Algorithms for intelligent automated evaluation of relevance of search queries results

Anna Tikhomirova [1] and Elena Matrosova[1]
[1]*National Research Nuclear University MEPhI*
(Moscow Engineering Physics Institute)
anna7909966@ya.ru, matrosova_ev@inbox.ru

Abstract
This paper is devoted to the problem of automated evaluation of relevance of search queries results. High relevance of search algorithm output is the base of effective large quantities of data processing, which is worked at by users of modern informational systems. Automated and reliable estimate of relevance of search queries results will give the opportunity to lower time expenditures for the best algorithm choice. The usage of improved from this perspective algorithms will allow to raise effectiveness and user satisfaction when dealing with automatic search systems in any activities.

Keywords: neural network, semantic analysis, algorithm, search query, teaching model, machine learning.

1 Introduction

The successful solution of current scientific and practical problems cannot be achieved without relevant and complete information about the state of the problem, the latest methods in the subject of research and trends in science development. Nowadays one cannot imagine a complete scientific work that didn’t use materials publicly available on the Internet or materials from private information systems. A quick and precise response to a search query is a crucial part of effective user operation and a competitive advantage of search engines.

The basis of the successful operation of a search engine is an effective algorithm of providing a response to a search query according to user expectations. Companies producing search engines and services are constantly carrying out research to improve algorithms in use [1]. The term 'neural network' was coined in the middle of 20th century. Nowadays neural networks are the main instrument of developers of intelligent search engines. Despite many capabilities of this instrument and its ability to self-learn in the process, however, all new or updated algorithms need to be tested before their practical application. It is done to ensure that users get the engine with an algorithm that works correctly and efficiently. This makes the problem of algorithm preliminary quality evaluation imperative for developers. Such quality evaluation is done by analyzing relevance of search query results.

2 Algorithm of forming and evaluating results of a search output

The workings of search engines that give access to information based on queries become more and more complicated. An enormous amount of information that can be found with one query demands intelligent ranking based not only on the presence of keywords but also on semantic correspondence of the material. Users grow more and more demanding to search results and if they feel that a search engine is not effective they will likely choose another search engine that gives them more relevant results.

Technology is constantly improving and researchers aim to use artificial intelligence to solve complicated multivariate problems [2]. For this problem it means using machine learning of search algorithms. In the context of this article machine learning is a system self-learning to find more and more relevant output results based on positive and negative examples. Self-learning means that a machine improves its work quality without any human involvement.

Result relevance is a characteristic of a degree of material (ranked and outputted by a search engine) correspondence with user expectations. In the context of the given problem it is semantic correspondence of a search query to an outputted document [4]. Since the problem is that results should correspond with human expectations, its solution can be successfully achieved by artificial intelligence systems that are based on self-learning machines optimizing search algorithms.

Considering this, it is suggested to look into a possibility of using neural networks to solve this problem. According to the suggested approach, to rate the quality of a search algorithm work one can apply the following diagram [5].

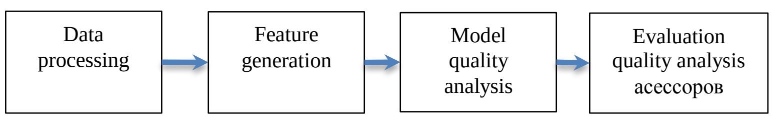

Figure 1: Diagram of the quality evaluation of a search algorithm

Data processing

This stage is highly important. User input can contain various mistakes and the system should transform any input into easy to analyze format and give relevant results. So, an effective search algorithm must be able to work with inputs of various qualities and successfully execute procedures such as spell-check, correction of lexical mistakes, unit conversion, lemmatization and stemming [6]. Development of such an algorithm requires not only machine learning specialists but also competent linguists [7].

Feature generation

Feature generation is a process of formalizing input data and identifying certain signs that can become input for a teaching model. In this stage should be realized the following procedures: identification and removal of noise words, generation of a ‘bag’ of words, use of TF-IDF. Work with the ‘bag’ of words means encoding all the words from a sampling and forming of a unified vocabulary. At the same time a space of dichotomic or serial variables is created, with one dimension for every time a word with a given index is present in the processed text (1).

$$D = D_1 \cup D_2 \cup \dots \cup D_n, \tag{1}$$

where

D_i is a set of words in an object i,
n is a number of objects.

