the problematic concerns the manner of aggregating these criteria.

During the experimentation of various criteria of extracts classification, we noted that some of them can be adapted to various contexts and for various summarization tasks. It is the case, for example, of the words coverage criterion of the extract. This criterion can differently relate to key-words, to structure words or to user question words according to whether it is a question of summarizing news or structured documents, or producing summaries based on user question.

These observations put the emphasis on the multitude of the possible criteria for the evaluation and the classification of a multi-document summary. This multitude poses the problem of the determination of the importance of one criterion compared to another on the level of the task of summaries classification.

In addition, the determination of this importance is a hard task. Indeed, and since human writing is handled, it is impossible to find invariants concerning the use of a uniform style for all the writings.

Moreover, the determination of the importance of one criterion compared to another does not obey explicit rules. A way of solving this problem is to attribute weights, in an empirical way, to the various criteria which are linearly combined. These weights can depend on the type of the original documents (e.g. scientific texts, blogs and biographies) and on the type of summarization task (e.g. summarization guided by user question and production of a very short summary).

The recourse to the application of a learning technique, therefore, is justified. Its application has, so,
as objective to determine the best combinations of criteria identifying the best summaries being given varied document collections as input and different summarization tasks. The recourse to this type of techniques aims, thus, to offer to our summarization method a better operational flexibility.

4. APPLICATION OF LEARNING TO THE EXTRACT’S CLASSIFICATION

The learning model which we chose to deduce the importance of extract’s classification criteria is that of naïve Bayes [19]. We notify that this model is characterized by its simplicity of implementation, by its relevance and by its taking into account of inadequate examples. The application of this model in the process of classification and selection of the best extracts recommends two stages: the first one consists in discretizing the used criteria; in the second stage, one must determine the probabilities of appearance of these criteria in the best extracts and this while using an important set of extracts. Let us note that this model is confronted with two constraints: the first is that of determination of an important set of sets used in the learning and test steps; whereas the second is articulated around the definition of automatic metrics allowing the determination of the best extracts.

Let’s note that our method recommends a generation process which produces a multitude of intermediate extracts to be classified in order to select their best. Indeed, the generation of several intermediate extracts (partitions), starting from the original texts, can offers a significant number of examples. This process allows us to overcome the first learning constraint related to the presence of an important set of examples. Indeed, it allows to deduce the elements contributing to the classifying of new examples. This constraint proved difficult to satisfy for the automatic summarization field. This difficulty is due to the fact that, on the one hand, there is not a concept of an ideal summary or extract and on the other hand, the corpora are limited in size since they generally derive from some international evaluation conferences or from isolated efforts of certain works in the field. Moreover, these conferences use human summaries as model summaries (whose production requires an important time).

So, a solution for these learning constraints is partially given by the vision which we proposed for the extraction process [20]. In addition, the automatic metrics of the summaries evaluation applied by the evaluation conferences supplement the solution suggestions to these constraints. Indeed, the use of these metrics can help to overcome the second learning constraint related to the definition of automatic metrics measuring the importance of the extracts. These metrics allow, thus, to automatically evaluating an extract or a summary and this by comparing it with several model summaries. Among these metrics, we chose the Rouge metric employed by the conference DUC for the summaries evaluation [21]. The scores determined by the Rouge system are based on two types of units:

- Model Units (MU): representing the n_grams of the words extracted starting from the human summaries used as model summaries;
- Peer Units (PU): which result from the decomposition of the summaries or the extracts generated by the systems in n_grams of words.

The Rouge_{n} metric uses the correspondence between the distribution of the words (n_grams) of a candidate summary (PU) and that of a whole of human summaries (MU). The computation formula of Rouge metric is as follows:

\[
\text{Rouge}_{n} = \frac{\sum_{C_{n}(\text{Model})} \sum_{n\_grams} \text{match(candiat,c)}}{\sum_{C_{n}(\text{Model})} \sum_{n\_grams}}
\]

(1)

Where match (candidate, c) represents the number of common n_grams between the system summary and the model summaries. The denominator of the equation represents the sum of the number of n_grams in model summaries.

