Article

Multi-Modal Spatio-Temporal Knowledge Graph of Ship Management

Yitao Zhang [1,†], Ruiqing Xu [1,†], Wangping Lu [1,†], Wolfgang Mayer [2], Da Ning [3], Yucong Duan [4,*], Xi Zeng [1,*] and Zaiwen Feng [1,5,6,7,8*]

1 College of Informatics, Huazhong Agricultural University, Wuhan 430070, China; xrq15564786383@webmail.hzau.edu.cn (R.X.)
2 Industrial AI Research Centre, University of South Australia, Mawson Lakes, SA 5095, Australia
3 No. 722 Research Institute of CSSC, Wuhan 430205, China
4 School of Computer Science and Technology, Hainan University, Haikou 570228, China
5 Hubei Three Gorges Laboratory, Wuhan 430070, China
6 Hubei Key Laboratory of Agricultural Bioinformatics, Wuhan 430070, China
7 Key Laboratory of Smart Farming Technology for Agricultural Animals, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China
8 Engineering Research Center of Intelligent Technology for Agriculture, Ministry of Education, Wuhan 430070, China
* Correspondence: duanyucong@hainanu.edu.cn (Y.D.); zengxi@mail.hzau.edu.cn (X.Z.); zaiwen.feng@mail.hzau.edu.cn (Z.F.)
† These authors contributed equally to this work.

Abstract: In modern maritime activities, the quality of ship communication directly impacts the safety, efficiency, and economic viability of ship operations. Therefore, predicting and analyzing ship communication status has become a crucial task to ensure the smooth operation of ships. Currently, ship communication status analysis heavily relies on large-scale, multi-source heterogeneous data with spatio-temporal and multi-modal features, which presents challenges for ship communication quality prediction tasks. To address this issue, this paper constructs a multi-modal spatio-temporal ontology and a multi-modal spatio-temporal knowledge graph for ship communication, guided by existing ontologies and domain knowledge. This approach effectively integrates multi-modal spatio-temporal data, providing support for subsequent efficient data analysis and applications. Taking the scenario of fishing vessel communication activities as an example, the query tasks for ship communication knowledge are successfully performed using a graph database, and we combine the spatio-temporal knowledge graph with graph convolutional neural network technology to achieve real-time communication quality prediction for fishing vessels, further validating the practical value of the multi-modal spatio-temporal knowledge graph.

Keywords: ship communication; multi-modal heterogeneous data; multi-modal spatio-temporal ontology; multi-modal spatio-temporal knowledge graph

1. Introduction

In recent years, the booming development of maritime activities such as oceanic travel, coastal aquaculture, and deep-sea mining exploration has led to an increasing number of ships, offshore platforms, and buoys, thereby driving the growing demand for high-speed and reliable maritime communication [1]. Currently, decision-making authorities urgently need to obtain real-time information on the location, activities, and communication status of ships in remote sea areas, so as to accurately predict the ship's future trends. The application of a large number of sensors, increased data storage capacity, cost-effective devices, and improved database management systems make it possible to predict the communication status of maritime vessels.

Various maritime activities have accumulated a large amount of ship situation data and communication status data, including ship positioning, attribute data and relevant geographic information, which are stored and managed through relational databases. Geographic information data related to ship movement and operation include ship operations and communication environments, natural environmental condition data, and human-related activity data, which are mainly used to accurately understand and monitor maritime activities and further provide significant data-driven technological innovation space for ship communication status analysis. Although many researchers have attempted to apply traditional big data analysis techniques for analyzing and predicting ship communication status based on the communication data in MySQL databases, there are still some limitations. For example, similarity-based methods, Kalman filtering, exponential smoothing, and fuzzy prediction have encountered challenges in parameter adjustment, establishing evaluation functions, limited learning capabilities, and susceptibility to subjective influences. In response to these issues, Bernhard Schölkopf et al. [2] proposed the analysis method of Support Vector Machines (SVM). This model exhibits fast convergence and the ability to address high-dimensional recognition problems, but it transforms originally simple problems into complex nonlinear regression problems. On the other hand, artificial neural networks possess strong adaptability, autonomous learning capabilities, excellent information retention, and optimization algorithms, making them more suitable for prediction tasks involving unstructured data and considering various interfering factors. However, artificial neural networks suffer from slow convergence and the potential to converge to local optima [3,4]. Moreover, these traditional methods primarily utilize ship positioning data for data mining and visualization, without considering the analysis of ship spatio-temporal activity processes and behavior patterns, as well as the application of domain knowledge.

In addition, maritime ship situation and communication status data, originating from diverse sources, often exhibit challenges in terms of large data volumes, inconsistent formats, and varying descriptions. Also, the voluminous trajectory data for ship activity with low knowledge density, making it difficult to perform in-depth knowledge mining [5]. Consequently, the unified organization, management, and application of such data become arduous tasks. As a NoSQL database, the graph database adopts a model of nodes, edges, and attributes to describe large-scale, multi-source heterogeneous data in a unified manner [6]. It can not only represent and processes complex semantic associations between data, but also mine and infer more knowledge for data analysis. Furthermore, the graph database also provides highly flexible and efficient query services. Recently, there has been a growing focus on the development of spatio-temporal knowledge graphs stored in graph databases, driven by the continuous improvement in computing efficiency for the attribute analysis of large interconnected datasets [7]. The robust integration capability of spatio-temporal knowledge graphs in managing temporal and spatial data resources offers effective solutions for these challenges. And spatio-temporal knowledge graph technology has found widespread application in various fields, including forest fire prediction. For example, Ge et al. proposed a forest fire prediction method that combines spatio-temporal knowledge graphs with machine learning models to efficiently extract the required features [8]. However, the application of this technology in analyzing and predicting ship communication quality has been limited. To address this gap, we have developed a spatio-temporal knowledge graph based on ship communication.

The contributions of this paper are briefly summarized as follows. (1) A multi-modal spatio-temporal knowledge graph is proposed to effectively address the issue of multi-source heterogeneous ship communication data management, aiming to facilitate data sharing and reuse. (2) A multi-modal spatio-temporal ontology is constructed by considering the multi-modal and spatio-temporal characteristics of ship communication data, which provides a more comprehensive and accurate data representation standard and improves the semantic interoperability between data. (3) Taking the communication scenario of fishing vessels going out to sea as an example, we validated the feasibility of using a

multi-modal spatio-temporal knowledge graph for communication quality prediction and communication knowledge query tasks.

2. Related Work

A knowledge graph is a data modeling method that represents knowledge as concepts, entities, and semantic relationships between them in the form of a graph [9]. However, the world contains a vast amount of dynamic and procedural knowledge that conventional static knowledge graphs cannot adequately represent [10]. Spatio-temporal knowledge graphs not only enable the representation of entities but also capture the spatio-temporal changes in those entities. By connecting multi-source spatio-temporal ship communication data and expert knowledge in a graph structure, spatio-temporal knowledge graphs facilitate the dynamic analysis and prediction of communication in monitoring scenarios involving heterogeneous data sources.

