Advances in Computational Systems, Algorithms, and Emerging Technologies (2022), 7, 1–18

ORIGINALRESEARCH

Knowledge Graph Construction and Link Prediction Using Graph Embedding Techniques: Applications in Recommender Systems

Carlos Eduardo Silva

Universidade Estadual de Mato Grosso do Sul, Faculdade de Computação, 879 Rua Pedro Celestino, Dourados, Mato Grosso do Sul, Brazil

Abstract

Recent advancements in knowledge graph construction and link prediction have significantly transformed the ways in which large-scale relational data are processed, analyzed, and utilized in complex real-world applications. By leveraging graph embedding techniques, it is possible to efficiently learn vector representations of entities and relations in a low-dimensional space, thereby enabling more accurate and scalable methods for inferring missing links and uncovering latent patterns. This approach holds particular relevance in recommender systems, where predicting potential connections among users, items, and contextual factors is critical to delivering precise and personalized suggestions. In this work, we undertake a thorough investigation of knowledge graph construction and link prediction, examining essential building blocks, structured representations, and advanced graph embedding methods that provide deep insights into complex relational data. We also discuss logical consistency requirements and the alignment of symbolic knowledge with high-dimensional numerical representations to ensure robust interpretability. Furthermore, we highlight the emerging trends and outstanding challenges in integrating graph-based recommendations, including scalability, explainability, and adaptability issues. Our analysis not only consolidates fundamental principles but also illustrates contemporary breakthroughs and open avenues for future research. Through this comprehensive exploration, the paper emphasizes how synergy between knowledge graphs and graph embedding techniques can drive next-generation recommender systems to offer unparalleled accuracy and impactful user experiences.

1. Introduction

Knowledge graphs have emerged as powerful tools for representing structured, semi-structured, and unstructured data within a unified framework (Venkatesan and Prabu 2019). They model real-world entities and their relationships in a manner that is both highly interpretable and amenable to machine learning techniques. Underpinning this framework is a focus on capturing heterogeneous, multi-relational data in graph-structured form, whereby nodes often denote entities or concepts, and edges capture various types of relationships. In contemporary applications, knowledge graphs have found extensive use in diverse fields such as recommender systems, semantic search, question answering, and drug discovery. Their adaptability lies in the ability to encode high-level abstractions, logical constraints, and richly interconnected information, thus offering a pathway toward deeper insights and more efficient discovery mechanisms.

The pivot toward leveraging knowledge graphs in large-scale systems is supported by the rapid evolution of computational frameworks, the availability of vast amounts of data from digital platforms, and algorithmic advances in machine learning (Khan and Ghani 2021). Nevertheless, designing effective approaches to extract, transform, and represent heterogeneous data remains a non-trivial challenge. The transformation of multi-modal data sources into a coherent knowledge graph calls for strategies that address entity disambiguation, relation extraction, knowledge consolidation, and data quality verification. Once a knowledge graph is in place, it opens the door to a range of operations aimed at deepening our understanding of the underlying structure. One of the most significant and widely studied operations in this space is link prediction, which seeks to discover missing or potential edges between entities.

Link prediction has been widely explored using both statistical and machine learning-based approaches, with methods ranging from heuristic-based similarity measures to deep learning-driven graph embedding techniques (Kahlon and Singh 2021). Traditional approaches often rely on structural properties of the graph, leveraging measures such as the Jaccard coefficient, Adamic-Adar index, and preferential attachment to infer potential links. These heuristics, while effective in certain domains, exhibit limitations when dealing with large-scale, sparse, or highly heterogeneous graphs. More sophisticated techniques exploit latent representations of nodes and relationships, capturing the underlying semantics of the graph using matrix factorization, random walk models, and deep neural networks. Representation learning in knowledge graphs has gained substantial traction, with models such as TransE, TransH, and RotatE learning low-dimensional vector embeddings that encapsulate relational properties, facilitating efficient link prediction.

Beyond link prediction, knowledge graph completion has emerged as a crucial research area, aiming to infer missing entities or relationships within an incomplete graph (Boutet et al. 2019). The challenge of knowledge graph completion is compounded by the noisy, incomplete, and dynamic nature of real-world data sources. Techniques for completion often extend upon link prediction models, incorporating additional contextual information to enhance inference accuracy. Methods such as graph neural networks (GNNs), knowledge graph embeddings, and reinforcement learning have demonstrated promising results in improving the robustness and generalizability of knowledge graph completion models. GNNs, in particular, leverage message-passing mechanisms that enable nodes to aggregate information from their neighbors, thereby capturing both local and global graph structures. When applied to knowledge graphs, GNNs facilitate inductive reasoning and generalization across unseen entities, making them powerful tools for knowledge graph completion and link prediction. (Kuntoğlu et al. 2021)

The integration of knowledge graphs with machine learning models has led to significant advancements in areas such as recommendation systems, where knowledge-aware models enhance personalization by incorporating structured knowledge about users, items, and contextual attributes. Traditional collaborative filtering techniques, which rely on user-item interaction matrices, suffer from sparsity issues and cold-start problems, making them suboptimal for large-scale recommendation tasks. Knowledge graphs provide a principled approach to addressing these challenges by encoding rich semantic relationships between entities, allowing for more informed and contextually relevant recommendations. Hybrid models combining knowledge graph embeddings with deep learning architectures have been shown to improve recommendation accuracy, capturing both explicit and implicit user preferences in a structured manner.

Similarly, in the domain of semantic search and information retrieval, knowledge graphs play a critical role in augmenting search results with contextually relevant information (Ahmad, Sunberg, and Humbert 2021). Conventional keyword-based search engines are limited in their ability to capture the semantics of user queries, often leading to imprecise or suboptimal results. By leveraging knowledge graphs, search engines can infer the intent behind a query, retrieve conceptually related documents, and provide more meaningful entity-based search results. Semantic search powered by knowledge graphs enhances disambiguation, entity linking, and contextual ranking, thereby improving user experience and information accessibility (Forbus et al. 2007).

The application of knowledge graphs extends beyond information retrieval and recommendations, permeating areas such as biomedical research, fraud detection, and automated reasoning. In biomedical applications, knowledge graphs are employed to model complex biological interactions, drug-target relationships, and disease ontologies, facilitating drug repurposing and precision medicine (Loh and Misselhorn 2019). The ability to integrate heterogeneous biomedical data sources into a structured knowledge graph enables researchers to uncover novel associations, generate testable hypotheses, and accelerate drug discovery pipelines. Techniques such as ontology alignment, knowledge fusion, and probabilistic inference are crucial in ensuring the reliability and interpretability of biomedical knowledge graphs.

Fraud detection is another domain where knowledge graphs have demonstrated remarkable efficacy. Traditional rule-based and machine learning models often struggle to capture complex patterns of fraudulent behavior, particularly in financial transactions and cybersecurity. Knowledge graphs provide a natural representation of transaction networks, allowing for anomaly detection and fraud identification through relational analysis (Bhardwaj, Jain, and Sood 2021). Graph-based fraud detection leverages techniques such as graph clustering, community detection, and subgraph matching to identify suspicious entities and transactions. The incorporation of temporal and dynamic graph modeling further enhances the capability to detect evolving fraud patterns in real-time.

Despite the wide-ranging applications and advantages of knowledge graphs, several challenges persist in their construction, maintenance, and scalability. One of the primary challenges is entity resolution, which involves identifying and merging duplicate entities across different data sources. Variations in entity representations, inconsistencies in naming conventions, and the lack of standardized identifiers contribute to entity resolution complexities. Machine learning-based entity resolution techniques, including embedding-based similarity measures and supervised learning approaches, have been developed to address these challenges (Izzo, Märtens, and Pan 2019). However, achieving high precision and recall in entity resolution remains an open problem, particularly in domains with high data variability.