It should be noted that Rouge_{n} is the general formula of the Rouge metric. One can thus obtain measurements of Rouge_{1} (1_gram), Rouge_{2} (2_grams), etc. Recent studies showed that Rouge_{2} measurement presents the best correlation with the human judgments [21]. So, we adopted this measure to quantify the importance of an extract.

The set of examples which we built is, therefore, made up of: a set of partitions (intermediary extracts) obtained starting from the original texts, on the one hand, and of Rouge_{2} values on the other hand. Let us note that intermediary extracts are generated while combining sentences of sources documents [4]. The Rouge_{2} values are calculated through the correspondence between the partitions of the same documentation and the model summaries of the collection in question. We also determine, for each partition, the values associated with the criteria in question. The main criteria which we adopted for the extract’s evaluation and classification process are the followings:

1. Length (SSL): Sum of the extract Sentences Lengths.
2. Pertinence (ATI): Average of the tf*idf values.
3. Coverage: According to the summarization task (summarization of multiple documents, of multiple blogs, oriented by user question,…), we can use:
   - CEK: Coverage of the Extract in Key words.
   - CWQ: Coverage of the extract in Words resulting from the user question.
   - C2W: Coverage of the extract in double key Words.
   - C3W: Coverage of the extract in triple key Words.
   - C4W: Coverage of the extract in quadruple key Words.
• COW: Coverage of the extract in key Words indicating the opinion.

4. Position criteria: According to the summarization task, we can use:
   - PSE: Average of the Positions of the extract Sentences in their original texts.
   - PQW: Average of the Positions of the Question Words in their original texts.
   - P1W: Average of the positions of the key words in their original texts.
   - P2W: Average of the positions of the double key words in their original texts.
   - P3W: Average of the positions of the triple key words in their original texts.
   - P4W: Average of the positions of the quadruple key words in their original texts.


6. Redundancy (RE): Informs about the redundancies of the extract sentences.

In order to be able to exploit the built set of examples, in a learning model, it is possible to discretize the values of each criterion. For that, we associated for each criterion a value on a scale varying from A to E. The Rouge$_2$ values obtained for the example set are as replaced by the values of this scale in the same way as the ones allotted to the various criteria. Thus, for the Rouge$_2$ criterion, we distinguish five classes going from extracts informatively weak to good extract (having good quality) and passing by the moderately weak, middle, and middling good extracts. This discretization depends, firstly, on the source collection, and secondly on the interval of density of each criterion.

For example, for the same collection, if the criterion associated with the Rouge$_2$ metric varies within the interval [0-0.15] and if one has a uniform density, then the class E has as an interval $[0.06-0.09]$ and classes B and A the respective intervals $[0.09-0.12]$ and $[0.12-0.15]$.

We mention that this theorem associates the posterior probability of an assumption H to three other probabilities according to the following formula [19]:

$$P(h \mid D) = P(D \mid h) \ast P(h) \ast P(D)$$

(2)

With $P(H) =$ probability that the assumption H is checked independently of the data D (this term is also called prior probability);

$P(D \mid H) =$ probability of observing the data D knowing that the assumption H is checked (this term is also called posterior probability);

$P(D) =$ probability of observing the data D independently of the assumption H (this term is also called obviousness).

Let us note that the application of this theorem, supposes beforehand, that the criteria were discretized and independent of each other. The naive bayesian model allows to determine, for a new example (extract), its probability of membership to the one of the classes defined by discretization of the Rouge$_2$ metric.

