Research on Bottleneck Identification and Decision Optimization of Public Welfare Development from the Perspective of Multi-Source Data Fusion

Han Wang, Jing Zhang
Liaoning Institute of Science and Technology, Benxi, Liaoning, China

Abstract:From the core perspective of multi-source data fusion technology, this paper addresses the practical pain points of public welfare undertakings, such as "data fragmentation, experience-based decision-making, and difficulty in quantifying effectiveness". Through literature analysis, case disassembly, and model construction, it identifies the core bottlenecks in data application of current public welfare undertakings, and further proposes a decision optimization path based on multi-source data fusion. The research finds that data barriers, insufficient technical adaptation, and weak value transformation capabilities are the key issues restricting the high-quality development of public welfare undertakings. The "data layer - feature layer - decision layer" three-level fusion model can effectively improve the accuracy, transparency, and sustainability of public welfare projects, providing theoretical references and practical paradigms for the digital transformation of public welfare undertakings.

Keywords: Multi-Source Data Fusion; Public Welfare Undertakings; Bottleneck Identification; Decision Optimization; Digital Transformation

1. Introduction

In recent years, China's public welfare undertakings have developed rapidly. By 2024, the number of registered charitable organizations nationwide has exceeded 16,000, and the annual total of charitable donations has exceeded 150 billion yuan. However, along with the expansion of scale, public welfare projects have also shown problems such as insufficient targeting, difficulty in tracing effects, and resource waste. The core crux lies in the dispersion of public welfare data among multiple entities, forming "data silos" that are difficult to support scientific decision-making.

Multi-source data fusion technology can integrate structured, semi-structured, and unstructured data to achieve in-depth exploration of data value. It has been proven effective in fields such as smart cities and precision medicine. Introducing it into public welfare undertakings is not only an inevitable choice to solve current development bottlenecks but also a core driver for the transformation of public welfare from experience-driven to data-driven. Theoretically, this research can fill the gap in the interdisciplinary field of multi-source data fusion and public welfare management, and construct a logical framework of data fusion, bottleneck identification, and decision optimization to enrich the theoretical system of public welfare digitalization. In practice, it can provide implementable data fusion solutions for charitable organizations and government supervision departments, helping public welfare projects accurately match needs, optimize resource allocation, enhance social credibility, and promote the high-quality development of public welfare undertakings.

2. Core Logic of Multi-Source Data Fusion

Multi-source data fusion refers to a technical system that collects, cleans, correlates, and analyzes data from different sources and types, eliminates data redundancy and conflicts, and forms comprehensive information with a value of "1+1>2" to support decision-making.

In the field of public welfare, data sources are extensive and diverse. They include not only structured data such as government policy documents, charitable organization project ledgers, and corporate donation records but also semi-structured data like volunteer service logs and project progress reports, as well as unstructured data such as public comments on social media, interview videos of beneficiary groups, and photos of disaster relief sites. These

data are stored in the information systems of different entities, with different formats, inconsistent standards, and even cases of duplicate or conflicting data. Multi-source data fusion technology is the key to solving these problems, and its core logic can be advanced layer by layer through three-level fusion to realize the value enhancement of data from disordered to ordered and from scattered to systematic.

2.1 Data Layer Fusion

As the basic link of multi-source data fusion, the data layer fusion mainly solves the problem of "whether data can be used". Due to the different data collection standards of various entities, raw data often has problems such as chaotic formats, missing information, and exposure of private information. For example, some donation records are stored in tabular form, while others are recorded in text; beneficiary group information may include sensitive content such as ID numbers and home addresses. Through standardized processing such as unifying data formats, supplementing missing information, and desensitizing private data, data layer fusion converts scattered raw data into a unified and standardized dataset, laying the foundation for subsequent fusion links. For instance, it unifies the amount data from different donation channels into the same numerical unit and encrypts the personal sensitive information of beneficiary groups to ensure that the data is operable under the premise of compliant use.

2.2 Feature Layer Fusion

Focusing on solving the problem of "what the data is used for", the feature layer fusion is a key link connecting the data layer and the decision layer. Based on standardized data, the feature layer fusion extracts key features of each data source through data mining technology and explores the hidden information value behind the data. Taking public welfare projects as an example, public satisfaction, focus points, and potential needs for the project can be extracted from social media public opinion data; indicators such as fund allocation efficiency and usage ratio can be analyzed from fund flow data; and the actual effects and existing deficiencies of project implementation can be summarized from the feedback data of beneficiary groups. The extracted feature information can not only directly reflect the core value of the data but also provide accurate information support for subsequent decision-making, avoiding judgment biases caused by one-sided information in the decision-making process.