The dimensionality of a feature space becomes equal to the number of unique not-noise words in the whole sampling, and the matrix of features becomes weak. One needs to take into account the word weight. Static measure TF-IDF can be used for that purpose.

After processing the data one will get a huge feature matrix. To work with it further one can use latent semantic analysis. It allows picking key features based on the revealed correlation between materials and words.

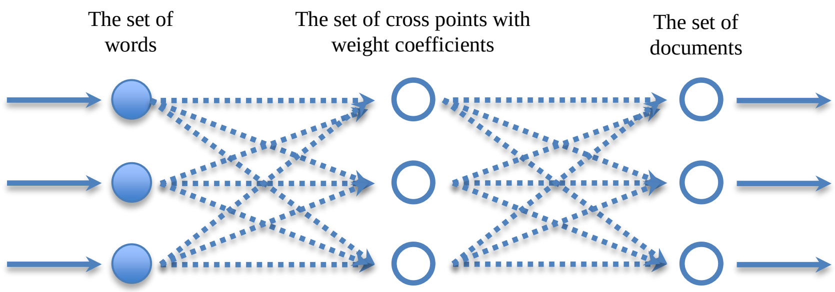

Figure 2: The representation of the latent semantic analysis

With machine learning, the space of features is expanded in the process of working with a natural language to get various heuristic statistics since they can contain hidden important information [8]. These features include the length of the text and the ration of the query length to the headline length.

It is necessary to make the procedure of normalization and centering of the features that prevents the network difficulty and computational effort from going up.

Model quality analysis

Algorithm quality evaluations, or quality metrics, can be calculated in different ways. One of the more frequently used methods is a loss function

$$Q(a, X^l) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(a, x_i), \tag{2}$$

that can determine the quality of an algorithm $a(x)$ on the plurality of objects X^l, and RMSE evaluation (3) [9].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - t)^2}{n}}, \tag{3}$$

where

y_i is a predicted value,
t is a target value.

Sample X^L is divided into two disjoint sub-samples: teaching X^l of length l and control X^k of length (4).

$$k = L - l \tag{4}$$

For every decomposition $n=1,...,N$ an algorithm is created and the work quality is rated based on the control sample $Q_n = Q(a, X^k)$. Then the arithmetic average Q_n for all the decompositions is calculated. It is the evaluation of a sliding control (5).

$$CV(\mu, X^L) = \frac{1}{N}\sum_{n=1}^{N} Q(\mu(X^l), X^k) \quad (5)$$

It is advisable to also trace base lines, i.e. to rate the quality by constant algorithms. These algorithms are not very time-consuming and one can judge the quality of their work by comparing their results with the results of a complicated model.

Assessors' evaluation quality analysis

The quality of an evaluation model can be determined by comparing it with assessor evaluation quality. Assessor is a person who is tasked by developers with evaluating the search engine based on how well the query and user expectations correspond with found materials. Generally, assessors' evaluations are discrete data. The procedure of assessors evaluating the quality of work of a search engine is, at its core, a task of group multicriteria evaluation of the given object, and after its completion one must find one unified consolidated evaluation [10]. Such evaluation can be found by different ways. The simplest one is calculating an arithmetical mean, but this evaluation can be incorrect, since assessors' qualification can differ from person to person. Assessors' different levels of computer literacy, different professional fields and education all directly influence their evaluation. But it must be said that the true authenticity cannot be achieved without their inherent differences. In that case it necessary to take this peculiarity into account, but at the same time to get an average evaluation, that is close to the true value.

To solve this problem one can take into account the weight of each assessor's evaluation and average them based on the analysis of deviation of an individual evaluation from the expectation value of the sample. With this method the evaluation that is the closest to the average mean in the group will get the highest weight.

For n number of evaluation criteria of the output relevance of m assessors there will be $m \times n$ evaluations that can be presented as a matrix (6) [11]:

$$Z = \begin{bmatrix} z_1^1 & \dots & z_n^1 \\ \dots & \dots & \dots \\ z_1^m & \dots & z_n^m \end{bmatrix} \quad (6)$$

where

z_j^i is an evaluation of the criterion j by assessor i.