Thus, the conditional probability so that the extract is classified in a class C of the Rouge index, for example, is given by the following formula:

$$P(C \mid A_1, A_2, …, A_n) = P(A_1, A_2, …, A_n \mid C) \ast P(C) \ast P(A_1, A_2, …, A_n)$$

(3)

If we suppose that there is independence between the criteria Ai, we can thus determine:

$$P(A_1, A_2, …, A_n \mid C) = P(A_1 \mid C) \ast P(A_2 \mid C) \ast … \ast P(A_n \mid C)$$

Thus, the new examples are classified in the class C if the product $P(C) \ast \Pi P(A_i/C)$ (for each i=1…n) is maximum, knowing that n indicates the number of criteria whereas C corresponds to the various classes allotted to the examples.

In the following section, we present the bases of examples and test which we tried out.

5. TEST CORPUS

As we have mentioned above, the difficulty in applying a learning model in an automatic multi-document summarization method is articulated around the definition of a learning and test corpus. Indeed, current works are generally focused on the sentences. They use rather a classification of each sentence to decide on its membership to the final extract [14], [22]. However, this idea collides, at the time of the learning step, with the difficulty in deciding, in a strict way, if a sentence belongs or not to the final extract. In order to remedy this problem, certain works use a degree of a sentence membership to the final extract and this while being based on commentators judgments. This solution poses the problem of the differences in annotation as well as the setting in scale of the notes allotted by the various commentators. This problem could be solved if one had an evaluation allowing to indicate the probability of membership of a source sentence to the final extract. But, the absence of this evaluation puts question marks on all the results obtained in this direction.

In the case of use of an approach centered around the extract, the problem of decision of the membership or not of a sentence to the summary is not posed any more. In addition, the use of metric of correlation between the extract and the human summaries (i.e. Rouge$_2$ metric), makes it possible to solve the problem of the extract’s importance evaluation.

Conferences DUC and TAC, facilitated to us the collection of a great corpus of examples. Indeed, the presence of the author's abstracts for the documentation provided on the one hand, and the “unlimited” number of partitions on the other hand, made it possible to generate the following corpora:

• A corpus resulting from DUC’04 treating multiple documents describing human biographies (C1 corpus);
• A corpus resulting from DUC'04 treating multiple documents which are manually and automatically translated from the Arabic to English language (C2 corpus);
• A corpus resulting from DUC'06 treating multiple documents provided with users questions (C3 corpus);
• A corpus of DUC'07 treating documents presenting the evolutionary events (C4 corpus);
• A corpus resulting from TAC'08 treating opinions in the multiple blogs (C5 corpus).

However, and in order to guarantee obtaining good extracts (of class A), certain genetic mechanisms of combination (archive) are introduced to ensure the recovery of the best extract during generations [18].

Let us note that we kept in these corpora the inadequate examples. These examples (extracts) have the same configurations of criteria but there are classified into two different classes. The presence of these inadequate examples is mainly due to the discretization carried out and to the superfluous borders between the classes which it generates.

6. MULTI-OBJECTIVE CLASSIFICATION OF EXTRACTS

The application of the learning algorithm makes it possible to determine the probable class of each extract without, however, determining the best extract (the final extract). To determine the final extract, we must select among the set of extracts of the best class (class A), the extract which maximizes the evaluation criteria. However, this operation cannot guarantee the presence of an extract which obtains the best values for all these criteria.

For this reason, we sought to determine a classification of the criteria with the aim of obtaining an importance order and this, to compare the extracts of the same class. Since this classification operates on several criteria, we use a multi-objective classification. This classification is based on the concept of dominance which compares each criterion of the solutions [23]. A solution (extract) dominates a second solution if the values obtained for each criterion are better, or if there is at least a criterion in the first solution, whose value is better, knowing that the remainders of the values of the other criteria are equal. Formally, we consider that a solution X dominates a solution Y and we note:
\[ \text{Dom}(X, Y) \] if for all criteria Ci, we have:
\[ Ci(X) > Ci(Y) \text{ or } \exists j \text{ such that } Ci(X) = Ci(Y) \quad (10) \]

Ci (X) design the value allotted to the solution (extract) X for the criterion Ci.