Another key challenge is knowledge graph reasoning, which involves deriving new facts from existing knowledge using logical inference, probabilistic reasoning, or neural-symbolic methods. Rule-based reasoning techniques, such as first-order logic and description logics, provide a formal foundation for knowledge inference but often suffer from scalability limitations. Probabilistic reasoning models, including Bayesian networks and Markov logic networks, offer a more flexible approach but require careful handling of uncertainty and computational efficiency. Recent advancements in neuro-symbolic AI aim to bridge the gap between symbolic reasoning and deep learning, enabling scalable and interpretable knowledge graph reasoning. (2019 scientific session of the society of american gastrointestinal and endoscopic surgeons (sages), baltimore, maryland, usa, 3-6 april 2019: podium abstracts. 2019)

The dynamic nature of real-world knowledge necessitates continuous updating and evolution of knowledge graphs. As new data sources emerge and existing knowledge changes, maintaining the consistency, accuracy, and completeness of a knowledge graph becomes a significant challenge. Techniques such as incremental graph updates, active learning, and automated knowledge fusion play a vital role in ensuring the long-term viability of knowledge graphs. Moreover, explainability and interpretability in knowledge graph-based models remain crucial considerations, particularly in high-stakes domains such as healthcare and legal reasoning. The development of explainable AI techniques tailored for knowledge graphs, including attention mechanisms, rule extraction, and graph visualization, is an active area of research. (Marcos-Pablos and García-Peñalvo 2018)

A key aspect of knowledge graph research involves benchmarking and evaluation, which requires robust datasets and well-defined metrics. Standard benchmarks such as FB15k, WN18, and YAGO have been widely used for evaluating knowledge graph embeddings and reasoning models. Performance metrics such as mean reciprocal rank (MRR), hit rate, and precision-at-k provide quantitative measures of model effectiveness. The following table presents a comparative analysis of popular knowledge graph embedding models, highlighting their core methodologies and performance characteristics.

Table 1. Comparison of Knowledge Graph Embedding Models

Model	Methodology	Embedding Space	Performance (MRR)
TransE	Translation-based	Euclidean	0.45
TransH	Hyperplane projection	Euclidean	0.47
TransR	Relation-specific projection	Euclidean	0.48
RotatE	Rotational embeddings	Complex	0.49
ConvE	Convolutional neural network	Euclidean	0.50

The future of knowledge graphs lies in the seamless integration with large-scale language models, enabling hybrid approaches that combine structured knowledge with unstructured textual information (Fadeel et al. 2018). Advances in retrieval-augmented generation (RAG), knowledge-enhanced pretraining, and multimodal knowledge fusion hold immense potential for pushing the boundaries of knowledge representation and reasoning. As knowledge graphs continue to evolve, their role in artificial intelligence, data science, and scientific discovery will remain indispensable, paving the way for more intelligent and interpretable AI systems.

Link prediction is of paramount importance in recommender systems, where predicting the relevance of an item to a user depends on uncovering latent relationships. By integrating graph embedding techniques, it is possible to map the nodes of a knowledge graph into a high-dimensional embedding space, in which geometric or algebraic properties correspond to relational patterns that might be difficult to capture using traditional methods. A well-designed embedding can capture subtle signals such as shared features, context-specific interactions, and transitivity in relationships (Lee, Park, and Lim 2019). These learned representations can then be employed by machine learning models to infer new edges or rank potential connections, thereby facilitating tasks like item recommendation, social network analysis, and anomaly detection.

Although the integration of knowledge graphs and graph embedding methods offers a powerful framework, several hurdles must be addressed. First, the complexity of knowledge graphs can escalate quickly with the addition of new entities, attributes, and relations, thus posing scalability challenges. Second, the interpretability of learned embeddings is crucial, particularly in sensitive domains like healthcare and finance. Interpretability can be further complicated by the black-box nature of certain deep learning models (Saini and Singh 2021). Third, reconciling logical constraints, consistency requirements, and symbolic representations with continuous embedding spaces is an ongoing challenge that calls for advanced methods in representation learning. In addition, ethical concerns regarding privacy and fairness arise when integrating personal user data into large-scale knowledge graphs. These issues necessitate careful policy considerations and the deployment of responsible artificial intelligence strategies.

The primary objective of this work is to present a comprehensive analysis of knowledge graph construction and link prediction, focusing on the synergy that arises when graph embedding techniques are systematically employed. We provide an overview of fundamental concepts, explore various approaches to embedding, and showcase practical applications in recommender systems (Giacobbe et al. 2018). We also propose logical and algebraic formulations that enhance the ability to capture intricate relational dependencies. By doing so, we aim to illuminate new directions for research, while acknowledging the theoretical and pragmatic challenges that await resolution. Ultimately, this paper underscores the transformative potential of knowledge graphs in driving next-generation recommendation engines and encourages the pursuit of more expressive, scalable, and interpretable embedding strategies.

2. Foundations of Knowledge Graph Construction

In constructing a knowledge graph, the first step is often the extraction of conceptual entities from raw data, which can range from unstructured text to semi-structured databases or fully structured relational tables. Each identified entity is uniquely represented within the graph, typically through an identifier, label, or canonical form. Relations are then specified, linking the entities in a manner consistent with the nature of their interactions (Tao et al. 2019). Importantly, many practical systems rely on schemas or ontologies that dictate the types of entities and relations allowed. These schemas serve as higher-level logical frameworks, ensuring consistency and interpretability throughout the construction process.

Formally, consider a knowledge graph to be a set $\mathcal{G} = (E, R, A)$, where $E = \{e_1, e_2, \ldots, e_n\}$ represents the set of entities, $R = \{r_1, r_2, \ldots, r_m\}$ denotes the set of relations, and $A \subseteq E \times R \times E$ specifies the set of triplets capturing the actual links among entities. In many cases, an entity $e \in E$ can itself be structured, containing attributes and sub-entities. At a higher level, to establish constraints and maintain logical consistency, we often incorporate a set Ω of axioms or rules that govern permissible configurations within the knowledge graph.

(Šećerov et al. 2019) **Entity Extraction:** The process of extracting entities typically involves one or more natural language processing pipelines, which may include named entity recognition, coreference resolution, and entity normalization. For instance, from a corpus of textual descriptions, terms like "smartphone," "camera," and "wireless networking" might be identified as salient entities if they correspond to relevant concepts in the domain. More advanced pipelines incorporate contextual embeddings that help disambiguate entities with overlapping lexical forms, reducing errors during graph construction. In some settings, specialized domain knowledge or external resources, such as ontologies, can be utilized to refine entity extraction (Sharma and Forbus 2010b). (A and B 2019)

Relation Extraction: Relation extraction entails identifying and classifying the associations that link entities. This task is inherently more complex than entity extraction, as it necessitates understanding the context in which the entities co-occur. Traditional approaches rely on rule-based or statistical techniques that parse the syntactic and semantic structure of sentences. More recently, neural network-based models have shown significant promise, with architectures like convolutional neural networks or transformers achieving high accuracy by capturing contextual cues. Commonly, relations are encoded as labeled edges connecting the relevant entities (Swamidason et al. 2020). For instance, one might encounter triplets such as $(\text{UserA}, \text{purchased}, \text{ItemB})$ or $(\text{ProteinX}, \text{interactsWith}, \text{ProteinY})$.

Ontology Alignment and Data Fusion: In many real-world scenarios, data fusion from multiple sources is required to build a more comprehensive knowledge graph. Heterogeneous data may come from different domains, each with its own ontology or schema. Ontology alignment techniques aim to harmonize these schemas by identifying equivalent classes, properties, or instances across knowledge repositories. This step is crucial to avoid redundant or contradictory representations and to ensure that the knowledge graph remains consistent (Priya and Dodagoudar 2018). Mathematically, this involves constructing a mapping ϕ between two sets of concepts C_1 and C_2, such that $\phi : C_1 \to C_2$ preserves the semantic relations between concepts. Identifying such a mapping may rely on lexical similarity measures, machine learning classifiers trained on labeled pairs of classes, or more advanced embedding-based approaches where semantically similar concepts are placed in close proximity within a vector space.

Graph Augmentation and Enrichment: Once an initial knowledge graph has been built, it is often beneficial to augment or enrich it. One potential avenue is to introduce new types of nodes or edges based on secondary data sources, such as user behavioral logs, sensor data, or domain-specific repositories (Cruz et al. 2021). Another is the use of rule-based or statistical inference engines to deduce implicit facts from explicitly stored ones. For example, if an ontology indicates that "camera" is a subcategory of "electronics," and the knowledge graph captures that a user purchased "camera," then a reasoner can infer that the user also purchased an "electronic device." This relational inference process can be formalized by means of forward-chaining or backward-chaining algorithms that traverse the graph, applying logical rules to produce new triplets.