2.3 Decision Layer Fusion

The decision layer fusion, which focuses on solving the problem of "how to use the data", is the ultimate embodiment of the value of multi-source data fusion. This link needs to combine specific public welfare goals (such as the needs of projects in different fields like disaster relief, student assistance, and environmental protection), input the key information extracted from the feature layer into corresponding decision models (such as regression analysis models and machine learning models), evaluate various possibilities of project implementation through model operations, and finally output the optimal plan. For example, in disaster relief public welfare projects, inputting data such as the population data of the disaster-stricken area, material demand characteristics, and donated material data into the decision model can yield the optimal material allocation plan; in student assistance projects, combining data such as the academic performance of beneficiary students, family economic status characteristics, and donated funds can formulate a precise student assistance resource allocation plan.

In the field of public welfare, the value of multi-source data fusion is not only reflected in the technical logic of the three-level fusion but also in connecting the data of donors, beneficiaries, implementers, and supervisors through this technology, breaking down data barriers between different entities, and realizing a full-process data closed loop from demand analysis at the project initiation stage, accurate matching during resource allocation, real-time tracking in the implementation process to effect evaluation after the project conclusion. This ensures that every link of the public welfare project is supported by data, which not only improves the scientificity and accuracy of public welfare decision-making but also provides a strong guarantee for the transparent and standardized development of public welfare undertakings.

3. Bottleneck Identification of Public Welfare Development from the Perspective of Multi-Source Data Fusion

Based on the three-level logic (data layer, feature layer, and decision layer) of multi-source data fusion and the analysis of practical cases of domestic public welfare organizations, the current public welfare undertakings still face various bottlenecks in promoting development with the help of data fusion. These bottlenecks correspond to different links of the fusion process and jointly hinder the effective release of data value.

3.1 Data Layer

The prominent data barriers between multiple entities have become the primary obstacle restricting data fusion, resulting in an extremely weak foundation for fusion. Specifically, there are three prominent problems: "unwillingness to fuse", "inability to fuse", and "fear of fusing".

The crux of "unwillingness to fuse" lies in the vague definition of data ownership and the differences in the interest demands of various entities. Data holders such as government departments, charitable organizations, and enterprises generally have a mindset of data privatization, worrying that data sharing may lead to information leakage, responsibility disputes, or damage to their own interests. For example, the subsistence allowance data of beneficiary groups held by local civil affairs departments involves sensitive information such as the economic status of beneficiary families. Due to concerns about data security and subsequent responsibility attribution, they are often unwilling to share it with public welfare organizations, making it difficult for public welfare projects to accurately connect with groups that truly need help.

The "inability to fuse" stems from the inconsistent data standards of various entities. Different public welfare platforms have significant differences in the definition of basic data such as the statistical caliber of the number of beneficiaries and the division standards of project progress. The formats and statistical dimensions of raw data are different, making it impossible to directly connect and integrate them. Additional efforts need to be invested in format conversion and caliber unification, which greatly increases the cost of fusion.

The "fear of fusing" is mainly limited by the imperfection of detailed rules for data security laws and regulations. Public welfare data contains a large amount of personal privacy information such as the home addresses of beneficiary groups and the identity information of donors. Although relevant laws and regulations have put forward requirements for data management, specific standards for data desensitization, security specifications for storage, and encryption methods for transmission have not yet been clarified. Fearing of crossing the compliance red line and causing privacy leakage risks, public welfare organizations adopt a cautious attitude towards promoting data fusion and dare not easily carry out cross-entity data integration.

3.2 Feature Layer

Problems such as insufficient technical adaptation and weak data value extraction capabilities are prominent. Most public welfare organizations, especially small and medium-sized ones, lack professional technical support, resulting in a large amount of data remaining in a "dormant" state and unable to be converted into valuable feature information.

From the perspective of technical tools, multi-source data fusion requires professional technical support such as data cleaning tools and feature extraction algorithms. However, more than 70% of small and medium-sized public welfare organizations do not have professional technical teams and are short of technical talents. Their daily data processing can only rely on Excel for simple statistical summary, and they cannot use professional tools for in-depth cleaning and feature mining of multi-source data.