The possible presence of several dominant solutions resulting from class A is due to the use of several criteria. Thus, and to carry out a classification of the dominant extracts, we carried out a study on the various corpora used at the time of the learning step. The purpose of this study is to determine, for the dominant extracts of the class A, the criteria which allow to obtain the best Rouge2 score and thus, the best correspondence with the model summaries.

We compare, thus, for each collection of a corpus Cr, the dominating extract having obtained the best Rouge2 score with the remainder of the dominant extracts of class A. This comparison enabled us to deduce an importance percentage as well as a classification of each criterion compared to the others criteria for all the corpus collections. Table 1 shows the results obtained by the comparison of the dominant extracts of class A and this, for the corpora studied at the time of the learning step.

Thus, being given a document collection Cd, the algorithm of production of the best extract Ef (final extract) is as follows:

**Algorithm ProdFinalExtract**

```
Begin
    i = 1, Ef=Ø
Repeat
    Archive=Ef
    To generate the set of intermediate extracts Edi= {Edi1, Edi2 ... EdiN} starting from Cd.
    To classify each extract Edi in its probable class (A or B or C or D or E) by using the classifier of Bayes.
    To seek among the extracts having for probable class A, the set of the dominant extracts Edi= {Edi1 ... EdiM}.
    Ef= Ef U EDI
    i = i+1
Until (File =Ef)
    If card (Ef)=1 Then
        Ef is the final extract
    Else
        Classify Ef according to the rank of priorities allotted to the multi-objective criteria of extracts classification
        Ef is the final extract
End if
End Algorithm
```

<table>
<thead>
<tr>
<th>Corpus</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>SSS</td>
<td>87%</td>
<td>92%</td>
<td>95%</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Pertinence</td>
<td>ATT</td>
<td>33%</td>
<td>36%</td>
<td>25%</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>CEK</td>
<td>72%</td>
<td>65%</td>
<td>57%</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CWQ</td>
<td>13%</td>
<td>16%</td>
<td>27%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(6)</td>
<td>(4)</td>
<td>(7)</td>
<td>(8)</td>
</tr>
<tr>
<td></td>
<td>C2W</td>
<td>4% (8)</td>
<td>3% (7)</td>
<td>8% (7)</td>
<td>6% (8)</td>
</tr>
<tr>
<td></td>
<td>(10)</td>
<td>(12)</td>
<td>(13)</td>
<td>(10)</td>
<td>(12)</td>
</tr>
</tbody>
</table>
The architecture of the ExtraNews system, which implements our method, is composed of the following seven modules (Figure 2):

1. **The Preliminary Module**: It allows extracting textual material from a HTML or XML file by removing tags. It also allows splitting the text into sentences and words.

2. **The Statistical Module**: It allows the calculation of the frequency of the non-empty words as well as the sort of words according to these frequencies. This module takes into account the derived forms in this calculation. The key-words coming from this module allows some criteria to participate in the evaluation of the extracts such as the coverage and the pertinence.

3. **The Linguistic Module**: It allows to correct the words frequencies while detecting the synonym and meronym words and thus by the use of WordNet2.

4. **The Filtering Module**: It allows reducing the sentences with the use of the inter-sentences dominance. This module has as a task to reduce the initial number of the handled sentences while keeping only those non-dominated.

5. **The Compression Module**: It allows reducing the textual material of the non-dominated sentences. This reduction is based on the elimination of some constituents which do not contain key-words.

6. **The Generation and Classification Module**: It allows generating a multitude of extracts which will be classified according to pre-determined statistical criteria. This classification is based on the learning technique and on the multi-objective classification detailed in this paper.

7. **The Re-Ordering Module**: It allows correcting the order between the sentences of the final extract so as to reestablish its coherence and cohesion.