Quality Assessment and Consistency Checking: Maintaining high-quality data in a knowledge graph is non-trivial, as inaccuracies can adversely affect downstream applications. A typical consistency check involves verifying that the triplets in the graph do not violate a set of constraints or logical rules (Das et al. 2019). For instance, if a schema states that the relation "purchased" must link a user entity to a product entity, a triplet associating a user with another user via "purchased" would constitute an inconsistency. Another strategy is to analyze the connectivity patterns of the graph, such as detecting disconnected or sparsely connected components that might signify missing data. These checks may be formalized through constraint satisfaction problem formulations, ensuring that the final knowledge graph is both coherent and accurate.

Taken as a whole, constructing a knowledge graph thus involves a multifaceted pipeline that must account for extraction, alignment, quality control, and ongoing refinement. These steps provide the substrate upon which higher-level tasks, such as link prediction, can be performed (Mukherjee et al. 2020). The structured representation embedded in a knowledge graph is crucial because it offers an interpretable and flexible mechanism for capturing relationships between entities—a property that underpins link prediction algorithms.

3. Graph Embedding Techniques for Link Prediction

Once a robust knowledge graph has been created, the challenge shifts to exploiting this representation to reveal missing links or to forecast the formation of new connections. Graph embedding techniques have proven exceptionally effective in addressing these needs, as they compress the complex network structure into a low-dimensional space, preserving essential relational properties. These learned embeddings can then be employed by machine learning models to classify, rank, or regress the likelihood of edges, effectively yielding link predictions. In what follows, we delve into the fundamental theory behind graph embeddings and explore a range of state-of-the-art methods.

3.1 Foundational Concepts in Graph Embedding

Graph embedding methods generally seek a function $f : E \cup R \to R^d$, mapping each entity and relation to a d-dimensional vector. Link prediction is then performed by defining a scoring function $\sigma : R^d \times R^d \to R$ (or a suitable extension that includes relations) that measures the plausibility of a triplet (e_i, r_k, e_j). In simpler formulations, the scoring function might evaluate the distance between the sum of e_i and r_k and the embedding of e_j, thereby encouraging consistent triplets to be closely aligned. (Djemame et al. 2018)

One notable approach is to construct adjacency or incidence matrices and then apply dimensionality reduction methods, such as matrix factorization or spectral decomposition. Alternatively, neural network-based techniques rely on stochastic gradient descent to iteratively refine embedding vectors. In all these strategies, the overarching objective is to preserve relevant proximity relationships. This can be encoded by a margin-based loss or binary cross-entropy loss that distinguishes valid triplets from negative samples.

Graph embedding techniques can broadly be categorized into translational models, tensor factorization methods, and deep learning-based approaches (He et al. 2019). Translational models, such as TransE, model relationships as vector translations in the embedding space, enforcing constraints such that $f(e_i) + f(r_k) \approx f(e_j)$. These methods, while computationally efficient, struggle with complex relational patterns such as one-to-many or many-to-many relationships. To address these issues, extensions like TransH and TransR introduce hyperplane projections and relation-specific transformation matrices, respectively. Tensor factorization methods, including RESCAL and TuckER, decompose the knowledge graph adjacency tensor into low-rank components, capturing high-order interactions among entities and relations.

More recent deep learning-based approaches leverage convolutional neural networks (ConvE), recurrent neural networks, or graph neural networks (GNNs) to extract richer representations from knowledge graphs (Gerós, Magalhães, and Aguiar 2020). ConvE, for example, reshapes embeddings into 2D matrices and applies convolutional filters to capture spatial dependencies. GNNs, on the other hand, perform message passing over the graph structure, enabling nodes to aggregate information from their neighbors, thereby learning context-aware embeddings.

A critical aspect of embedding-based link prediction is the choice of loss functions. Margin-based ranking loss, commonly used in translational models, is defined as follows:

$$\mathcal{L} = \sum_{(e_i,r_k,e_j)\in\mathcal{T}} \sum_{(e_i',r_k,e_j')\in\mathcal{T}'} \max(0, \gamma + \sigma(e_i, r_k, e_j) - \sigma(e_i', r_k, e_j'))$$

where $\mathcal{T}$ represents the set of valid triplets, $\mathcal{T}'$ represents corrupted triplets, and γ is a margin hyperparameter. This formulation ensures that the plausibility score of a true triplet is higher than that of a negative triplet by at least γ. (Yang et al. 2018)

Alternatively, binary cross-entropy loss is frequently employed in deep learning-based models, where predictions are treated as classification problems. Given a scoring function $\sigma(e_i, r_k, e_j)$ that outputs a probability of a triplet being valid, the loss is computed as:

$$\mathcal{L} = -\sum_{(e_i,r_k,e_j)\in\mathcal{D}} y_{ijk} \log(\sigma(e_i, r_k, e_j)) + (1 - y_{ijk}) \log(1 - \sigma(e_i, r_k, e_j))$$

where y_{ijk} denotes the ground truth label (1 for valid triplets, 0 for negative samples). This loss function effectively trains the model to distinguish between valid and invalid triplets based on observed data.

The quality of learned embeddings is heavily dependent on the negative sampling strategy. In knowledge graphs, negative samples are typically generated by randomly corrupting entities in triplets (i.e., replacing e_i or e_j with a random entity) (Bouteiller and Charlety 2019). However, uniform random sampling often results in unrealistic or uninformative negatives. Hard negative sampling, which prioritizes triplets that closely resemble true facts, has been shown to significantly improve model performance. This can be achieved using adversarial sampling, where the most confusing negative samples are selected dynamically during training.

Another crucial factor influencing knowledge graph embeddings is the choice of embedding space and distance metric. While Euclidean embeddings are commonly used, alternative representations in hyperbolic and complex spaces have demonstrated improved performance, particularly in hierarchical or multi-relational graphs (Bakheet and Al-Hamadi 2020). Hyperbolic embeddings, for instance, are well-suited for capturing hierarchical structures due to their exponential growth properties, enabling more efficient representation of taxonomic relationships. Complex-valued embeddings, as used in models like ComplEx and RotatE, introduce additional degrees of freedom by modeling relations as rotations in the complex plane, thereby capturing asymmetric and compositional relationships more effectively.

To compare different embedding methods, a set of standard benchmark datasets is widely used in the research community. These datasets include WordNet (WN18), Freebase (FB15k), and YAGO, each presenting unique challenges such as sparse connectivity, large-scale multi-relational links, and entity disambiguation. The following table provides a comparison of key properties of these datasets: (Elmoulat et al. 2021)

Table 2. Comparison of Benchmark Knowledge Graph Datasets

Dataset	# Entities	# Relations	# Triplets
WN18	40,943	18	151,442
FB15k	14,951	1,345	592,213
YAGO3-10	123,182	37	1,079,040

Despite the significant progress in graph embedding techniques, several challenges remain. One of the key limitations is the trade-off between expressiveness and scalability. While high-dimensional embeddings can capture intricate relationships, they require extensive computational resources and are prone to overfitting. Low-dimensional representations, on the other hand, often struggle to encode complex relational structures. Techniques such as knowledge distillation, where a smaller model is trained to mimic a larger model's embeddings, have been proposed to address this issue.

Another challenge lies in dynamic knowledge graphs, where entities and relationships continuously evolve over time (Blezek et al. 2021). Most existing embedding methods assume a static graph structure, making them inadequate for applications requiring temporal reasoning. Recent advances in temporal knowledge graph embeddings, such as time-aware extensions of TransE and GNN-based temporal models, aim to address this gap by incorporating timestamps into the embedding process (Sharma and Forbus 2010a).

Incorporating external knowledge sources is another frontier in knowledge graph embeddings. While most methods rely purely on structural information, augmenting embeddings with textual descriptions, ontological constraints, or multimodal data (e.g., images, audio) can significantly enhance their utility. Pretrained language models, such as BERT and GPT, have been increasingly integrated with knowledge graphs to improve entity representations by leveraging contextual semantics. (Kumar and Rao 2018)

As knowledge graph embedding methods continue to evolve, their applications are expanding beyond traditional link prediction tasks. Emerging areas such as explainable AI, fairness in knowledge graphs, and neuro-symbolic reasoning are gaining traction. The development of interpretable embeddings that provide human-understandable justifications for predictions is particularly important in domains such as healthcare and legal reasoning. Similarly, addressing biases in knowledge graphs, which often arise due to imbalanced data distributions, remains a critical research challenge.