From the perspective of feature dimensions, the current application of data in the public welfare field has obvious limitations. Most of the focus is on fund data, such as the statistics of donation amounts and the calculation of expenditure ratios, while ignoring the important value of non-fund data. Taking disaster relief public welfare projects as an example, real-time image data of disaster-stricken areas can directly reflect the severity of the disaster and the urgency of needs in different areas, providing key references for material deployment. In student assistance projects, the learning trajectory data of beneficiary students can be used to track and evaluate the actual impact of the project on students' academic performance for a long time. However, such non-fund data is often not included in the scope of feature extraction, resulting in one-sided information in the feature layer, which cannot fully reflect the actual

situation of public welfare projects and is difficult to support subsequent accurate decision-making.

3.3 Decision Layer

The problem of disconnection between data and public welfare goals is prominent, and the efficiency of data value transformation is low. Even if some public welfare organizations overcome the obstacles in the previous links to achieve data fusion, they still face the dilemma of "data being unused". The core reason is that the decision layer fusion fails to be closely combined with public welfare goals.

On the one hand, the adaptability of decision models is poor. Most of the existing data decision models are derived from the commercial field, such as precision marketing models in the e-commerce industry. These models are centered on maximizing commercial interests and do not fully consider the special attribute of public welfare undertakings that prioritize social benefits. For example, if a single indicator of fund use efficiency is simply used to determine the priority of public welfare projects, it may lead to the priority selection of projects with low implementation difficulty, fast fund turnover but limited social benefits, while shelving projects that are complex in implementation, relatively low in fund use efficiency but related to the urgent needs of groups, which deviates from the original intention of public welfare undertakings.

On the other hand, there is a lack of effect feedback mechanisms. The effects of public welfare projects are often lagging. For example, student assistance projects need three to five years or even longer to fully reflect their impact on the future development of beneficiary students. However, current decisions mostly rely on short-term data, such as immediate indicators like project completion rate and fund utilization rate. There is a lack of a closed loop of long-term data tracking and feedback, making it impossible to timely capture changes in the long-term effects of projects. As a result, decision optimization lacks a scientific basis, and it is difficult to adjust subsequent strategies according to the actual impact of projects, further reducing the supporting role of data fusion in public welfare decision-making.

4. Decision Optimization Path of Public Welfare Undertakings Based on Multi-Source Data Fusion

In response to the bottlenecks faced by public welfare undertakings at the data layer, feature layer, and decision layer, and combined with the three-level logic of multi-source data fusion, the following optimization paths can be constructed from three dimensions (data integration, feature extraction, and decision implementation) to promote the transformation of public welfare decision-making towards scientificity and accuracy.

4.1 Construct a Government-Led and Industry-Coordinated Public Welfare Data Sharing System at the Data Layer

The core of optimization at the data layer lies in breaking down data barriers and laying the foundation for fusion. The civil affairs department needs to take the lead and work with charitable industry associations to formulate public welfare data classification and sharing standards, clearly defining the scope of sharable data, restricted sharable data, and non-sharable data. Sharable data includes basic information such as project types and beneficiary areas; restricted sharable data needs to undergo desensitization processing, such as anonymized information of beneficiary groups; non-sharable data focuses on core privacy content. By clearly defining these scopes, the concerns of various entities regarding "unwillingness to fuse" and "inability to fuse" can be addressed.

At the same time, referring to the construction model of the national government service platform, a national public welfare data sharing platform should be built. Entities such as government departments, charitable organizations, and enterprises can access the platform through authorization, realizing "one-time data collection and multiple reuses". This greatly reduces the data acquisition cost of small and medium-sized public welfare organizations and avoids resource waste caused by repeated data collection.

In addition, it is necessary to improve the data security mechanism, introduce blockchain technology to ensure that public welfare data is traceable and non-tamperable, and at the same time clarify the security responsibilities of data users, requiring them to implement data encryption storage measures and establish a responsibility system for illegal use. This eliminates the concerns of various entities about "fear of fusing" from both technical and

institutional aspects and ensures the safe and compliant sharing of data.

4.2 Develop Lightweight and Scenario-Specific Public Welfare Data Feature Extraction Tools at the Feature Layer

The key to optimization at the feature layer is to improve the data value extraction capabilities of public welfare organizations and activate "dormant" data.

On the one hand, we can cooperate with technology enterprises to develop dedicated data toolkits for the public welfare field, such as public welfare data cleaning and feature extraction plug-ins. These toolkits need to simplify the operation process, adopt a visual interface design, and support one-click generation of feature reports. This allows public welfare staff without a technical background to easily use them, complete data cleaning and feature extraction without relying on professional technical teams, and reduce the threshold for technology use.

