A new literature search model for multi data objective programming

Li Kan
Teaching and research support centre, Dalian Naval Academy, Dalian, Liaoning, 116018, China
Corresponding author’s e-mail: likandl@126.com

Abstract: Aiming at the problem that the current literature search algorithms only provide literature clues and original texts, which can not realize the analysis and evaluation, it is the formal comparison and cannot compare the actual contents of new papers and existing literatures; the objective programming target algorithm is introduced to calculate the "copy ratio" of the substantive content by minimizing the deviation between the objective function and the target value; small sample verification is carried out It is feasible and can open up a new idea for the upgrading of literature search tools.

1. Preface
Literature search is a kind of formal inquiry of literature clues and original texts for the needs of specific topics or novelty and duplicate search of periodical agencies. It can not analyze and evaluate the real content of documents. Expert evaluation method can solve the above problems, and evaluate the creativity, advanced nature, novelty and practicality of the evaluated subjects based on their professional knowledge, practical experience, comprehensive analysis ability and professional knowledge. However, the role of experts can not be replaced by general scientific and technological intelligence personnel. Moreover, due to the high labor cost and low efficiency, it is impossible to complete the substantive content search under big data. It is necessary to find a precise search model based on computer technology.

2. Concept definition

2.1 literature
Literature is the most effective means of recording, accumulating, disseminating and inheriting knowledge. It is the most basic and main source of information in human social activities. It is also the knowledge with historical value and research value, and the most basic means of exchanging and disseminating information. People usually understand that literature refers to the sum of knowledge recorded in books, periodicals and ordinances.

2.2 literature search
It refers to the process of obtaining documents according to the needs of study and work. In modern times, literature refers to articles and books with historical value or important books and materials related to a certain discipline. With the development of modern network technology, literature retrieval is more completed by computer technology. Literature search in narrow sense refers to the process of searching and obtaining specific related documents from a large number of organized documents. In a broad sense,
literature search includes two processes: document storage and search. Information storage is to gather a large number of disordered literature information. According to the appearance and content characteristics of information sources, it is processed by sorting, classifying, concentrating and indexing to make it systematized and orderly. According to certain technical requirements, a database or retrieval system with retrieval function is built for people to search and use.

2.3 goal planning
The optimization of more than one objective function in a given region is studied. The quality of a program needs to be compared with multiple objectives, which are sometimes inconsistent or even contradictory. Under a set of constraints, multiple objective functions are maximized (or minimized). Objective programming can be regarded as a special compromise model of multi-objective optimization problems. In multi-objective decision-making problems, it is assumed that the decision-maker designs an objective value for each objective, and its idea is to minimize the deviation (positive deviation, negative deviation or positive deviation) between each objective function and the objective value.

3. Design of literature search database and field search

3.1 structure chart of retrieval database
Computer retrieval system is generally divided into two types: retrieval database (commercial type) and search engine (network type). In terms of information value and integrity of retrieval function, retrieval database is better than network retrieval. The structure diagram is only for the retrieval database, as shown in Figure 1. The main items of each document in the source literature are extracted and combined together to form an item. This item can replace the full text and greatly concentrate the information content of the original text. This process is called description process.

![Figure 1 composition structure of retrieval database](image)

3.2 field search design
Field limited search refers to the search of a limited search term within one or more fields in a database record. Retrieval can be divided into two types: suffix and prefix. The former corresponds to the basic index and reflects the subject content of the literature; the latter corresponds to the auxiliary index, reflecting the external characteristics of the literature. The position operator defines the relative position between words. In the process of retrieval, field operator is sometimes used to limit the range of search fields. Field retrieval is divided into basic field retrieval (see Table 1) and auxiliary field retrieval (see Table 2).

Basic fields
### In field retrieval, suffix can be used to limit the basic field and prefix to restrict the auxiliary field.

### 4. Literature search modeling of multi data objective programming

#### 4.1 multi objective planning

In many practical search problems, there are usually multiple incompatible targets, which need to be considered at the same time. Multi objective programming is defined as maximizing (or minimizing) multiple objective functions under a set of constraints. Its general form is as follows:

\[
\max \{f_1(X), f_2(X), ..., f_m(X)\} \\
\text{s. t.} \\
g_j(X) \leq 0, j = 1, 2, ..., p,
\]

\[X = (x_1, x_2, ..., x_n)\] is a n dimension decision vector, \(f_i(X)\) is the objective function, \(i = 1, 2, ..., m, g_j(X) \leq 0\) is the system constraint \(j = 1, 2, ..., p\).

When the objective function is in conflict state, there is no optimal solution to optimize all the objective functions simultaneously. In this case, we use the concept of efficient solution to show that it is impossible to improve any objective function without sacrificing other objective functions. Specifically, a solution \(X^*\) is called an efficient solution. If \(X \in S\) does not exist, \(f_i(X) \geq f_i(X^*), i = 1, 2, ..., m\) and the unequal sign hold for at least one sequence number \(j\).