Overall, knowledge graph embeddings represent a foundational component in modern AI systems, enabling enhanced reasoning, retrieval, and decision-making capabilities (Allgeier et al. 2018). With continued advancements in representation learning, graph-based neural architectures, and hybrid AI paradigms, the future of knowledge graph embeddings promises even greater scalability, interpretability, and generalization to real-world applications.

3.2 Translational Models

A seminal line of research in graph embedding for link prediction introduces "translational" models. For instance, TransE, one of the earliest proposals, represents each relation as a vector "translation" that operates on an entity vector. Specifically, the plausible triplet (e_i, r_k, e_j) suggests

$$\mathbf{e}_i + \mathbf{r}_k \approx \mathbf{e}_j,$$

where $\mathbf{e}_i, \mathbf{r}_k, \mathbf{e}_j \in R^d$. The goal is to minimize a distance metric, often the L1 or L2 norm, between $\mathbf{e}_i + \mathbf{r}_k$ and $\mathbf{e}_j$. This simple yet powerful notion has been further refined in subsequent models, including TransH, TransR, and TransD, all aiming to handle complex relations such as 1-to-N, N-to-1, and N-to-N by mapping entities and relations into various latent subspaces. (Dang et al. 2021)

For example, TransR introduces a relation-specific projection matrix $\mathbf{W}_k \in R^{d\times d}$ that transforms entity embeddings into a relation-specific space:

$$\mathbf{e}'_i = \mathbf{W}_k\mathbf{e}_i, \quad \mathbf{e}'_j = \mathbf{W}_k\mathbf{e}_j.$$

Hence, the translational property manifests as

$$\mathbf{e}'_i + \mathbf{r}_k \approx \mathbf{e}'_j,$$

thereby allowing different relations to specialize their embeddings. Such extensions are particularly beneficial in domains featuring diverse relation patterns.

3.3 Neural Network-Based Embeddings

Another family of techniques employs deep neural networks to capture higher-order and non-linear dependencies. Methods such as DistMult, ComplEx, and RotatE rely on embedding vectors (or complex-valued embeddings) combined with specialized scoring functions (Alotaibi, Asghar, and Ahmad 2021). DistMult scores a triplet (e_i, r_k, e_j) by computing

$$\sigma(e_i, r_k, e_j) = \mathbf{e}_i^\top \operatorname{diag}(\mathbf{r}_k)\, \mathbf{e}_j,$$

where $\operatorname{diag}(\mathbf{r}_k)$ is a diagonal matrix formed from $\mathbf{r}_k$. Meanwhile, ComplEx extends this concept into the complex domain, enabling the capture of asymmetric relations by using the Hermitian dot product. RotatE interprets each relation as a rotation in the complex plane, leading to a formulation:

$$\mathbf{e}_j = \mathbf{e}_i \circ \mathbf{r}_k,$$

where $\circ$ denotes element-wise (Hadamard) multiplication in the complex space.

Neural architectures often incorporate negative sampling strategies (Fardet, Quaresima, and Bottani 2019). For each valid triplet, several "corrupted" triplets are generated by replacing entities or relations with random alternatives. These corrupted triplets represent negative samples, motivating the network to distinguish plausible connections from spurious ones. The final model parameters are then learned via gradient-based optimization, typically employing Adam or related methods.

3.4 Graph Convolutional and Attention-Based Methods

Recent progress in graph neural networks (GNNs) has spurred new approaches for link prediction. These architectures propagate information along edges, using techniques such as convolution or attention to aggregate neighborhood features (Mahabadi and Besmi 2020). A fundamental building block in these methods is the aggregation function, which integrates information from neighboring nodes to update the embedding of a target node. A generalized mechanism might be written as:

$$\mathbf{h}_i^{(l+1)} = \phi\left(\mathbf{W}^{(l)} \cdot \mathrm{AGG}\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\}\right),$$

where $\mathbf{h}_i^{(l)}$ is the embedding of node i at layer l, $\mathrm{AGG}\{\cdot\}$ is an aggregation function (e.g., sum or mean), and ϕ is a non-linear activation. Graph attention networks refine this mechanism by introducing attention coefficients that weigh the contribution of each neighbor separately.

After a stack of such layers, the embeddings obtained are then used in a link prediction task. In knowledge graph scenarios, specialized GNN frameworks handle distinct relation types by parameterizing the message passing logic with relation-specific weights. Combining these network architectures with multi-relational or meta-path-based strategies has been shown to be highly effective, particularly in large heterogeneous networks. (Koohi-Var and Zahedi 2018)

3.5 Optimization and Regularization

Regardless of the specific embedding algorithm, controlling model complexity and overfitting remains an important concern. Techniques such as L_2 regularization, dropout, and norm constraints on embeddings are frequently applied. Some methods leverage parameter sharing across relations that are semantically related, reducing the dimensionality of the parameter space. Additionally, hyperparameter tuning, including the choice of embedding dimension d, learning rate, and the number of negative samples per positive triplet, significantly impacts model performance. In practice, cross-validation on a validation subset of links is used to refine these hyperparameters (Li, Du, and Shen 2020). Early stopping criteria that monitor performance metrics, such as mean rank or hits-at-k, are commonly employed to prevent overfitting (Abhishek and Rajaraman 2005). In summary, the evolving landscape of graph embedding techniques supports a rich array of possibilities for capturing multi-relational data and performing link prediction. By integrating translational models, deep neural embeddings, and graph neural architectures, researchers and practitioners can exploit increasingly nuanced signal in knowledge graphs. The expressiveness of such techniques is a driving force behind the success of link prediction, especially in modern recommender system applications that require flexible modeling of heterogeneous data and user-item interactions.

4. Applications in Recommender Systems

Recommender systems serve as a prime beneficiary of knowledge graph and link prediction research (Majumdar, Mitra, and Bhattacharya 2021). Traditional collaborative filtering or content-based recommender systems rely on user-item ratings or features, but often fail to account for the complex network of relationships among users, items, and contextual information. By contrast, knowledge graphs extend this foundation through multi-relational data modeling, unearthing rich contextual and structural cues. Consequently, combining knowledge graphs with advanced graph embedding methods offers a powerful approach for generating highly accurate and personalized recommendations.

4.1 Integrating Multi-Relational Data

In a typical e-commerce context, knowledge graphs might incorporate users, items, categories, tags, reviews, brand affiliations, social connections, and so forth. Each of these elements is linked by one or more relation types, forming a multi-relational network (Ansari et al. 2019). For instance, consider a fragment of a knowledge graph that includes triplets:

$$(\text{User}_1, \text{purchased}, \text{Item}_A), \quad (\text{Item}_A, \text{hasCategory}, \text{Category}_C), \quad (\text{User}_1, \text{followed}, \text{User}_2).$$

Such relationships open a path to discover new items relevant to User_1 by not only examining purchase history but also leveraging the contextual information that Item_A belongs to a specific category or that User_2 has similar preferences.

The utility of multi-relational data is further realized when user-specific and item-specific attributes are integrated. For instance, embedding approaches can encode user demographics and item metadata, while capturing social interactions that may influence user preferences. In essence, knowledge graphs transform recommendation from a matrix-completion perspective into a rich link prediction scenario, wherein the recommendation is the anticipation of future or unobserved edges between users and items.

4.2 Graph Embedding as a Recommendation Backbone

Advanced embeddings provide a natural backbone for recommender systems by encoding latent factors (Jabeen, Gao, and Andreae 2019). In the simplest approach, one obtains entity and relation embeddings via a technique like TransE or DistMult, and then uses a suitable scoring function to predict the likelihood that a user connects to an item. One might adopt a triple scoring function $\sigma(u, r, i)$, where u denotes a user entity, i denotes an item entity, and r denotes the "purchased" or "rated" relation. Potential recommendations can be ranked in descending order of $\sigma(u, r, i)$.