In the research involving the analysis and prediction of maritime communication and activities through the utilization of a knowledge graph, Liu et al. [11] predicted missing nodes in a maritime knowledge graph by link prediction in the knowledge graph. However, they fell short in leveraging temporal and spatial information as guiding factors to achieve more robust predictions. A dynamic method for predicting knowledge graph links was proposed in [12] for identifying navigation scenarios at sea. Dynamic knowledge graphs are used to capture the evolution of entities such as ships, ports, and countries. The study was limited to conducting rudimentary experiments in positional prediction. To achieve the event and attribute predictions as described in the article, substantial reliance on extensive expert knowledge is indispensable. Wen et al. [13] introduced a semantic model of ship behavior (SMSB) to describe the behavior and status of ships on a route, including sailing, anchoring, and stopping. The status is recognized and established by rules, and the potential behavior is inferred by a dynamic Bayesian network (DBN). Liu [5] improved the SEM (Simple Event Model) model based on the core idea of "process-event-behavior" and designed a ship activity ontology model. The semantic information of trajectories is extracted using the Stop/Move model and geographic correlation relationship, and the relationship between ship sudden events and normal events is extracted using a deep learning model to complete instance-level filling. Ren et al. [14] used information mining technology to perform spatio-temporal and event correlation analysis on the historical information of ships, forming a knowledge graph analysis system for vertical domain intelligence information on foreign military ship activities. But their works were limited to basic applications such as querying and visualization, without incorporating inferential reasoning into the ship activity knowledge graph.

Overall, the application of knowledge graph-driven data analysis technology to solve maritime multi-scenario prediction tasks has become an important trend at present, but there is still a lack of research on the management and analysis of ship communication data. Therefore, this paper designs and constructs a multi-modal space-time knowledge graph to uniformly organize and manage large-scale multi-source heterogeneous ship communication data, and provide knowledge support for subsequent data analysis and application.

3. Method

The ship communication data used in this paper were collected and obtained by means of ship-to-shore communication and automatic collection technology, and stored in the MySQL database. Through further analysis and processing of data, we aim to improve ship management and decision-making capabilities. Specifically, ship communication data can be divided into three categories: ship navigation data, communication basic resource data, and audio–visual image data. Ship navigation data cover important information such as track points, speed, sea area, and weather conditions. The communication basic resource database data include platform basic data, equipment resource basic data, and communication history data (used frequency band and signal strength and other information).

For multi-source heterogeneous ship communication data, this paper proposes a multi-modal spatio-temporal knowledge graph to integrate ship communication data with multi-modal and spatio-temporal characteristics, which provides outstanding data management support for subsequent data analysis and application tasks, such as the ship communication quality prediction task. As shown in Figure 1, multi-modal spatio-temporal ontology construction is for clear global concepts and semantic relations between concepts of multi-source heterogeneous ship communication data, and to achieve semantic interoperability between data. Based on the multi-modal spatio-temporal ontology, the method automatically maps ship communication data to the ontology by automatic semantic modeling and further achieves the aim of organizing and representing the data into a multi-modal spatio-temporal knowledge graph with hierarchical structure and correlation. Moreover, in order to more naturally organize and represent the relationships between entities, while facilitating more efficient data querying, this article adopts a graph database based on a graphical structure to store a multi-modal spatio-temporal knowledge graph containing complex information such as time and geographic location.
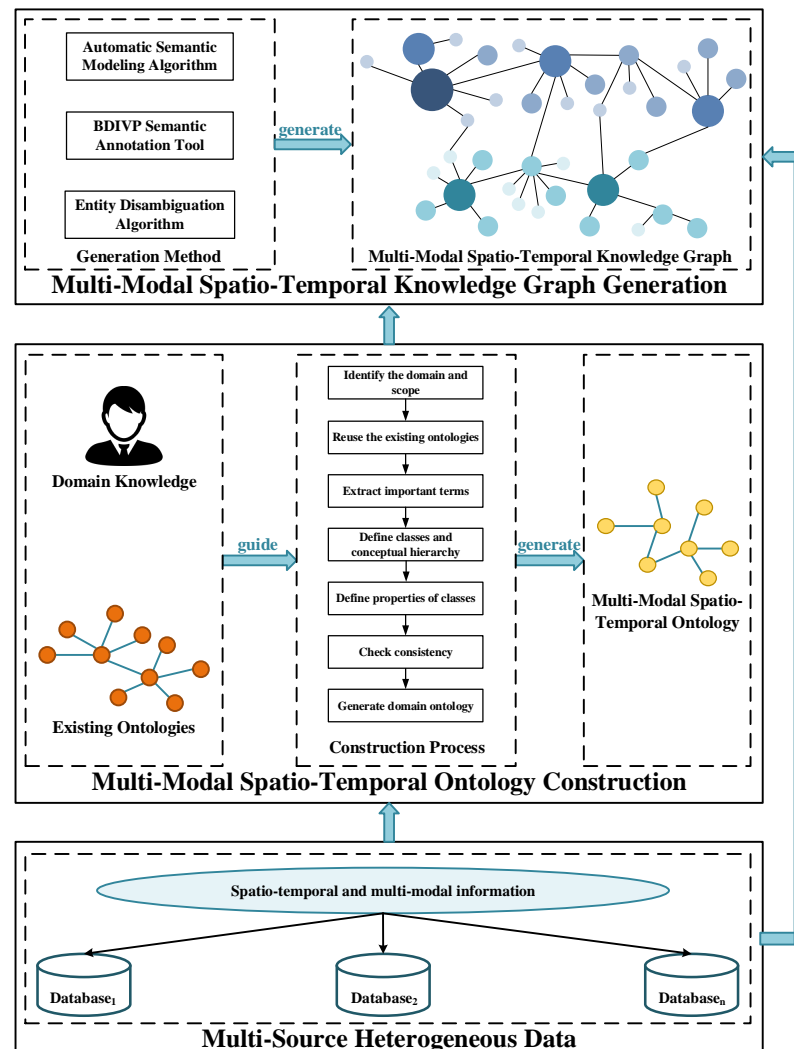
Figure 1. Multi-modal spatio-temporal ontology and knowledge graph construction framework.

3.1. Multi-Modal Spatio-Temporal Ontology Construction

Guided by the domain knowledge of ship communication, we consider reusing the existing ontologies (e.g., the known time ontology, geographic space ontology, and event ontology) to enhance the quality and efficiency of multi-modal spatio-temporal ontology construction. We apply different techniques to extract important terms from multi-modal

data in databases for class and property definitions in the multi-modal spatio-temporal ontology. And then, consistency checking is performed to obtain the ultimate multi-modal spatio-temporal ontology for establishing semantic association among heterogeneous data sources in ship communication.

3.1.1. Preliminary Preparation

Analyzing and determining the domain and scope of the ontology to be constructed from the existing data and the purpose of using ontology plays an important role. Not only can this step ensure that the designed ontology meets the practical needs of the application, but it is of great significance to the development and maintenance of the ontology.

There are "platform (*Plat*)", "sensor (*Sensor*)", "communication equipment (*Equipment*)", "event (*Event*)", "weather (*Weather*)", "area (*Area*)", and other relevant ship communication data from MySQL databases. Several types of data information, such as *Event* and *Area*, involve time and spatial information. For example, *Area* includes such spatial information as longitude, latitude, and height. Furthermore, *Sensor* and *Equipment* will generate multi-modal data, such as images and videos during ship communication. These data are of great significance to the analysis and prediction of ship communication situations. Therefore, it can be determined that the ontology constructed in this paper needs to achieve unified constraints and correlation integration of multi-source heterogeneous, spatio-temporal multi-modal data in ship communication, and further provide outstanding data management support for subsequent ship communication situation analysis and prediction tasks.

