International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Domain Knowledge Integration and Distillation using text Mining

Shashi Pal Singh[1], Ritu Tiwari[2], Sanjeev Sharma[3], Aastha Porwa[4], Shrishti Choudhary[5]

Submitted: 15/05/2024 **Revised**: 28/06/2024 **Accepted**: 07/07/2024

Abstract: Imagine a media organization striving to combat the spread of misinformation during an election season. The ability to synthesize and simplify diverse knowledge sources becomes crucial for accurately predicting and identifying fake news. In today's dynamic world, this need extends across all disciplines. This paper explores two fundamental processes essential for achieving such synthesis: domain knowledge integration, which harmonizes information from various sources, and distillation, which extracts essential insights. Through a comprehensive literature review and detailed case studies, we examine the methodologies, challenges, and benefits associated with these processes. Embracing domain knowledge integration and distillation enables organizations to streamline operations, enhance strategic planning, and gain a competitive edge. The results reveal significant improvements in fake news detection accuracy, streamlined operations, and increased stakeholder confidence.

Moreover, the application of these processes in real-world scenarios underscores their practical value. By leveraging these methodologies, media organizations can better allocate resources, respond more quickly to emerging threats, and maintain public trust. This research thus provides a robust framework for other sectors facing similar challenges, highlighting the universal applicability of these techniques. The insights gained from this study are poised to influence future strategies in combating misinformation, promoting a more informed and resilient society. This research significantly contributes to understanding these methodologies, ensuring originality and unwavering integrity in its findings.

Keywords: *Integration ,Distillation , Data Preprocessing , NLP, Confusion Matrix , Classification*.

1. Introduction

In today's rapidly evolving world, blending and simplifying different types of knowledge are essential for making smart decisions and solving problems in many areas. Combining information from various places (like books, studies, or experts) is called domain knowledge integration. After that, distillation helps pull out the most important ideas from this mix.These methods are like important frameworks for managing big amounts of information and making organizations work better. Domain knowledge integration and distillation have become indispensable tools for facilitating informed decision-making and problem-solving across a multitude of disciplines. The effective utilization of domain-specific knowledge is paramount in navigating the complexities of modern challenges. Domain knowledge integration involves the assimilation of information from various sources, while distillation processes aim to extract actionable insights from this consolidated knowledge pool. These methodologies serve as crucial frameworks for managing the ever-increasing volumes of information and enhancing organizational efficiency.

Domain knowledge integration involves the comprehensive assimilation of data, expertise, and insights from diverse sources such as research papers, industry reports, and expert opinions.This approach empowers organizations to cultivate a comprehensive grasp of intricate matters and craft judicious decisions.Following integration, distillation processes are employed to extract essential insights and trends from the integrated knowledge pool. These insights are crucial for guiding strategic initiatives, driving innovation, and solving complex problems effectively.

The topic aims to bridge the gap between raw data and intelligent behavior. It does this by incorporating and leveraging domain knowledge into AI models. Imagine domain knowledge as the specialized understanding, rules, and insights specific to a particular field, gleaned from expert opinions, scientific literature, or real-world observations. Domain Knowledge Integration and Distillation integrates this knowledge into the design, training, or inference process of AI models.

This document embarks on an exploration of the principles,methodologies and real world applications underlying domain knowledge integration and distillation, recognizing their pivotal roles in driving innovation and improving operational effectiveness. Through a comprehensive review of existing literature, methodologies, and real-world applications, our aim is to furnish readers with a nuanced understanding of these concepts. By delving into the intricacies of domain

[1] *AAIG, Center for Development of Advanced Computing (C-DAC), Pune. India. shashis@cdac.in*
[2]*Indian Institute of Information Technology (IIIT) Pune, India. ritu@iiitp.ac.in*
[3] *Indian Institute of Information Technology (IIIT) Pune, India. sanjeevsharma@iiitp.ac.in*
[4]*Banasthali Vidyapith, Jaipur, India. aasthaporwal6@gmail.com*
[5]*Banasthali Vidyapith, Jaipur, India. srianshi17@gmail.com*

knowledge integration and distillation, organizations can gain valuable insights into optimizing their decision-making processes and fostering a culture of continuous improvement.

Through an examination of diverse methodologies and practical case studies, this research endeavor seeks to shed light on the significance of domain knowledge integration and distillation in contemporary organizational settings. By elucidating the methodologies, challenges, and potential benefits associated with these processes, we aim to equip stakeholders with the requisite knowledge and tools to navigate the complexities of modern decision-making. Ultimately, by incorporating domain knowledge integration and distillation principles, organizations can streamline processes, deepen comprehension, and secure a competitive advantage within their industries.

Ultimately, by embracing domain knowledge integration and distillation, organizations can streamline operations, improve strategic planning, and gain a competitive edge in today's dynamic business environment. As we embark on this exploration, we remain committed to advancing knowledge and contributing meaningfully to the ongoing discourse surrounding these critical methodologies.

Fig. 1. Domain Knowledge in Machine Learning

In summary, Domain Knowledge Integration and Distillation are vital processes that empower individuals and organizations to harness the richness of domain-specific information, creating a foundation for innovation, problem-solving, and continuous improvement.