Extensions of this concept use GNN-based embeddings that account for the local neighborhood of a user, their social links, or their historical interactions. After a GNN processes the user's immediate neighbors to produce a refined embedding, a scoring function is again applied to evaluate item compatibility (Mantovan and Nanni 2020). Such strategies can incorporate user-user similarity, item-item relationships, or meta-path structures, providing a more holistic perspective than conventional collaborative filtering. Formally, for a user node u, a GNN layer updates its embedding $\mathbf{h}_u$ based on neighbors $\mathcal{N}(u)$, which may include items and other users. This iterative update can be represented as

$$\mathbf{h}_u^{(l+1)} = \phi\left(\sum_{v\in\mathcal{N}(u)} \alpha_{u,v}^{(l)}\mathbf{h}_v^{(l)}\right),$$

where $\alpha_{u,v}^{(l)}$ are learned attention coefficients. The final embedding $\mathbf{h}_u^{(L)}$ then facilitates link predictions with candidate items.

4.3 Cold-Start and Sparsity Problems

Traditional recommender systems often struggle with cold-start issues, where new users or items have minimal interaction data, as well as sparsity problems arising when the ratio of observed interactions to possible interactions is low. Knowledge graphs partially mitigate these challenges by enriching each entity with additional relational structure. Instead of relying exclusively on user ratings or item popularity, a knowledge graph can integrate side information (e.g., textual descriptions, attribute relationships, brand connections, or user demographics) (Clement et al. 2019). Link prediction algorithms can use these auxiliary links to extrapolate latent preferences even when direct user–item interaction data is sparse.

The advantage of knowledge graph embeddings in this context is the capacity to exploit secondary relations to learn robust representations, effectively acting as a smoothing mechanism. A new user with minimal purchase history might still be associated with their geographic location, age range, or declared interests, all of which provide signals that can guide the recommendation process. Similarly, a new item can be associated with a product category, brand, or similar items. Consequently, the knowledge graph compensates for the lack of direct user–item interactions, reducing the detrimental impact of data sparsity. (Gaudrie et al. 2020)

4.4 Explainability and Interpretability

Explainability is crucial in recommender systems, particularly in domains such as healthcare or finance where trust and accountability are paramount. Knowledge graphs lend themselves to better interpretability since they intrinsically structure relationships between entities. When using a knowledge graph-based model, one can trace the "path" of influence leading to a recommendation. For instance, if a user is recommended Item_X, the system may highlight that the user follows someone who purchased an item in the same category as Item_X. Alternatively, it might show that multiple items in the user's purchase history share attributes with Item_X. In more formal terms, interpretability can be aided by path-based reasoning methods that find high-scoring meta-paths from the user to the recommended item within the knowledge graph. The presence of a coherent path offers a tangible explanation, grounded in the underlying relationships.

4.5 Scalability Considerations

Despite these benefits, deploying knowledge graph-based recommender systems in large-scale industrial contexts raises significant scalability concerns (Braik, Al-Zoubi, and Al-Hiary 2020). Real-world recommendation platforms often involve tens or hundreds of millions of users and items, resulting in an immense quantity of entities and relations. Distributed architectures and efficient sampling strategies become indispensable for both graph construction and embedding training. Approximate neighbor search or partitioning techniques may be employed to reduce computational overhead. Moreover, incremental or streaming updates to the graph are often needed, particularly in dynamic environments where user-item interactions unfold in real time. Approaches that can rapidly update the embeddings, or that leverage incremental learning, hold promise in such scenarios. (Hariri and Narin 2021)

4.6 Empirical Evidence in Industrial Use Cases

A growing body of work documents the impact of knowledge graph-based methods in diverse industrial applications. Global e-commerce platforms have reported substantial improvements in click-through rate and conversion by employing graph embeddings that integrate product attributes, user demographics, and social influences. Streaming services for music and video have incorporated knowledge graphs to capture genre, artist, and user preference networks, facilitating cross-domain recommendations. Preliminary investigations in ride-sharing platforms suggest the viability of knowledge graphs to integrate demand patterns, geographic data, and user feedback, although the scope of such systems is still evolving (Basu et al. 2006). Overall, the synergy between knowledge graphs and link prediction through graph embeddings provides a robust framework to tackle the complexities of modern recommender systems (Visvikis et al. 2019). By seamlessly integrating heterogeneous, multi-relational data, these methodologies address cold-start and sparsity issues, enhance interpretability, and exhibit scalability with appropriate architectural choices. The next section delves into the challenges and future directions that may guide continued research and application in this rapidly developing field.

5. Challenges and Future Directions

While knowledge graphs and graph embedding techniques hold great potential for powering advanced link prediction mechanisms, several key challenges and research gaps remain to be addressed. By systematically examining these challenges, we can outline promising avenues for future investigation.

5.1 Data Quality and Trustworthiness

One immediate concern is the quality and trustworthiness of data within knowledge graphs (Bao et al. 2021). Link prediction is fundamentally reliant on the integrity of the underlying relations; inaccuracies or ambiguities in graph construction can undermine even the most sophisticated embedding models. Automated extraction of entities and relations from unstructured sources often introduces noise, leading to erroneous links or incomplete coverage of relevant facts. In critical domains such as healthcare or finance, even minor inaccuracies can lead to significant consequences. Research efforts focusing on robust extraction pipelines, uncertainty modeling, and outlier detection can help mitigate these issues. For instance, probabilistic knowledge graph representations may introduce confidence scores for each edge, enabling a more nuanced approach to link prediction that accounts for data reliability. (Koopman et al. 2020)

5.2 *Context-Aware Embeddings*

Many practical scenarios demand context-aware embeddings that adapt to temporal, geospatial, or situational variations. For example, in recommender systems, user preferences may shift over time or differ depending on the user's location. Similarly, scientific knowledge graphs may need to capture evolving findings and relationships. Incorporating temporal or spatial dimensions into knowledge graph embeddings can be accomplished by extending the embedding function $f : E \cup R \rightarrow R^d$ to include additional parameters for time or location. One approach introduces time-specific embeddings, such that an entity or relation has different vectors for different time segments. Another line of research involves spatio-temporal GNNs that capture changes in both space and time (Chen et al. 2020). Beyond just modeling dynamics, enabling real-time updates to embeddings as new data arrive is pivotal in many industrial contexts.

5.3 *Human-in-the-Loop Systems*

Despite progress in automated graph construction and link prediction, human expertise often plays a pivotal role in validating or refining inferences. Domain experts, crowd-sourcing platforms, or specialized curators may intervene to approve or dispute automatically generated links. This human-in-the-loop paradigm can significantly enhance the reliability of knowledge graphs. It can also introduce feedback loops that continuously improve model accuracy, as embeddings are retrained or fine-tuned based on human inputs (Gonçalves et al. 2018). Future research directions include techniques for optimizing where and when to solicit human feedback, how to effectively incorporate domain knowledge, and how to design user interfaces that allow experts to interact with complex knowledge graphs.

5.4 *Interpretable and Explainable Models*

As knowledge graph-based systems gain traction in domains with strict accountability requirements, the need for interpretable and explainable models becomes increasingly pressing. While knowledge graphs naturally lend themselves to partial interpretability, black-box embedding models may obscure how certain link predictions are generated. Ongoing research aims to develop methods to clarify the roles of specific dimensions in embedding spaces, identify influential relations, or reconstruct logical pathways. For instance, neural logic programming frameworks seek to fuse neural embeddings with explicit logical reasoning, offering textual or rule-based explanations for link predictions. Another trend involves representational methods that maintain human-readable annotations alongside learned embeddings, enabling domain experts to interrogate and validate model outputs. (Casagrande et al. 2020)

5.5 *Fairness and Privacy Concerns*

Fairness and privacy are emerging imperatives in recommender systems and broader AI applications. Knowledge graphs, particularly those constructed from personal user data or sensitive domains, raise concerns about potential biases and unauthorized inferences. Biased embeddings may systematically disadvantage certain user groups, while privacy leaks can occur if hidden attributes are inferred through link prediction. Addressing fairness requires implementing constraints or regularization strategies that equalize representation across demographic segments or mitigate undesirable correlations. Privacy-aware methods aim to obfuscate or perturb data to safeguard user identity without severely degrading model performance (Mir et al. 2018). Techniques like federated learning may enable partial training of embeddings on user devices, limiting data transfer to central servers. A rigorous approach to these concerns involves formulating constraints in logic or symbolic form that ensure compliance with domain regulations or ethical guidelines.