Then one must calculate the arithmetical mean value for evaluations of each criterion (7):

$$\overline{\overrightarrow{z^{(A)}}} = \frac{1}{m}\sum_{i=1}^{m} \overrightarrow{z^{(i)}} = \frac{1}{m}\left(\sum_{i=1}^{m} \overrightarrow{z_1^{(i)}}, \dots, \sum_{i=1}^{m} \overrightarrow{z_n^{(i)}}\right) = \left(\overline{z_1^{(A)}}, \dots, \overline{z_n^{(A)}}\right) \quad (7)$$

where

$\overline{z_j^{(A)}}$ is an average evaluation of criterion j.

It is suggested to assume the arithmetical mean of the evaluations plurality as the true value. For each assessor we have the evaluation of dispersion of the introduced random variate (8).

$$\sigma^{(i)2} = \frac{1}{n-1}\sum_{j=1}^{n}(z_j^{(i)} - z_j^{(A)})^2 \quad (8)$$

Then we calculate the sum of reciprocal values of deviation dispersions of all assessors (9):

$$\sum_{i=1}^{m}\frac{1}{\sigma^{(i)2}} \quad (9)$$

and the weight coefficient for each (10):

$$w^{(i)} = \frac{1}{\sigma^{(i)2}} / \sum_{i=1}^{m}\frac{1}{\sigma^{(i)2}} \quad (10)$$

After calculation the weights we calculate the refined criteria evaluations as average weighted evaluation with the account of various levels of assessor evaluations' error (11):

$$\overline{z_j^{(B)}} = \sum_{i=1}^{m} w^{(i)}\, z_j^{(i)} \quad (11)$$

The suggested method of assessor evaluations analysis can increase the evaluation relevance authenticity. Since assessors' opinions are taken as a comparison standard (accounting for errors), the standard variant must also be correctly mathematically processed.

3 Conclusions

The problem of replacing manual labour with automated system is always difficult and complicated. Human evaluation of any object (in this article it's the relevance of a search engine output) must be transformed into an algorithm realizable by artificial intelligence. Assessor's evaluation cannot be taken as is, since there is always a factor of varying competencies that makes the evaluation result not uniform enough and demanding corrections.

The successful solution of this problem is in the interests of not only developers of various automated search engines that are nowadays used everywhere from small businesses and network organizations to science information libraries, but also of average users of various services who are interested in reducing the time and effort consumed in processing the search information that doesn't meet their expectations.

4 Acknowledgments

This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Professional Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).

References

[1] Evgeniy Gabrilovich and Shaul Markovitch. Feature generation for text categorization using world knowledge. *In IJCAI*, volume 5, pages 1048–1053, 2005.

[2] Samsonovich A.V. Functional possible biologically inspired by cognitive architectures. *XVII All-russian scientific-technical conference "Neuroinformatics-2015"*. National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, 2015.

[3] Samsonovich, A.V., Klimov, V.V., Rybina, G.V. Biologically Inspired Cognitive Architectures (BICA) for Young Scientists: Proceedings of the First International Early Research Career Enhancement School (*FIERCES 2016*) ISBN: 978-3-319-32553-8 (Print) 978-3-319-32554-5 (Online)

[4] Birger Hjorland. The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology*, 2010.

[5] Skorokhodov I.S., Tikhomirova A.N. Key stages of text processing and feature generation in text classification. *Problems of Modern Science and Education*, 2016, 15(57), p. 18-22.

[6] Vimala Balakrishnan and Ethel Lloyd-Yemoh. Stemming and lemmatization: Acomparison of retrieval performances. *IACSIT*, 2014.

[7] Balandina, A., Chernyshov, A., Klimov, V., Kostkina, A. Usage of language particularities for semantic map construction: Affixes in Russian language. *International Symposium on Neural Networks*. ISNN 2016: Advances in Neural Networks – ISNN 2016 pp 731-738

[8] Klimov, V.V., Chernyshov, A.A., Shchukin, B.A. Composition of web-services based on semantic description . *WEBIST 2015 - 11th International Conference on Web Information Systems and Technologies, Proceedings* .

[9] J.M. Altman, D.G.; Bland. Statistics notes: measurement error. *BMJ*, 1996.

[10] Tikhomirova A.N., Sidorenko E.V. Optimization of the process of scientific and technical expertise projects in nanobiomedical technologies. *Nanotechnics*, № 1 (29), 2012 p.26-28

[11] Kryanev A.V., Tikhomirova A.N., Sidorenko E.V. Group expertise of innovative projects using the Bayesian approach. *Economics and mathematical Methods*, 2013, 49 (2), p.134-139.