There are three methods to reuse existing ontologies for improving the quality and efficiency of ontology construction: (1) extending existing ontologies, (2) reusing existing ontologies, and (3) integrating multiple existing ontologies. In this paper, we design a multi-modal spatio-temporal ontology based on ship communication by integrating multiple existing ontologies.

Specifically, multi-modal spatio-temporal ontology is specifically designed to integrate and describe time and spatial information. There is an existing Time Ontology in OWL [15] that provides a clear, formal, and standardized description of time concepts and the relations between them. Therefore, we incorporate concepts *Instant* and *DateTimeDescription* from the Time Ontology in OWL to define the time-related description in our ontology. A set of geospatial data types, functions, and predicates have been defined in the existing ontology-based query language extension GeoSPARQL [16] for processing geospatial data. We can abstract geospatial concepts and relations between concepts from GeoSPARQL to form the geospatial ontology. For example, *Coordinate Reference System (CRS)* is used to determine the position and shape of geospatial data. Additionally, we reuse the structure of existing event ontology to define the ship communication event description and its related semantic information. Table 1 illustrates the reusing concepts and their corresponding descriptions.

The construction of a multi-modal spatio-temporal ontology involves extracting important terms from the data and analyzing their context to accurately identify their meanings and relations. This process is crucial for guiding the core structure design and semantic relations construction of the ontology. As the ontology to be constructed is a multi-modal spatio-temporal ontology, it is necessary to extract professional terms and spatio-temporal information-related terms from ship communication data, with a specific focus on extracting important terms from multi-modal data. The MySQL databases contain multi-modal information, including text, images, and videos. Therefore, this paper considers the application of various technical methods to extract high-quality important terms for constructing a comprehensive and accurate multi-modal spatio-temporal ontology, taking into account both the structural information of the MySQL databases and the characteristics of the multi-modal data. Specifically, various techniques, such as natural language processing, image processing, and video analysis can be employed to extract relevant terms from the

data. The extracted terms can then be utilized to construct a comprehensive and accurate ontology that effectively represents the spatio-temporal information.

Table 1. Reusing concepts and corresponding descriptions.

Name	Description
Instant	Representing the occurrence time of a communication event.
DateTimeDescription	Describing time information including year, month, day, hour, minute, second, etc.
CRS	Indicating coordinate reference systems (e.g., "WGS84", "UTM," etc.) for determining the position and shape of geospatial data.
Event	Linking dynamic ship communication information.

Extract terms from structure information of data

The data type and storage structure related to equipment information in the MySQL database are presented in Table 2. To extract important terms, we analyze and parse the structure and content information of the relational data table in the MySQL databases. The table name "communication equipment (Equipment)" is identified as an important concept term, as well as column names, such as "identifier (id)", "name", "nation", "type", "model", "description", "image", and "status" are also recognized as terms.

Table 2. Equipment data storage form in MySQL database.

```
Communication Equipment(Equipment)(
    id    INT PRIMARY KEY AUTO_INCREMENT,        – Equipment ID, self-increasing
    name    VARCHAR(64) NOT NULL,                – Equipment name
    nation    VARCHAR(64) NOT NULL,              – The country of the equipment
    type    VARCHAR(64) NOT NULL,                – Equipment type
    model    VARCHAR(64) NOT NULL,               – Equipment model
    description    TEXT NOT NULL,                 – Equipment description
    image    BLOB,                               – Equipment image
    status    VARCHAR(64) NOT NULL,              – Equipment malfunction situation
)
```

For the content information in the table, we employ data preprocessing (including data cleaning and segmentation) and text mining techniques such as term frequency statistics and TF-IDF algorithms to automatically extract important terms. For example, shortwave communication equipment terms such as "7300 type" and "726 shortwave communication equipment" are extracted from the "model" attribute.

Extract terms from multi-modal data

Ship communication involves multimedia modal data, including images, audio, and video. It is crucial to establish comprehensive correlations among the different modalities while integrating them.

- For image data in ship communication, both image recognition and manual annotation technology can be comprehensively applied to identify frequent image regions and use their labels as terms, where image recognition technology includes image preprocessing such as image denoising and image enhancement, as well as image segmentation and object detection techniques. Manual annotation technology, on the other hand, is employed to annotate objects and scenes in images, facilitating the extraction of relevant terms.
- For audio data in ship communication, the extraction of terms can be accomplished using audio analysis and manual annotation technology. Similar to extracting terms from images, we can annotate objects and scenes in audio data to obtain relevant terms. After automatic speech recognition and speech-to-text conversion, audio classification

and active speech detection techniques can be used to detect frequent entity targets from audio data and define their labels as terms.

- For video data in ship communication, image recognition can be employed to identify frequent image regions and assign their labels as terms after video preprocessing operations including video frame segmentation, inter-frame difference, image denoising, and image enhancement. Additionally, manual annotation technology can be utilized to annotate objects and scenes in the video, enabling the extraction of relevant terms.

3.1.2. Definition of Classes and Conceptual Hierarchy

To improve the semantic expression ability and inductive integration ability of an ontology, it is crucial to define classes and the hierarchical structure between them. This entails determining the parent–child relations between classes and ensuring that the classes possess an appropriate level of generality to encompass and describe a specific range of instances.

Guided by domain knowledge in ship communication, we identify classes and their hierarchical structure from the more general terms of table names and column names. Initially, we establish preliminary unified concepts and semantic relations for describing ship communication data by defining *Plat*, *Equipment*, *Event*, *Image*, *Video*, *Instant*, and *Env* classes as shown in Figure 2. Since ships, submarines, and shore stations are all entities carried by communication equipment, they have common attributes and functions in the communication system. In order to maintain conceptual consistency, they are abstracted into "*Plat*" classes. Moreover, in ship communication events, the attribute parameters of communication equipment and the spatial information of the ship will change over time. In the instantiation process, the communication equipment entities and spatial information (environment) entities at different times need strict one-to-one correspondence. Therefore, we design a "*State*" [17] class connected with *Plat* class, which describes time series information, spatial position information, communication equipment information, and other attributes of ships in different times and spaces. Based on the time node of the latest communication event that occurred during the voyage, a *State* class node is added to the communication subject participating in the event, which connects the spatial position information of the current communication subject and the resource data of various communication equipment installed on the communication entity. As a result, considering the occurrence of communication events, the "*State*" node is added to represent the communication situation of the entity during the corresponding period/moment.

Plat	ships and other platforms info	State	state info of the platform at a given time
Event	communication event info	Equipment	communication equipment info on platforms
Instant	communication event time info	Env	environment info under a certain state
Image	image modal data	Video	video modal data

Figure 2. Core classes in the multi-modal spatio-temporal ontology.