5.6 Multi-Modal and Cross-Domain Integration

Modern knowledge graphs often must integrate multi-modal data sources, such as images, audio clips, or sensor readings, as well as cross-domain information. A single entity in one domain may correspond to an entity in another domain under a different ontology or naming convention. Handling such integrative tasks demands new embedding paradigms that are robust to cross-lingual or cross-modal inconsistencies (Mohammed et al. 2021). Strategies might include joint embedding models that unify textual, visual, and relational features. Alternatively, gating mechanisms within GNNs can dynamically select the most relevant modality for each link prediction task.

5.7 Benchmarking and Standardization

While various benchmarks exist for tasks like link prediction (e.g., FB15k, WN18, YAGO datasets), there remains a gap in standardized evaluations that replicate real-world complexities of large-scale, dynamic, and often noisy knowledge graphs. Developing more comprehensive benchmarks that reflect diverse domains and realistic data conditions will support progress in the field and enhance comparability. Furthermore, standardized metrics that go beyond ranking accuracy—encompassing interpretability, fairness, and real-time adaptability—are needed to more fully capture the capabilities of modern graph embedding techniques. (Edwards et al. 2021)

5.8 Novel Algorithmic Paradigms

Finally, there is continued interest in pushing beyond existing paradigms. Approaches that combine symbolic reasoning and neural embeddings are especially promising, as they can reconcile the interpretability of rule-based systems with the flexibility and robustness of deep networks. Additionally, quantum-inspired or hyperbolic embeddings show early promise in capturing hierarchical or complex relational patterns. Hierarchical embeddings may represent entity classes with hyperbolic geometries that preserve transitive and multi-level relationships more naturally. Meanwhile, emergent paradigms like neural-symbolic integration aim to fuse logical constraints directly into embedding objectives, thus ensuring that learned representations remain consistent with domain-specific rules. (AlZubi, Al-Maitah, and Alarifi 2021)

In summation, the future of knowledge graph construction and link prediction hinges on addressing issues related to data quality, contextual and temporal dynamics, interpretability, fairness, and scalability. These challenges highlight the importance of interdisciplinary collaboration among experts in machine learning, graph theory, domain-specific areas, and policy to realize the full potential of these technologies. As research in graph embeddings matures, it stands poised to underpin a new generation of intelligent systems capable of nuanced reasoning and robust predictive performance.

6. Conclusion

In this paper, we have examined the processes and methodologies central to knowledge graph construction and link prediction, highlighting the role of graph embedding techniques as a powerful tool for inference and discovery. Knowledge graphs present a flexible and richly structured means of representing multi-relational data drawn from heterogeneous sources, offering a robust substrate for tasks such as entity disambiguation and relation extraction (Monaca et al. 2019). Once built, these knowledge graphs can be leveraged for link prediction, enabling the automatic inference of unobserved connections in domains ranging from social networks and e-commerce platforms to scientific research.

A critical component in this endeavor is the development of effective graph embedding methods, which compress high-dimensional, intricate graph structures into low-dimensional vector spaces while preserving essential relational properties. We explored translational models, neural network-based methods, and graph neural network architectures, elucidating how these techniques can yield state-of-the-art performance on link prediction benchmarks. Moreover, we highlighted how recommender systems, in particular, benefit substantially from link prediction in knowledge graphs, as they can harness additional relational cues to mitigate cold-start and sparsity issues and provide interpretable suggestions to end-users.

Despite these advancements, numerous challenges remain unresolved (Sufang 2020) (Sharma, Witbrock, and Goolsbey 2016). The quality and completeness of knowledge graphs remain fundamental concerns, as do algorithmic scalability and interpretability. Additionally, ensuring fairness and privacy preservation is paramount when recommendations draw on sensitive user data. Ongoing research aims to address these issues by introducing context-aware, time-evolving embeddings, more refined human-in-the-loop validation processes, and novel approaches that balance symbolic reasoning with deep learning. New paradigms for multi-modal integration, cross-domain fusion, and standardization of benchmarks also continue to emerge. By confronting these challenges and capitalizing on future developments, knowledge graph-based link prediction systems promise to transform the landscape of intelligent applications. Recommender systems are emblematic of the broader potential of these technologies, demonstrating the value of comprehensive, graph-based representations in delivering enhanced personalization, scalability, and interpretability across a multitude of real-world settings. (Mosca et al. 2021)

References

A, Vijayalakshmi, and Rajesh Kanna B. 2019. Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications* 79, no. 21 (January 11, 2019): 15297–15317. https://doi.org/10.1007/s11042-019-7162-y.

Abhishek and V Rajaraman. 2005. A computer aided shorthand expander. *IETE Technical Review* 22 (4): 267–272.

Ahmad, Shakeeb, Zachary N. Sunberg, and J. Sean Humbert. 2021. End-to-end probabilistic depth perception and 3d obstacle avoidance using pomdp. *Journal of Intelligent & Robotic Systems* 103, no. 2 (September 17, 2021): 1–18. https://doi.org/10.1007/s10846-021-01489-w.

Allgeier, Stephan, Andreas Bartschat, Sebastian Bohn, Sabine Peschel, Klaus-Martin Reichert, Karsten Sperlich, Marcus Walckling, et al. 2018. 3d confocal laser-scanning microscopy for large-area imaging of the corneal subbasal nerve plexus. *Scientific reports* 8, no. 1 (May 10, 2018): 7468–7468. https://doi.org/10.1038/s41598-018-25915-6.

Alotaibi, Fahad, Muhammad Zubair Asghar, and Shakeel Ahmad. 2021. A hybrid cnn-lstm model for psychopathic class detection from tweeter users. *Cognitive Computation* 13, no. 3 (March 10, 2021): 709–723. https://doi.org/10.1007/s12559-021-09836-7.

AlZubi, Ahmad Ali, Mohammed Al-Maitah, and Abdulaziz Alarifi. 2021. Cyber-attack detection in healthcare using cyber-physical system and machine learning techniques. *Soft Computing* 25, no. 18 (June 26, 2021): 12319–12332. https://doi.org/10.1007/s00500-021-05926-8.

Ansari, Mohd Yousuf, Amir Ahmad, Shehroz S. Khan, Gopal Bhushan, and null Mainuddin. 2019. Spatiotemporal clustering: a review. *Artificial Intelligence Review* 53, no. 4 (July 15, 2019): 2381–2423. https://doi.org/10.1007/s10462-019-09736-1.

Bakheet, Samy, and Ayoub Al-Hamadi. 2020. Chord-length shape features for license plate character recognition. *Journal of Russian Laser Research* 41, no. 2 (March 26, 2020): 156–170. https://doi.org/10.1007/s10946-020-09861-1.

Bao, Qiangwei, Gang Zhao, Yong Yu, Sheng Dai, and Wei Wang. 2021. The ontology-based modeling and evolution of digital twin for assembly workshop. *The International Journal of Advanced Manufacturing Technology* 117, nos. 1-2 (July 29, 2021): 395–411. https://doi.org/10.1007/s00170-021-07773-1.

Basu, Anupam, et al. 2006. Iconic interfaces for assistive communication. In *Encyclopedia of human computer interaction,* 295–302. IGI Global.

Bhardwaj, Charu, Shruti Jain, and Meenakshi Sood. 2021. Transfer learning based robust automatic detection system for diabetic retinopathy grading. *Neural Computing and Applications* 33, no. 20 (May 7, 2021): 13999–14019. https://doi.org/10.1007/s00521-021-06042-2.

Blezek, Daniel J., Lonny Olson-Williams, Andrew D. Missert, and Pangiotis Korfiatis. 2021. Ai integration in the clinical workflow. *Journal of digital imaging* 34, no. 6 (October 22, 2021): 1435–1446. https://doi.org/10.1007/s10278-021-00525-3.

Bouteiller, Pauline Le, and Jean Charlety. 2019. Semi-supervised multi-facies object retrieval in seismic data. *Mathematical Geosciences* 52, no. 6 (September 17, 2019): 817–846. https://doi.org/10.1007/s11004-019-09822-8.

Boutet, Alexandre, Robert Gramer, Christopher J. Steele, Gavin J B Elias, Jürgen Germann, Ricardo Maciel, Walter Kucharczyk, Ludvic Zrinzo, Andres M. Lozano, and Alfonso Fasano. 2019. Neuroimaging technological advancements for targeting in functional neurosurgery. *Current neurology and neuroscience reports* 19, no. 7 (May 30, 2019): 42–42. https://doi.org/10.1007/s11910-019-0961-8.