**8. EVALUATION**

In order to evaluate the ExtraNews system implementing the proposed method, we have evaluated our participation in the DUC’04-TAC’08 conferences. We used the rest of the official corpus which was not used in the learning step. The following table presents an overview of our result before and after adding the learning step.

### Table 1: Classification of the Criteria of the Extracts Evaluation

<table>
<thead>
<tr>
<th>Module</th>
<th>CE</th>
<th>RSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>COW</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td>Abs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSD</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td>(3)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>PQW</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>PDS</td>
<td>17%</td>
<td>15%</td>
</tr>
<tr>
<td>(5)</td>
<td>(5)</td>
<td></td>
</tr>
<tr>
<td>PZW</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>(7)</td>
<td>(7)</td>
<td></td>
</tr>
<tr>
<td>PZW</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td></td>
<td>(12)</td>
<td></td>
</tr>
<tr>
<td>PZW</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td></td>
<td>(12)</td>
<td></td>
</tr>
<tr>
<td>CE</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSE</td>
<td>Abs</td>
<td>Abs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2 WorldNet: An Electronic Lexical Database Available for Different Languages.
As we have shown in this evaluation, we can conclude that the integration of the learning step has improved the results of our system and the Rouge2 value (with learning) was increased by 0.03-0.05 (which represents an amelioration of 35%-70% for the initial value of Rouge2: without learning).

9. CONCLUSION

In this paper, we started by presenting the principal basic elements of our method which considers the process of extraction as an optimization problem. It is then a matter of selecting the best extracts, among multiple intermediary extract, which maximizes a set of classification criteria. Let us note that the criteria of classification are defined on the extract (granularity of comparison) as a whole entity.

In the second part of this paper, we detailed the criteria adapted to our vision and allowing to evaluate the quality of the extracts. We exploited these criteria in a learning model where it is a question of using a naive bayesian classifier in order to quantify the importance of the criteria of the extracts classification.

The choice of the final extract is obtained following a multi-objective classification. This classification allows to determine, in a first stage the dominant extracts, then to select the best one.

We also focused the numerical aspect which characterized the extraction process of our method. This method also distinguishes the development of symbolic and numerical features. These features were integrated to improve the quality of the extraction. In particular, our method recommends a step of filtering the sentences of the documents sources in order to reduce their number. Filtering is based on the concept of dominance between sentences. Our method is also based on a step of compression of the filtered sentences in order to eliminate the useless textual material. This compression is based on the use of a set of compression rules and of a parser. We also considered necessary to integrate a step of reordering of the extracts sentences, whose goal is to correct the order of the extract sentences.

References:


Table 2: Evaluation Results of the Extra News System

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Size</th>
<th>Without Learning</th>
<th>Rank</th>
<th>With Learning</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (DUC’04)</td>
<td>30</td>
<td>0.076</td>
<td>4/14</td>
<td>0.115</td>
<td>1/14</td>
</tr>
<tr>
<td>C2 (DUC’05)</td>
<td>5</td>
<td>0.090</td>
<td>1/11</td>
<td>0.121</td>
<td>1/11</td>
</tr>
<tr>
<td>C3 (DUC’06)</td>
<td>20</td>
<td>0.068</td>
<td>22/36</td>
<td>0.118</td>
<td>3/36</td>
</tr>
<tr>
<td>C4 (DUC’07)</td>
<td>7</td>
<td>0.070</td>
<td>21/32</td>
<td>0.120</td>
<td>3/32</td>
</tr>
<tr>
<td>C5 (TAC’08)</td>
<td>10</td>
<td>0.090</td>
<td>7/19</td>
<td>0.123</td>
<td>2/19</td>
</tr>
</tbody>
</table>


[19] B. Lavoie, Apprentissage Bayésien, Synthèse de Lectures, Séminaire Spécial sur L’Apprentissage Automatique, Université du Québec à Montréal, Un-Published.