Based on the above eight core classes, we defined *Ship*, *Submarine*, and *Shore Station* as subclasses of *Plat* class as shown in Figure 3. There are six subclasses including *Data Transfer Event*, *Communication Enhancement Event*, *Disconnection Event*, *Connection Device Event*, *Voice Call Event*, and *Message Event* for the *Event* class in Figure 4. As can be observed from Figure 5, *Equipment* class has these five subclasses such as *Communication Repeater Equipment*, *Satellite Communication Equipment*, *Radio Equipment*, *Optical Communication Equipment*, and *Radio Navigation Equipment*, where these subclasses also have lower subclasses, such as *GPS Receiver* of *Radio Navigation Equipment* class.
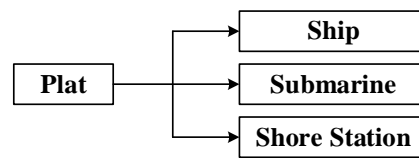
```
         ┌──────────────────┐
    ┌───▶│       Ship       │
    │    └──────────────────┘
┌──────┐ │    ┌──────────────────┐
│ Plat │─┼───▶│    Submarine     │
└──────┘ │    └──────────────────┘
    │    ┌──────────────────┐
    └───▶│  Shore Station   │
         └──────────────────┘
```

Figure 3. *Plat* class and its corresponding subclasses.

```
              ┌────────────────────────────────────┐
         ┌───▶│         Data Transfer Event         │
         │    └────────────────────────────────────┘
         │    ┌────────────────────────────────────┐
         ├───▶│        Disconnection Event          │
         │    └────────────────────────────────────┘
         │    ┌────────────────────────────────────┐
┌───────┐├───▶│         Voice Call Event            │
│ Event │┤    └────────────────────────────────────┘
└───────┘│    ┌────────────────────────────────────┐
         ├───▶│          Message Event              │
         │    └────────────────────────────────────┘
         │    ┌────────────────────────────────────┐
         ├───▶│      Connection Device Event        │
         │    └────────────────────────────────────┘
         │    ┌────────────────────────────────────┐
         └───▶│ Communication Enhancement Event     │
              └────────────────────────────────────┘
```
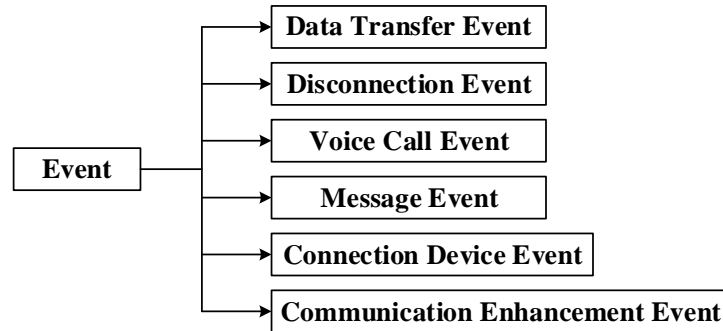
Figure 4. *Event* class and its corresponding subclasses.

```
                  ┌──────────────────┐            ┌──────────────────────────┐
             ┌───▶│ Optical Communi- │       ┌───▶│  Fiber Optic Transceiver │
             │    │ cation Equipment │───────┤    └──────────────────────────┘
             │    └──────────────────┘       └───▶│   Fiber Optic Modulator  │
             │                                     └──────────────────────────┘
             │    ┌──────────────────┐            ┌──────────────────────────┐
             ├───▶│  Radio Navigation│       ┌───▶│          Radar           │
             │    │    Equipment     │───────┤    └──────────────────────────┘
             │    └──────────────────┘       └───▶│       GPS Receiver       │
             │                                     └──────────────────────────┘
             │    ┌──────────────────┐            ┌──────────────────────────┐
             │    │                  │       ┌───▶│   Shortwave Transmitter  │
┌──────────┐ │    │                  │       │    └──────────────────────────┘
│Equipment │─┼───▶│ Radio Equipment  │───────┤───▶│    Shortwave Receiver    │
└──────────┘ │    │                  │       │    └──────────────────────────┘
             │    └──────────────────┘       ├───▶│ Ultra-shortwave Transmit.│
             │                               │    └──────────────────────────┘
             │                               └───▶│  Ultra-shortwave Receiver│
             │                                     └──────────────────────────┘
             │    ┌──────────────────┐            ┌──────────────────────────┐
             │    │Satellite Communi-│       ┌───▶│      Satellite Phone     │
             ├───▶│ cation Equipment │───────┤    └──────────────────────────┘
             │    └──────────────────┘       ├───▶│       Satellite TV       │
             │                               │    └──────────────────────────┘
             │                               └───▶│Satellite Internet Equip. │
             │                                     └──────────────────────────┘
             │    ┌──────────────────┐            ┌──────────────────────────┐
             │    │  Communication   │       ┌───▶│        Amplifier         │
             └───▶│Repeater Equipment│───────┤    └──────────────────────────┘
                  └──────────────────┘       ├───▶│       Transponder        │
                                             │    └──────────────────────────┘
                                             └───▶│         Repeater         │
                                                   └──────────────────────────┘
```
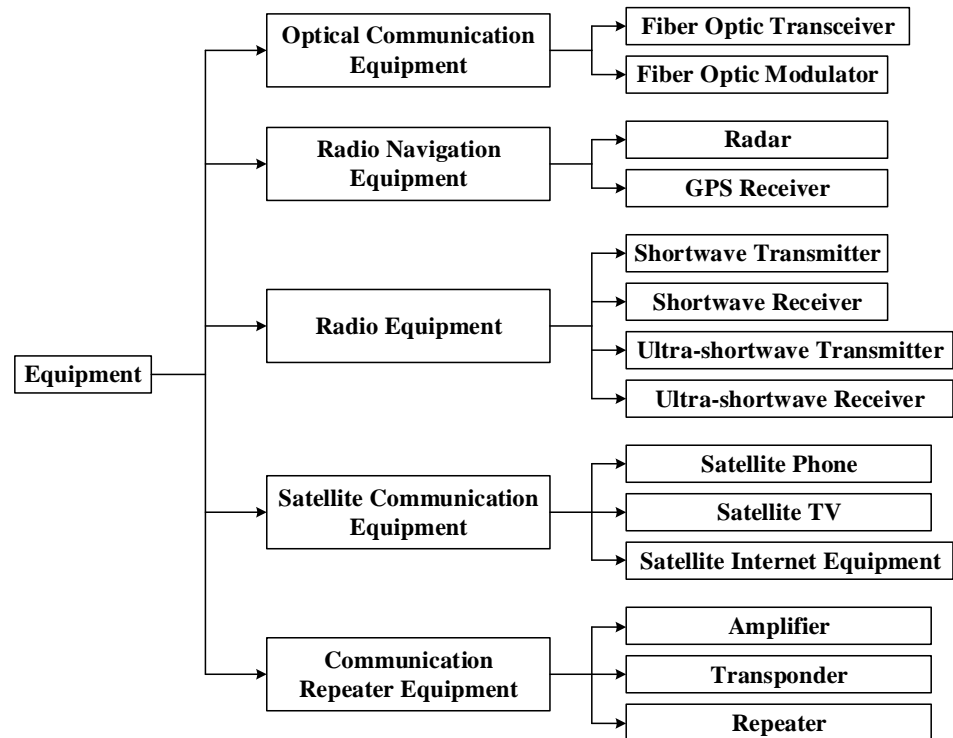
Figure 5. *Equipment* class and its corresponding subclasses.

3.1.3. Definition of Class Properties

The properties of a class include object properties and data properties, where object properties are used to describe the relations between classes, while data properties are used to describe relations between a class and its property values. When defining the properties of classes in ontology, it is necessary to determine both the object properties between classes and the data properties between a class and its property values. Additionally, it is important to identify the domain and range of these properties.

In general, verbs or verb phrases can serve as the basis for property naming. Some property terms have already been obtained while extracting terms from MySQL databases, such as "hasLongitude" and "hasLatitude". And these properties can be defined as data properties of *CRS* class to describe its specific properties about "longitude" and "latitude", whose domain is the *CRS* class and range is a float.