Braik, Malik, Hussein Al-Zoubi, and Heba Al-Hiary. 2020. Artificial neural networks training via bio-inspired optimisation algorithms: modelling industrial winding process, case study. *Soft Computing* 25, no. 6 (November 23, 2020): 4545–4569. https://doi.org/10.1007/s00500-020-05464-9.

Casagrande, Luan Carlos, Luiz Antonio Buschetto Macarini, Daniel Bitencourt, Antônio Augusto Fröhlich, and Gustavo Medeiros de Araújo. 2020. A new feature extraction process based on sfta and dwt to enhance classification of ceramic tiles quality. *Machine Vision and Applications* 31, no. 7 (September 24, 2020): 1–15. https://doi.org/10.1007/s00138-020-01121-1.

Chen, Xuanwei, Wei Chen, Liang Hou, Huosheng Hu, Bu Xiangjian, and Qingyuan Zhu. 2020. A novel data-driven rollover risk assessment for articulated steering vehicles using rnn. *Journal of Mechanical Science and Technology* 34, no. 5 (April 30, 2020): 2161–2170. https://doi.org/10.1007/s12206-020-0437-4.

Clement, Patricia, Marco Castellaro, Thomas W. Okell, David L. Thomas, C. Gorgolewski, Stefan Appelhoff, Jan Petr, Michael A. Chappell, and Henk J M M Mutsaerts. 2019. Asl-bids, the brain imaging data structure extension for arterial spin labeling. *Magma (New York, N.Y.)* 32, no. Suppl 1 (September 24, 2019): 107–233. https://doi.org/10.1007/s10334-019-00754-2.

Cruz, Francisco, Richard Dazeley, Peter Vamplew, and Ithan Moreira. 2021. Explainable robotic systems: understanding goal-driven actions in a reinforcement learning scenario. *Neural Computing and Applications* 35, no. 25 (August 28, 2021): 1–18. https://doi.org/10.1007/s00521-021-06425-5.

Dang, Fu-Rong, Jintao Tang, Kunyuan Pang, Ting Wang, Shasha Li, and Xiao Li. 2021. Constructing an educational knowledge graph with concepts linked to wikipedia. *Journal of Computer Science and Technology* 36, no. 5 (September 30, 2021): 1200–1211. https://doi.org/10.1007/s11390-020-0328-2.

Das, Amita, Priti Das, Sourav Panda, and Sukanta Sabut. 2019. Detection of liver cancer using modified fuzzy clustering and decision tree classifier in ct images. *Pattern Recognition and Image Analysis* 29, no. 2 (June 17, 2019): 201–211. https://doi.org/10.1134/s1054661819020056.

Djemame, Safia, Mohamed Batouche, Hamouche Oulhadj, and Patrick Siarry. 2018. Solving reverse emergence with quantum pso application to image processing. *Soft Computing* 23, no. 16 (June 28, 2018): 6921–6935. https://doi.org/10.1007/s00500-018-3331-6.

Edwards, Luke J., Kerrin Pine, Gunther Helms, and Nikolaus Weiskopf. 2021. Rational approximation of the ernst equation for dual angle r1 mapping revisited: beyond the small flip-angle assumption. *Magma (New York, N.Y.)* 34, no. Suppl 1 (September 18, 2021): 45–46. https://doi.org/10.1007/s10334-021-00947-8.

Elmoulat, Meryem, Lahcen Ait Brahim, Abderrahman Elmahsani, A. Abdelouafi, and Mohammed Mastere. 2021. Mass movements susceptibility mapping by using heuristic approach. case study: province of tétouan (north of morocco). *Geoenvironmental Disasters* 8, no. 1 (August 23, 2021): 1–19. https://doi.org/10.1186/s40677-021-00192-0.

Fadeel, Bengt, Lucian Farcal, Barry Hardy, Socorro Vázquez-Campos, Danail Hristozov, Antonio Marcomini, Iseult Lynch, Eugenia Valsami-Jones, Harri Alenius, and Kai Savolainen. 2018. Advanced tools for the safety assessment of nanomaterials. *Nature nanotechnology* 13, no. 7 (July 6, 2018): 537–543. https://doi.org/10.1038/s41565-018-0185-0.

Fardet, T, A Quaresima, and S Bottani. 2019. Dense: modeling neuronal morphology and network structure in silico. *BMC neuroscience* 20, no. Suppl 1 (November 14, 2019): 22–95. https://doi.org/10.1186/s12868-019-0538-0.

Forbus, Kenneth D, Christopher Riesbeck, Lawrence Birnbaum, Kevin Livingston, Abhishek Sharma, and Leo Ureel. 2007. Integrating natural language, knowledge representation and reasoning, and analogical processing to learn by reading. In *Proceedings of the national conference on artificial intelligence,* 22:1542. 2. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Gaudrie, David, Rodolphe Le Riche, Victor Picheny, Benoit Enaux, and Vincent Herbert. 2020. Modeling and optimization with gaussian processes in reduced eigenbases. *Structural and Multidisciplinary Optimization* 61, no. 6 (February 29, 2020): 2343–2361. https://doi.org/10.1007/s00158-019-02458-6.

Gerós, Ana, Ana Magalhães, and Paulo Aguiar. 2020. Improved 3d tracking and automated classification of rodents' behavioral activity using depth-sensing cameras. *Behavior research methods* 52, no. 5 (March 30, 2020): 2156–2167. https://doi.org/10.3758/s13428-020-01381-9.

Giacobbe, Maurizio, Giuseppe Pellegrino, Marco Scarpa, and Antonio Puliafito. 2018. An approach to implement the "smart office" idea: the smartme energy system. *Journal of Ambient Intelligence and Humanized Computing* 14, no. 12 (May 8, 2018): 1–19. https://doi.org/10.1007/s12652-018-0809-0.

Gonçalves, Enyo José Tavares, Marcos de Oliveira, Ingrid Teixeira Monteiro, Jaelson Castro, and João Araújo. 2018. Understanding what is important in istar extension proposals: the viewpoint of researchers. *Requirements Engineering* 24, no. 1 (July 20, 2018): 55–84. https://doi.org/10.1007/s00766-018-0302-5.

Hariri, Walid, and Ali Narin. 2021. Deep neural networks for covid-19 detection and diagnosis using images and acoustic-based techniques: a recent review. *Soft computing* 25, no. 24 (August 24, 2021): 1–18. https://doi.org/10.1007/s00500-021-06137-x.

He, Kangli, Holger Hermanns, Hengyang Wu, and Yixiang Chen. 2019. Connection models for the internet-of-things. *Frontiers of Computer Science* 14, no. 3 (December 7, 2019): 143401–. https://doi.org/10.1007/s11704-018-7395-3.

Izzo, Dario, Marcus Märtens, and Binfeng Pan. 2019. A survey on artificial intelligence trends in spacecraft guidance dynamics and control. *Astrodynamics* 3, no. 4 (July 31, 2019): 287–299. https://doi.org/10.1007/s42064-018-0053-6.

Jabeen, Shahida, Xiaoying Gao, and Peter Andreae. 2019. Semantic association computation: a comprehensive survey. *Artificial Intelligence Review* 53, no. 6 (November 20, 2019): 3849–3899. https://doi.org/10.1007/s10462-019-09781-w.

Kahlon, Navroz Kaur, and Williamjeet Singh. 2021. Machine translation from text to sign language: a systematic review. *Universal Access in the Information Society* 22, no. 1 (July 3, 2021): 1–35. https://doi.org/10.1007/s10209-021-00823-1.

Khan, Nida Saddaf, and Muhammad Sayeed Ghani. 2021. A survey of deep learning based models for human activity recognition. *Wireless Personal Communications* 120, no. 2 (May 7, 2021): 1593–1635. https://doi.org/10.1007/s11277-021-08525-w.

Koohi-Var, Tahereh, and Morteza Zahedi. 2018. Cross-domain similarity assessment for workflow improvement to handle big data challenge in workflow management. *Journal of Big Data* 5, no. 1 (July 23, 2018): 1–20. https://doi.org/10.1186/s40537-018-0135-6.