#### 4.2 establishment of literature search model

In the multi-objective decision-making problem, it is assumed that the decision-maker designs an objective value for each objective. The idea is to minimize the deviation (positive deviation, negative deviation or positive deviation) between each objective function and the objective value. One goal can only be achieved at the expense of others, which are generally incompatible. Therefore, establish a priority between these incompatible goals, and rank all objectives according to this priority, so as to achieve as many goals as possible. In order to balance multiple conflicting objectives, the literature search problem can be modeled as a goal programming model according to the decision maker's goal value and priority structure. The general form can be expressed as follows:

If the decision maker integrates \(m\) objective functions together, the preference function can be maximized under the same constraints. One compromise model is established by weighting the objective function, and the other is to minimize the distance function from \(((f_1(x), f_2(x), ..., f_m(x))\) to an ideal
vector \((f_1^*, f_2^*, ..., f_m^*)\), where \(f_i^*\) is the optimal value of the \(i\) target without considering other targets, \(i = 1, 2, ..., m\). Another is to use human-computer interaction to find compromise solutions.

\[
\begin{align*}
\min & \sum_{j=1}^{l} P_j \sum_{i=1}^{m} (u_{ij} d_i^+ + v_{ij} d_i^-) \\
\text{s.t.} & \quad f_i(X) + d_i^- - d_i^+ = b_i, \quad i = 1, 2, ..., m \\
& \quad g_j(X) \leq 0, \quad j = 1, 2, ..., p \\
& \quad d_i^-, d_i^+ \geq 0, \quad i = 1, 2, ..., m
\end{align*}
\]

Among them, \(P_j\) priority factor represents the relative importance of each target, and for all \(j\), there is a weight factor for \(P_j \geq P_{j+1}\), \(u_{ij}\) corresponding to the priority factor \(j\) for the positive deviation of the \(i\) goal, \(v_{ij}\) for the priority factor \(j\), the weight factor for the negative deviation of the \(i\) goal, and the positive deviation of \(d_i^+\) target \(i\) from the target value are defined as follows:

\[
d_i^+ = \begin{cases} f_i(X) - b_i, & f_i(X) > b_i \\ 0 & f_i(X) \leq b_i \end{cases}
\]

The negative deviation of \(d_i^-\) target \(i\) from the target value is defined as:

\[
d_i^- = \begin{cases} 0, & f_i(X) \geq b_i \\ b_i - f_i(X), & f_i(X) < b_i \end{cases}
\]

\(X\) is the \(n\) dimension decision vector, \(f_i\) is the function in the objective constraint, \(g_i\) is the function in the system constraint, \(b_i\) is the target value of objective \(i\), \(l\) is the priority number, \(m\) is the number of objective constraints, \(p\) is the number of system constraints.

\[
\text{lex min} \left\{ \sum_{i=1}^{m} (u_{il} d_l^+ + v_{il} d_l^-), \ldots, \sum_{i=1}^{m} (u_{il} d_l^+ + v_{il} d_l^-) \right\}
\]

\[
\text{s.t.} \quad f_i(X) + d_i^- - d_i^+ = b_i, \quad i = 1, 2, ..., m \\
& \quad g_j(X) \leq 0, \quad j = 1, 2, ..., p \\
& \quad d_i^-, d_i^+ \geq 0, \quad i = 1, 2, ..., m
\]

Where \(\text{lexmin}\) is to minimize the target vector in lexicographic order.

5. Conclusion

Literature full-text search has the advantages of large amount of information, comprehensive and sufficient query, but the amount of information returned is too much, users need to select their own demand information. This model can intelligently compare the actual content of new papers and existing literature. After small sample verification, the "copy ratio" of the real content can be calculated by the target planning target algorithm, which can open up a new idea for the upgrading of literature search tools.

References

[1] Zhou Xiaolan, Jin Sheng, Xie Hong. Science and technology information retrieval and application [M]. Beijing: China Electric Power Press, 2012:5-205

[2] Wang Liping, Gu Guoqing. A new course of document information retrieval [M]. Beijing: People's Posts and Telecommunications Press, 2014:1-138

[3] Li Qiang, Chen Xue, Bai Jun. practical technology of literature retrieval [M]. Beijing: People's Posts and Telecommunications Press, 2014:1-250
[4] Li Kan. Validity analysis and tracking of articles published in academic journals [M]. Science and Technology Literature Press, 2020:1-286

[5] Ren Fujun, Zhai Jiequan. Introduction to science and technology communication and popularization [M]. China Science and Technology Press, 2014:1-296