In the face of the case where object properties cannot be automatically extracted to establish semantic relations between classes, we adopt the "verb + class name" method to define object properties. For example, the object property "hasInstant" can be defined to describe the time of the communication event, whose domain is the *Event* class and range is the *Instant* class. In the term extraction phase, we also define the object properties "hasSubject" and "hasObject" for the *Event* class and the *Equipment* class based on foreign keys, where the domain of these two properties is the *Event* class and range is the *Equipment* class. In addition, for multi-modal data such as images, audio, and video, we define the object properties including "hasImage", "hasAudio", and "hasVideo" to establish the association between multi-modal data classes and other classes.

3.1.4. Consistency Check and Generation of Ontology

After the above steps, the definition of classes and related properties in the multi-modal spatio-temporal ontology has been basically completed. Where classes and properties in the ontology is used to organize multi-source, heterogeneous, and spatio-temporal multi-modal data. Following the above method, there may be contradictions or inconsistencies between the classes and properties defined from the multi-modal data in the databases. We primarily consider the following aspects to perform the ontology's consistency check. The ultimate multi-modal spatio-temporal ontology based on ship communication after consistency check is described in Figure 6.

(1) Check whether the classes and properties in the ontology match the actual data in the data source.
(2) Check whether the relationships between classes and properties in the ontology are consistent.
(3) Check whether the definitions in the ontology are consistent. If the same classes and properties are defined in different data sources using different naming or definition methods, machine learning (such as text similarity algorithms, clustering algorithms, etc.) or manual review should be used to standardize them to ensure the consistency of the ontology.
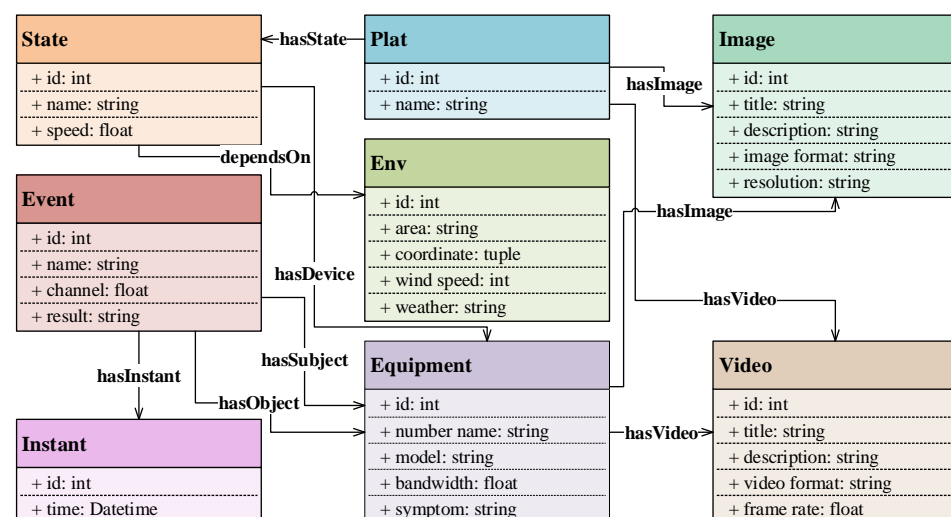
Figure 6. The ultimate multi-modal spatio-temporal ontology based on ship communication.

3.2. Multi-Modal Spatio-Temporal Knowledge Graph Generation

Knowledge graphs are large semantic networks, which encode relations between real-world facts through nodes and edges associated to semantic entities. One of the important reasons for integrating ship communication data into the knowledge graph is that they are helpful for downstream prediction tasks due to the ability of knowledge reasoning. We apply different techniques to extract knowledge from multi-modal data

with spatio-temporal information and represent it in the multi-modal spatio-temporal knowledge graph.

3.2.1. Knowledge Extraction from Unstructured Data

This paper employs the Transformer technology [18], known for its exceptional performance in feature extraction tasks, to accomplish the semantic extraction task of multi-modal unstructured data, encompassing text, image, and speech modalities. Initially, the raw textual, image, and speech data are preprocessed and transformed into formats compatible with the Transformer model. Then, the processed input sequences are fed into the Transformer encoder through the encoder and decoder of the Transformer, and the features of each modality data are extracted through the internal multi-head attention mechanism and other modules of the encoder. Following encoding, the input sequence proceeds to the decoder, which is followed by specific output heads for multi-modal data and is used to output the extracted triplets from each modality, providing data support for constructing a multi-modal knowledge graph.

Data Preprocessing

Due to the different characteristics of data in different modalities, it is necessary to preprocess the data of each modality and convert it into an input format that can be accepted by the Transformer encoder. The subsequent section will provide an overview of the data preprocessing steps for each modality.

For an image, its storage format in a computer is composed of individual pixels. Typically, each pixel of an image (assuming it is single-channel) is treated as a token, and its corresponding embedding operation is performed. Then, the embedding result is added to the corresponding positional encoding to obtain the final image embedding. However, for a single-channel image with a size of 224×224, treating each pixel as a token would result in an input length of 50,176 for the Transformer, which is too large and leads to an excessively large number of model parameters, making the model cumbersome and requiring more computational resources and time during training.

To address this issue, we adopt a patch-based method, which involves dividing the original image into small patches and treating each patch as a token, as shown in Figure 7. For a three-channel color image, its size format is [224, 224, 3]. The size of each patch is set to $16 \times 16 = 256$. Therefore, the original image can be divided into $(224/16)^2 = 196$ patches. Since it is a three-channel image, the size of each patch is $256 \times 3 = 768$. Hence, after patch processing, a $224 \times 224 \times 3$ image can be transformed into tokens with a size of 196×768, where num_token = 196 and token_dim = 768, which is in line with the input format required by Transformer. Before inputting it into the Transformer encoder, positional embedding needs to be added. After the aforementioned two steps of processing, the data can be inputted into the Transformer encoder for further processing.
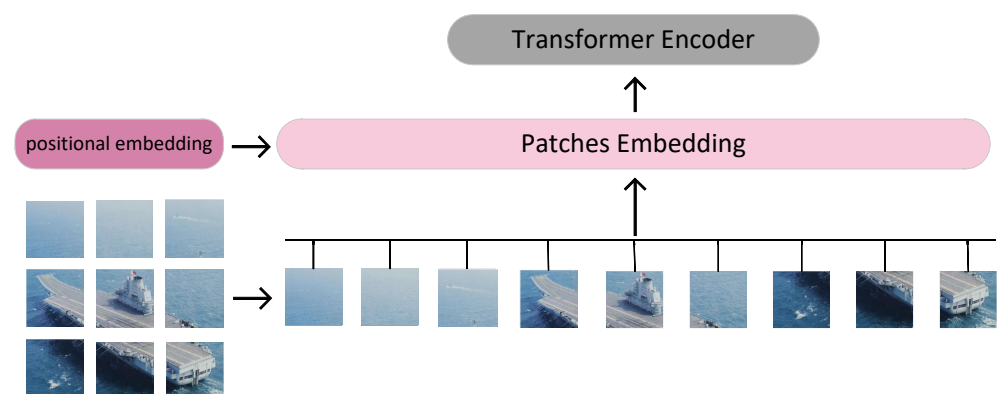

Figure 7. Patch-based image preprocessing.