Koopman, Mandy, Quentin Peter, Renée I. Seinstra, Michele Perni, Michele Vendruscolo, Christopher M. Dobson, Tuomas P. J. Knowles, and Ellen A. A. Nollen. 2020. Assessing motor-related phenotypes of caenorhabditis elegans with the wide field-of-view nematode tracking platform. *Nature protocols* 15, no. 6 (May 20, 2020): 2071–2106. https://doi.org/10.1038/s41596-020-0321-9.

Kumar, M. Ravi, and Y. Srinivasa Rao. 2018. Epileptic seizures classification in eeg signal based on semantic features and variational mode decomposition. *Cluster Computing* 22, no. 6 (February 16, 2018): 13521–13531. https://doi.org/10.1007/s10586-018-1995-4.

Kuntoğlu, Mustafa, Emin Salur, Munish Kumar Gupta, Murat Sarikaya, and Danil Yu. Pimenov. 2021. A state-of-the-art review on sensors and signal processing systems in mechanical machining processes. *The International Journal of Advanced Manufacturing Technology* 116, no. 9 (July 6, 2021): 2711–2735. https://doi.org/10.1007/s00170-021-07425-4.

Lee, Chang Joo, Sang Kyoo Park, and Myo Taeg Lim. 2019. Multi-target tracking and track management algorithm based on ufir filter with imperfect detection probability. *International Journal of Control, Automation and Systems* 17, no. 12 (September 23, 2019): 3021–3034. https://doi.org/10.1007/s12555-018-0439-5.

Li, Deng, Xin Du, and Ji-zhong Shen. 2020. Web page classification based on heterogeneous features and a combination of multiple classifiers. *Frontiers of Information Technology & Electronic Engineering* 21, no. 7 (July 29, 2020): 995–1004. https://doi.org/10.1631/fitee.1900240.

Loh, Wulf, and Catrin Misselhorn. 2019. Augmented learning, smart glasses and knowing how. *AI & SOCIETY* 35, no. 2 (March 18, 2019): 297–308. https://doi.org/10.1007/s00146-019-00881-3.

Mahabadi, Aminollah, and Mohammad Reza Besmi. 2020. Risk-aware service level agreement modeling in smart grid. *Multimedia Tools and Applications* 80, no. 1 (September 8, 2020): 1433–1456. https://doi.org/10.1007/s11042-020-09787-5.

Majumdar, Parijata, Sanjoy Mitra, and Diptendu Bhattacharya. 2021. Iot for promoting agriculture 4.0: a review from the perspective of weather monitoring, yield prediction, security of wsn protocols, and hardware cost analysis. *Journal of Biosystems Engineering* 46, no. 4 (November 25, 2021): 440–461. https://doi.org/10.1007/s42853-021-00118-6.

Mantovan, Lorenzo, and Loris Nanni. 2020. The computerization of archaeology: survey on artificial intelligence techniques. *SN Computer Science* 1, no. 5 (August 14, 2020): 1–32. https://doi.org/10.1007/s42979-020-00286-w.

Marcos-Pablos, Samuel, and Francisco José García-Peñalvo. 2018. Information retrieval methodology for aiding scientific database search. *Soft Computing* 24, no. 8 (October 12, 2018): 5551–5560. https://doi.org/10.1007/s00500-018-3568-0.

Mir, Junaid, Dumidu S. Talagala, Anil Fernando, and Hemantha Kodikara Arachchi. 2018. A comprehensive study and performance evaluation of hdr video coding. *Arabian Journal for Science and Engineering* 44, no. 3 (October 8, 2018): 2427–2444. https://doi.org/10.1007/s13369-018-3583-6.

Mohammed, Abdul Hanan Khan, Hrag-Harout Jebamikyous, Dina Nawara, and Rasha Kashef. 2021. Iot text analytics in smart education and beyond. *Journal of Computing in Higher Education* 33, no. 3 (August 31, 2021): 779–806. https://doi.org/10.1007/s12528-021-09295-x.

Monaca, Ubaldo la, Serena Bertagna, Alberto Marino, and Vittorio Bucci. 2019. Integrated ship design: an innovative methodological approach enabled by new generation computer tools. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 14, no. 1 (September 26, 2019): 59–76. https://doi.org/10.1007/s12008-019-00612-4.

Mosca, Sara, Claudia Conti, Nicholas Stone, and Pavel Matousek. 2021. Spatially offset raman spectroscopy. *Nature Reviews Methods Primers* 1, no. 1 (March 11, 2021): 1–16. https://doi.org/10.1038/s43586-021-00019-0.

Mukherjee, Anwesha, Shreya Ghosh, Aabhas Behere, Soumya K. Ghosh, and Rajkumar Buyya. 2020. Internet of health things (ioht) for personalized health care using integrated edge-fog-cloud network. *Journal of Ambient Intelligence and Humanized Computing* 12, no. 1 (June 8, 2020): 943–959. https://doi.org/10.1007/s12652-020-02113-9.

Priya, B. Divya, and G. R. Dodagoudar. 2018. An integrated geotechnical database and gis for 3d subsurface modelling: application to chennai city, india. *Applied Geomatics* 10, no. 1 (February 5, 2018): 47–64. https://doi.org/10.1007/s12518-018-0202-x.

Saini, Aradhya, and Dharmendra Singh. 2021. Dronertef:development of a novel adaptive framework for railroad track extraction in drone images. *Pattern Analysis and Applications* 24, no. 4 (June 29, 2021): 1549–1568. https://doi.org/10.1007/s10044-021-00994-w.

Šećerov, Ivan, Stevan Savić, Dragan Milošević, Daniela Arsenović, Dragan Dolinaj, and Srdjan Popov. 2019. Progressing urban climate research using a high-density monitoring network system. *Environmental monitoring and assessment* 191, no. 2 (January 21, 2019): 89–. https://doi.org/10.1007/s10661-019-7210-0.

Sharma, Abhishek, and Kenneth D Forbus. 2010a. Graph-based reasoning and reinforcement learning for improving q/a performance in large knowledge-based systems. In *2010 aaai fall symposium series.*

———. 2010b. Modeling the evolution of knowledge and reasoning in learning systems. In *2010 aaai fall symposium series.*

Sharma, Abhishek, Michael Witbrock, and Keith Goolsbey. 2016. Controlling search in very large commonsense knowledge bases: a machine learning approach. *arXiv preprint arXiv:1603.04402.*

Sufang, Wang. 2020. An adaptive ensemble classification framework for real-time data streams by distributed control systems. *Neural Computing and Applications* 32, no. 9 (February 24, 2020): 4139–4149. https://doi.org/10.1007/s00521-020-04759-0.

2019 scientific session of the society of american gastrointestinal and endoscopic surgeons (sages), baltimore, maryland, usa, 3-6 april 2019: podium abstracts. 2019. *Surgical endoscopy* 33, no. Suppl 1 (February 27, 2019): 1–77. https://doi.org/10.1007/s00464-019-06703-3.

Swamidason, Iwin Thanakumar Joseph, Sravanthi Tatiparthi, V. M. Arul Xavier, and C. S. C. Devadass. 2020. Exploration of diverse intelligent approaches in speech recognition systems. *International Journal of Speech Technology* 26, no. 1 (November 25, 2020): 1–10. https://doi.org/10.1007/s10772-020-09769-w.

Tao, Jianhua, Jian Huang, Ya Li, Zheng Lian, and Mingyue Niu. 2019. Semi-supervised ladder networks for speech emotion recognition. *International Journal of Automation and Computing* 16, no. 4 (May 2, 2019): 437–448. https://doi.org/10.1007/s11633-019-1175-x.

Venkatesan, R., and S. Prabu. 2019. Hyperspectral image features classification using deep learning recurrent neural networks. *Journal of medical systems* 43, no. 7 (June 4, 2019): 216–216. https://doi.org/10.1007/s10916-019-1347-9.

Visvikis, Dimitris, Catherine Cheze Le Rest, Vincent Jaouen, and Mathieu Hatt. 2019. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *European journal of nuclear medicine and molecular imaging* 46, no. 13 (July 6, 2019): 2630–2637. https://doi.org/10.1007/s00259-019-04373-w.

Yang, Xi, Zhuo Song, Chengkun Wu, Wei Wang, Li, Wei Zhang, Lingqian Wu, and Kai Lu. 2018. Constructing a database for the relations between cnv and human genetic diseases via systematic text mining. *BMC bioinformatics* 19, no. 19 (December 31, 2018): 528–528. https://doi.org/10.1186/s12859-018-2526-2.