On the other hand, it is necessary to focus on the needs of different public welfare scenarios and design targeted feature dimensions. In the disaster relief scenario, key features such as disaster-stricken area, number of affected people, traffic accessibility, and material gap are extracted to provide a basis for material deployment and rescue arrangement; in the student assistance scenario, features are extracted around dimensions such as students' academic performance, family economic status, school resources, and project participation duration to help evaluate the effect of student assistance and optimize resource allocation; in the environmental protection scenario, features such as pollution concentration, governance area, residents' satisfaction, and ecological restoration progress are extracted to support the implementation tracking and effect evaluation of environmental protection projects. Through scenario-specific feature extraction, it is ensured that the feature information fully meets public welfare needs and provides accurate support for decision-making.

4.3 Establish a Public Welfare Decision Model with Social Benefit Priority and Multi-Indicator Linkage at the Decision Layer

The optimization at the decision layer needs to solve the problem of disconnection between data and public welfare goals and improve the efficiency of data value transformation.

First of all, a multi-objective decision indicator system should be constructed to break through the limitation of a single fund efficiency indicator. Indicators should be set from three dimensions: social benefits, efficiency, and sustainability. The social benefit dimension includes indicators such as the coverage rate of beneficiary groups and demand matching degree; the efficiency dimension covers indicators such as fund utilization rate and project completion cycle; the sustainability dimension focuses on indicators such as public participation and the stability of corporate cooperation. Through multi-dimensional indicators, the value of public welfare projects is comprehensively measured, and the deviation of decision-making towards short-term interests is avoided.

Secondly, it is necessary to develop adaptive models for different decision scenarios. In the project approval stage, a three-dimensional model of "demand matching degree - social benefit - implementation difficulty" is used to screen projects, giving priority to projects with high demand, high social benefits, and low implementation difficulty to ensure that resources are invested in key areas; in the resource allocation stage, a model of "number of beneficiaries - urgency - regional balance" is used to allocate funds, avoiding excessive concentration of resources in areas with easy implementation and ensuring the fairness and accuracy of resource allocation.

Finally, a long-term effect feedback closed loop should be established to conduct short-term, medium-term, and long-term data tracking of public welfare projects. Short-term tracking focuses on project progress within one year, medium-term tracking pays attention to the preliminary effects within 1-3 years, and long-term tracking monitors the far-reaching impact for more than 3 years. The feedback data from each stage is re-input into the decision model to continuously optimize decision-making strategies, forming a continuous iteration mechanism of "decision-making - implementation - feedback - optimization" and promoting the continuous improvement of public welfare decision-making.

5. Conclusions and Prospects

Through the analysis from the perspective of

multi-source data fusion, this paper finds that the core bottlenecks restricting the development of current public welfare undertakings are prominent data barriers at the data layer, insufficient technical adaptation at the feature layer, and inefficient value transformation at the decision layer. The three-level optimization path of constructing a data sharing system, feature extraction tools, and a multi-objective decision model can effectively solve these bottlenecks and promote the transformation of public welfare undertakings from experience-driven to data-driven.

Future research can further focus on two directions: one is to explore the combination of large AI models and public welfare data fusion, such as using generative AI to automatically generate public welfare project decision suggestions; the other is to optimize the data fusion model for small-sample data scenarios of grass-roots public welfare organizations (such as county-level organizations), for example, introducing transfer learning technology to make multi-source data fusion technology better serve grass-roots public welfare undertakings.

Acknowledgments

This paper is supported by 2026 College Student Innovation and Entrepreneurship Project of Liaoning Institute of Science and Technology - "Time and Space Trap Script Killing + Cross-Border Integration" (No. 202611430067).

References

[1]Wang Ming, Li Yong. China Charity Development Report (2024) [M]. Beijing: Social Sciences Academic Press (China), 2024.

[2]Zhang Li, Liu Jun. Research on the Application of Multi-Source Data Fusion Technology in the Field of Public Services [J]. Journal of Intelligence, 2023, 42(5): 123-129.

[3]Ministry of Civil Affairs. Measures for the Disclosure of Charitable Organization Information (Revised Version) [Z]. 2022.

[4]Tencent Public Welfare. 2023 Public Welfare Data Fusion Practice Report [R]. Shenzhen: Tencent Public Welfare Foundation, 2024.