The processing of textual data is relatively straightforward. Firstly, the raw corpus is formatted into sentences. Then, each word in each processed sentence is embedded to obtain an embedding sequence. This sequence consists of multiple embedding tokens, where each token represents the embedding of a word. Afterward, a [CLS] token and a [SEP] token are added to the beginning and end of the embedding sequence, respectively. Before inputting the sequence into the model, padding processing is performed and a corresponding padding mask vector is constructed. The purpose of padding processing is to maintain a consistent length of input sentences. Since the length of text varies, Pad tokens are added to shorter texts to make the sentence lengths consistent, which facilitates subsequent model processing and computation. The processed embedding sequence, positional embedding, and semantic embedding are integrated to form the final embedding input vector. The embedding input vector is then inputted into the model for feature extraction, and a classification layer is used to classify each output token. Finally, the predicted results of each token are post-processed to achieve the entire named entity recognition task.

For the speech modality, its processing can be briefly summarized as mapping the raw speech signal into a continuous space. Specifically, the feature sequence of the speech is transformed into the corresponding character sequence for subsequent operations and calculations. Since the speech sequence can be described as a two-dimensional spectrogram with a time axis and a frequency axis, its feature sequence is usually several times longer than the character sequence. When reading spectrograms, humans rely on the correlation between different frequencies over time to predict pronunciation. Therefore, focusing on the time and frequency axes may be advantageous for modeling the temporal and spectral dynamics in the spectrogram. We choose convolutional neural networks to exploit the structural locality of the spectrogram and alleviate length mismatch across time, ultimately transforming it into an input sequence that can be accepted by the Transformer encoder.

Encoder

In the encoder part, we employed three different Transformer encoders to fully extract the feature information of each modality, namely text, image, and speech. The adoption of distinct Transformer encoders is based on the unique features of each modality. Using different encoders can train the model parameters to better fit the needs of each modality. However, these three Transformer encoders all retain the characteristics of Transformer, such as dot-product attention and multi-head attention mechanism. The Transformer encoder for text and speech modalities retains the position-wise feedforward neural network, while the Transformer encoder for the image modality uses a multilayer perceptron and employs GeLU as its activation function, abandoning the traditional ReLU activation function in traditional Transformers.

Decoder

To better accomplish the tasks of entity recognition and relation extraction in semantic extraction, we used different decoders based on the Transformer encoder and employed different linear layers in the final hidden output state to perform the tasks. For enhanced entity recognition, we incorporated a conditional random field (CRF) decoder in the decoder part. This allows for the better utilization of dependencies among different labels. For a given feature sequence $s = [s_1, s_2, \ldots, s_T]$ and its corresponding gold label sequence $y = [y_1, y_2, \ldots, y_T]$, where $Y(s)$ represents the valid label sequence, the probability value of y can be calculated by the Equation (1):

$$P(y \mid s) = \frac{\sum_{t=1}^{T} e^{f(y_{t-1}, y_t, s)}}{\sum_{y'}^{Y(s)} e^{f(y'_{t-1}, y'_t, s)}} \tag{1}$$

where $f(y_{t-1}, y_t, s)$ calculates the transition score from y_{t-1} to y_t and the score of y_t. The optimization goal is to maximize $P(y|s)$, and the Viterbi algorithm is used to find the path with the maximum probability during decoding.

In relation extraction, the identification of relations can be viewed as a classification problem. Entity relation types are generally mutually exclusive, although there are a few non-mutually exclusive relations, which account for a low proportion and can be artificially decomposed into mutually exclusive relations. Since Softmax is well-suited for handling mutually exclusive multi-classification problems, a Softmax classifier is employed to classify the output generated by the Transformer encoding layer.

3.2.2. Knowledge Extraction from Structured Data

Ship communication data are mostly stored in the structured form, and extracting knowledge from these data manually requires considerable human cost, and expertise and can be error-prone. The mapping between the data source and the domain ontology can be represented as a semantic network, also known as a semantic model, which describes the implicit semantic relations in the data source according to the concepts and relations defined in the domain ontology. The constructed semantic model can be used to automatically transform the data source to RDF triples for publishing to the knowledge graph. In this paper, we apply an automatic semantic modeling algorithm including seed semantic model generation and seed semantic model amending these two steps to obtain the most plausible semantic model [19], and further complete the task of extracting knowledge from structured data.

Seed Semantic Model Generation

For the input ship communication data source, we first find all candidate semantic types for each attribute in the data source, and then generate a candidate semantic model for it by using the Steiner tree algorithm. In summary, there are two sub-steps, that is semantic labeling and relation discovery to obtain the initial seed semantic model. In the semantic labeling phase, we employ the SemanticTyper algorithm proposed by Krishnamurthy et al. [20] to annotate the semantic types for source attributes. Based on the annotated semantic types, we obtain a candidate semantic model by modeling the relations between them using the Steiner tree algorithm [21].

Seed Semantic Model Amending

There are some missed substructures and wrong relations in the obtained seed semantic model after the first step. To improve the quality of the generated semantic model for the input data source, we use TF-IDF cosine similarity and other measurement machine learning methods to distinguish some ambiguous relations by analyzing data source information. Meanwhile, some incorrect substructures in the seed semantic model can be detected by matching model fragments in an existing relevant knowledge graph. After removing incorrect relations and substructures, with the help of the existing relevant knowledge graph, we can obtain the most plausible semantic model by adding potentially missed substructures using the modified frequent subgraph mining algorithm [22].

As shown in Figure 8, we have completed the construction of a multi-modal spatio-temporal knowledge graph based on ship communication through the processing steps of unstructured knowledge extraction based on transformer and structured data knowledge extraction based on automatic semantic modeling.
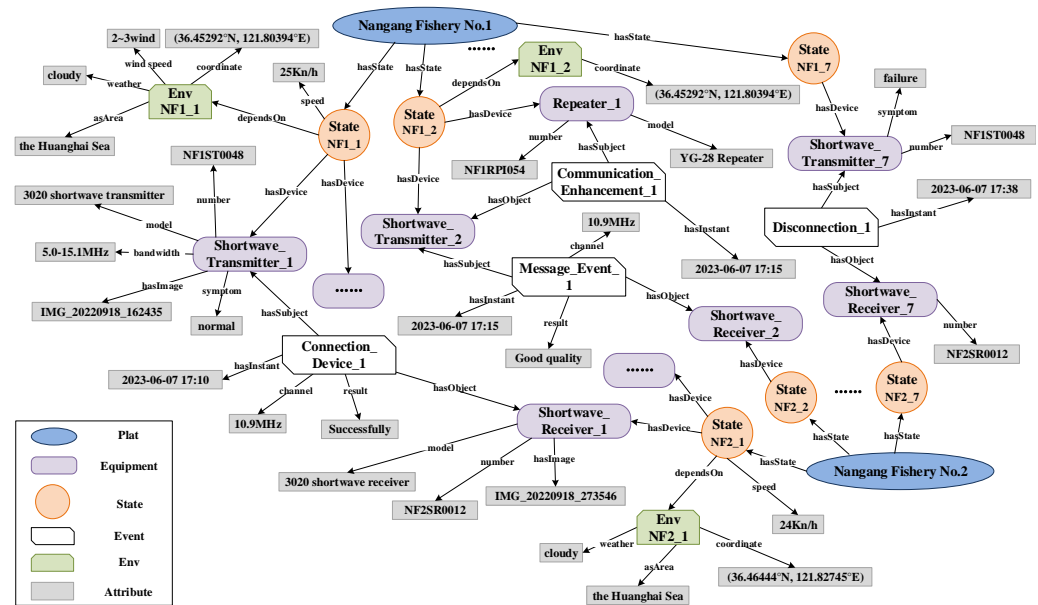
Figure 8. Multi-modal spatio-temporal knowledge graph.

3.3. Multi-Modal Spatio-Temporal Knowledge Graph Storage

The spatio-temporal knowledge graph constructed in this study is stored using Neo4j, a high-performance NoSQL graph database designed for storing structured data on a network rather than in traditional tables. Neo4j exhibits remarkable scalability, enabling the efficient processing of billions of nodes, relationships, and attributes on a single machine. It can also be horizontally scaled across multiple machines to facilitate parallel processing. By leveraging the node-based storage model and establishing relationships between nodes, we are able to construct intricate nested and interconnected unstructured data structures. This approach effectively caters to the storage requirements of multi-level nested spatio-temporal scene data models.

3.3.1. Time Expression Model Based on Neo4j Graph Database

Time expressions in spatio-temporal data models can be categorized into three types: interval-based methods, point-based methods, and time-based methods. Interval-based methods partition time into discrete intervals, which are defined by their relationships, such as 'before' or 'after'. Point-based methods represent time as specific moments when entity objects exist or events occur. In this study, a time-based method that integrates both point-based and interval-based approaches is employed. The interval-based method is expressed using a link table in the graph database, while the point-based method is represented using a timeline tree. By combining these two techniques, a comprehensive time-based method is achieved, enabling the direct inclusion of time as an attribute of node entities based on the domain model.

3.3.2. Spatial Expression Model Based on Neo4j Graph Database

The Neo4j graph database incorporates the Neo4j Spatial extension plugin, which facilitates the representation of spatial data using nodes and relationships. The underlying methodology of the Neo4j Spatial plugin involves the construction of an R-tree, a powerful library that enables Neo4j to perform comprehensive spatial operations. This plugin supports the import of ESRI Shapefile files and OSM data, enabling the representation of diverse geometric shapes, including points, lines, polygons, and more. Additionally, it enables the execution of topological operations, such as containment, coverage, and intersection, on spatio-temporal data.

4. Discussion

Case 1. The ship communication quality prediction task holds significant importance in enhancing the reliability and safety of maritime communication, thereby ensuring the security of maritime activities. Specifically, this task aids ship crews in better route planning and communication strategies, reducing the risks related to communication interruptions or failures and ultimately enhancing communication reliability. In emergency scenarios, ship communication quality prediction tasks can help anticipate the performance of emergency call signal transmissions under different environmental conditions. These predictions enable maritime rescue organizations to effectively determine the timing and direction of rescue operations, thereby improving overall operational efficiency.

The knowledge graph constructed in this paper organizes multi-modal data describing the historical communication tasks of ships from a spatio-temporal perspective and uses communication event entities to represent the communication situations of ships during sea voyages. Specifically, communication events are divided into Data_Transfer_Event, Communication_Enhancement_Event, Disconnection_Event, Connection_Device_Event, Voice_Call_Event, and Message_Event, which include attributes such as occurrence time, communication channel, and communication result. The communication result describes the communication quality of the above events. Therefore, the ship communication quality prediction problem can be modeled as a task of completing missing attribute values of communication events in the multi-modal spatio-temporal knowledge graph.

Suppose a fishing boat named "Nanggang Fishery No. 1" as shown in Figure 9 is fishing in the Huanghai Sea and suddenly encounters severe winds and waves, resulting in significant damage to the vessel. The crew urgently needs to send a distress signal to the nearby fishing boat "Nanggang Fishery No. 2". However, due to the complex maritime environment, the communication result attribute is missing in the communication event of this fishing boat, making it impossible to determine the quality of this communication.

Figure 9. Ship communication quality prediction based on the multi-modal spatio-temporal knowledge graph.

As shown in Figure 9, the blue area describes the communication scenario where "At 8 June 2023 6:14, Nangang Fishery No. 1 encountered cloudy weather with strong winds of level 10–11 in the Huanghai Sea (37.25278° N, 120.46378° E) area and needed to use the NF1ST0048 shortwave transmitter to send rescue information to the nearby Nangang Fishery No. 2 fishing vessel (37.43496° N, 121.0438° E)". In this scenario, it is necessary to predict the communication quality result of this Message_Event to guide the rescue ship to

adjust communication methods and rescue direction and complete the rescue operation in a timely manner.

Through the analysis of historical data, it is evident that environmental factors such as weather conditions, wind speed, and geographic coordinates at a specific time can indeed impact the transmission quality of communication signals. Moreover, considering the extensive coverage of the Huanghai Sea and the high volume of maritime traffic, it is plausible that this fishing vessel may experience interference from other ships during communication. Thus, a deep learning approach based on graph convolutional neural networks can be employed to effectively address the missing communication result attribute for this specific communication event. This approach will enable the prediction of the communication quality, providing valuable insights into the effectiveness of the communication process.

We trained a graph convolutional neural network model on historical data to address missing values in ship communication events. Subsequently, we utilized environmental factors, communication equipment parameters, and other relevant variables at the given time point as features to predict the likelihood of poor communication quality in similar communication events. Such predictions are crucial as they directly impact the transmission quality of emergency signals in comparable environments. Leveraging these results, we can effectively guide Nangang Fishery No. 2 to make timely adjustments to the rescue time and direction, ultimately enhancing the efficiency of rescue operations and ensuring the safety of crew members.

Case 2. Compared to traditional relational databases, graph databases model based on entities and relationships in the real world, and their expression is more intuitive and concise. Graph databases are very suitable for querying and analyzing complex relationships with multiple levels and diversity; relational databases are complex and inefficient in handling complex relationship queries, especially when involving multitable associations or recursive queries. Cypher is a property graph data query language implemented in the graph database Neo4j [23]. Cypher query language provides the basis for data correction, analysis, and expansion for the ship knowledge graph system. The following will introduce in detail the operation content and implementation method of querying ship knowledge based on Cypher language. Cypher query language relies on matching graph patterns. The MATCH keyword is used to specify the search pattern, the WHERE keyword is used in conjunction with the MATCH keyword to add predicate constraints to the matching pattern, and the RETURN keyword is used to return result variables. Below are two examples of querying ship communication knowledge.

Example 1. Query the sailing speed of Nangang Fishery No.1 at 17:10 on 7 June 2023.

```
MATCH (x1:Plat)-[:hasState]->(x2:State)-[:hasDevice]->(:Equipment)
    <-[]-(:Event)-[:hasInstant]->(x3:Instant)
WHERE x1.name = ''Nanggang Fishery No.1'' AND x3.time = ''2023-06-07 17:10''
RETURN x2.speed as Speed
```

Query Result

```
Speed: 25Kn/h
```

Example 2. Query the communication status of Nangang Fishery No.1 and Nangang Fishery No.1 at 17:10 on 7 June 2023.

```
MATCH(x1:Plat)-[*]-(x2:Event)-[*]-(x3:Plat)
WHERE x1.name = ''Nanggang Fishery No.1'' AND x2.time = ''2023-06-07 17:10''
    AND x3.name = ''Nangang Fishery No.2''
RETURN x2.result as Result
```

Query Result

```
Result: Successfully
```

5. Conclusions

This paper proposes an effective approach to integrate ship communication data with spatio-temporal and multi-modal features by constructing a multi-modal spatio-temporal knowledge graph, which provides excellent data management support for subsequent data applications and situational awareness.

To address the heterogeneity among multi-source data, we establish the association between data by constructing a multi-modal spatio-temporal ontology based on existing ontologies, which guides the integration and aggregation of multi-modal spatio-temporal information and facilitates information sharing and reuse. We use different techniques to extract ship communication knowledge and obtain a high-quality multi-modal spatio-temporal knowledge graph for structured and unstructured data. Taking the communication scenario of fishing boats going to sea as an example, we use the constructed multi-modal spatio-temporal knowledge graph and graph convolutional neural network model to predict communication quality, and the query tasks for ship communication knowledge were accomplished through a graph database.

In future work, we intend to further improve the performance of the proposed approach by incorporating more advanced techniques, such as deep learning and natural language processing. Moreover, we will evaluate the proposed approach in real-world scenarios to verify its effectiveness and applicability. Overall, the proposed multi-modal spatio-temporal knowledge graph has significant potential to enhance ship communication management.

Author Contributions: Writing—original draft, Y.Z., R.X. and W.L.; Writing—review & editing, W.M., D.N., Y.D., X.Z. and Z.F. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Xia, T.; Wang, M.M.; Zhang, J.; Wang, L. Maritime internet of things: Challenges and solutions. *IEEE Wirel. Commun.* **2020**, *27*, 188–196. [CrossRef]
2. Scholkopf, B.; Sung, K.K.; Burges, C.J.; Girosi, F.; Niyogi, P.; Poggio, T.; Vapnik, V. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans. Signal Process.* **1997**, *45*, 2758–2765. [CrossRef]
3. Yao, X.; Hu, N.; Zhou, L.; Li, Y. Ore blending of underground mines based on an immune clone selection optimization algorithm. *J. Univ. Sci. Technol. Beijing* **2011**, *33*, 526–531 .
4. Wang, Z.; Zeng, Y.; Wang, J.; Hu, Z. Intrusion Detection Based on Improved BP Neural Network Based on Improved Beetle Swarm Optimization. *Sci. Technol. Eng.* **2020**, *20* , 13249–13257.
5. Jianxiang, L. Method Research on Construction of Ship Activity Graph and Visual Analysis. 2022. Available online: https://kns.cnki.net/kcms2/article/abstract?v=sw50xB5SLWyOcJc9Di9GPF-VnUZLKJZdfWDoXKHkLqDJ6s_h6FZrr7JWtb Zyk2lMv941pfiFpOUa7ylmBz2_90tp4aUfWKKq0RCPntpwR3BjqgQZe8Q81PcLA6kbqCaUvHr49pArKVw=&uniplatform=N ZKPT&language=CHS (accessed on 17 August 2023).
6. Batra, S.; Tyagi, C. Comparative analysis of relational and graph databases. *Int. J. Soft Comput. Eng.* **2012**, *2*, 509–512.
7. Del Mondo, G.; Stell, J.G.; Claramunt, C.; Thibaud, R. A Graph Model for Spatio-temporal Evolution. *J. Univers. Comput. Sci.* **2010**, *16*, 1452–1477.

8.   Ge, X.; Yang, Y.; Peng, L.; Chen, L.; Li, W.; Zhang, W.; Chen, J. Spatio-temporal knowledge graph based forest fire prediction with multi source heterogeneous data. *Remote Sens.* **2022**, *14*, 3496. [CrossRef]

9.   Lu, F.; Yu, L.; Qiu, P. On geographic knowledge graph. *J. Geo Inf. Sci.* **2017**, *19*, 723–734.

10.   Guan, S.; Cheng, X.; Bai, L.; Zhang, F.; Li, Z.; Zeng, Y.; Jin, X.; Guo, J. What is event knowledge graph: A survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *99*, 1–20. [CrossRef]

11.   Liu, P.; Chen, F.; Ma, J.; Zhang, J. Research on Prediction of Link Embedding in Maritime Knowledge Graph. In Proceedings of the 2021 2nd International Conference on Electronics, Communications and Information Technology (CECIT), Sanya, China, 27–29 December 2021 ; IEEE: New York, NY, USA , 2021; pp. 1036–1040.

12.   Everwyn, J.; Mouaddib, A.I.; Zanuttini, B.; Gatepaille, S.; Brunessaux, S. Link Prediction on Dynamic Attributed Knowledge Graphs for Maritime Situational Awareness. In Proceedings of the Conference Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA), Saint-Étienne, France, 30 June–1 July 2019 .

13.   Wen, Y.; Zhang, Y.; Huang, L.; Zhou, C.; Xiao, C.; Zhang, F.; Peng, X.; Zhan, W.; Sui, Z. Semantic modelling of ship behavior in harbor based on ontology and dynamic bayesian network. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 107. [CrossRef]

14.   Ren, H.; Luo, F. Research on the knowledge graph analysis system of ship activity law. *Ship Sci. Technol.* **2022**, *44*, 159–164.

15.   Grüninger, M. Verification of the OWL-time ontology. In Proceedings of the The Semantic Web–ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, 23–27 October 2011; Proceedings, Part I 10; Springer: Berlin/Heidelberg, Germany, 2011; pp. 225–240.

16.   Battle, R.; Kolas, D. Geosparql: Enabling a geospatial semantic web. *Semant. Web J.* **2011**, *3*, 355–370. [CrossRef]

17.   Wang, S.; Zhang, X.; Ye, P.; Du, M.; Lu, Y.; Xue, H. Geographic knowledge graph (GeoKG): A formalized geographic knowledge representation. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 184. [CrossRef]

18.   Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; Association for Computational Linguistics: Toronto, ON, Canada, 2020; pp. 38–45.

19.   Xu, J.; Mayer, W.; Zhang, H.; He, K.; Feng, Z. Automatic Semantic Modeling for Structural Data Source with the Prior Knowledge from Knowledge Base. *Mathematics* **2022**, *10*, 4778. [CrossRef]

20.   Ramnandan, S.K.; Mittal, A.; Knoblock, C.A.; Szekely, P. Assigning semantic labels to data sources. In Proceedings of the Semantic Web, Latest Advances and New Domains: 12th European Semantic Web Conference, ESWC 2015, Portoroz, Slovenia, 31 May–4 June 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 403–417.

21.   Taheriyan, M.; Knoblock, C.A.; Szekely, P.; Ambite, J.L. Learning the semantics of structured data sources. *J. Web Semant.* **2016**, *37*, 152–169. [CrossRef]

22.   Elseidy, M.; Abdelhamid, E.; Skiadopoulos, S.; Kalnis, P. Grami: Frequent subgraph and pattern mining in a single large graph. *Proc. VLDB Endow.* **2014**, *7*, 517–528. [CrossRef]

23.   Francis, N.; Green, A.; Guagliardo, P.; Libkin, L.; Lindaaker, T.; Marsault, V.; Plantikow, S.; Rydberg, M.; Selmer, P.; Taylor, A. Cypher: An evolving query language for property graphs. In Proceedings of the 2018 International Conference on Management of Data, Houston, TX, USA, 10–15 June 2018; pp. 1433–1445.