2. Literature Review

Domain knowledge integration and distillation are pivotal methodologies in contemporary organizational settings, facilitating informed decision-making and problem-solving across various domains. Existing research provides valuable insights into the methodologies, challenges, and benefits associated with these processes.

Author	Title	Description
LinWang,Kuk-Jin Yoon	Knowledge Distillation and Student-Teacher Learning for Visual Intelligence: A Review and New Outlooks	The techniques for transferring knowledge from larger teacher models to smaller student models to enhance efficiency in visual intelligence tasks
Mertins and Jochem(2006) Integrating knowledge	Management and enterprise resource planning	The importance of integrating knowledge from diverse domains within organizations to support effective decision-making processes
Lin and Lee (2010)	Domain Knowledge Driven Intelligent Data Auditing Model	Integrates data mining technology with audit domain knowledge to create an intelligent data auditing mode
Weigang Lu,Ziyu Guan,Yaming Yang(2024)	AdaGMLP: AdaBoosting GNN-to-MLP Knowledge Distillation	Proposes AdaGMLP (AdaBoosting GNN-to-MLP Knowledge Distillation) to address challenges such as insufficient training data and incomplete test data.
Qing Xu,Min Wu ,Xiaoli Li and Kezhi Mao, Zhenghua Chen	Distilling Universal and Joint Knowledge for Cross-Domain Model Compression on Time Series Data	Proposes AdaGMLP (AdaBoosting GNN-to-MLP Knowledge Distillation) to address challenges such as insufficient training data and incomplete test data. Discusses integrating UDA (unsupervised domain adaptation) techniques with KD (knowledge distillation) for cross-domain model compression
Himel Das Gupta,Victor S. Sheng(2022)	A Roadmap to Domain Knowledge Integration in Machine Learning	Highlights the limitations of machine learning in representing human knowledge and emphasizes the importance of domain knowledge

M. F. Mridha(2021)	A Comprehensive Review on Fake News Detection With Natural Language Programming	It examines the effectiveness of different models and approaches, highlighting their strengths and limitations in tackling misinformation

3. Methodology

The methodology for domain knowledge integration and distillation involves a series of essential steps designed to synthesize and extract actionable insights from a variety of knowledge sources. This comprehensive approach encompasses the following key components:

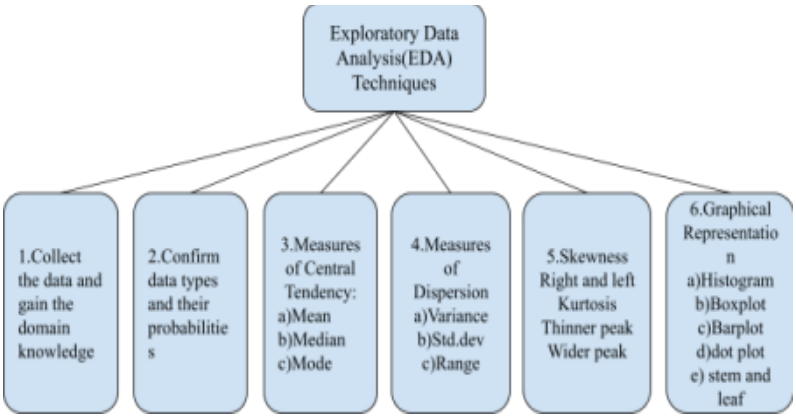
Problem Definition: Clearly define the problem or objective that the integration and distillation process aims to address. This step will give understanding to the specific requirements, desired outcomes, and constraints of the problem domain.

Data Collection and Acquisition: The first step involves gathering relevant data and information from various sources, including research papers, industry reports, expert opinions, internal organizational data, social media, and public datasets. It's crucial to ensure data quality, relevance, and diversity to capture a comprehensive understanding of the problem.

Data Preprocessing: Before integration and distillation, data preprocessing is often necessary to clean, transform, and standardize the collected data. This step includes handling missing values, outlier detection, normalization, and encoding categorical variables. Data preprocessing enhances data quality and consistency, thus improving the accuracy and reliability of subsequent analyses.It includes data visualization, tokenization , stemming, lemmatization and stop words.,

- Data visualization: It is a graphical representation of data in pictorial format such as graphs ,charts and tables.
- Tokenization: Tokenization enables machines to comprehend and process text effectively while retaining its original meaning. By breaking down text into tokens, algorithms gain the capability to recognize patterns more efficiently. This ability is essential as it empowers machines to interpret and respond to human input. For example, when presented with the word "running", a machine doesn't view it as a single unit but rather as a series of tokens that it can analyze and extract significance from.
- Stemming: It is essential to standardize the text by getting rid of repetitions and using stemming to return words to their basic form in order to create a strong model.Stemming is used in word analysis, tagging systems, indexing, text mining SEOs, and information retrieval.
- Lemmatization: Lemmatization is the process of combining all of a word's inflected forms into one group so that they can be examined as a single unit. Lemmatization, as opposed to stemming, gives words context by concatenating words with related meanings into a single word. Preprocessing text involves lemmatization as well as stemming.
- Stopwords: In many applications, stopwords are essential because they help us concentrate on the most important words in a language by eliminating words that are frequently used. For example, if the search word in the context of a search engine is "how to develop information retrieval applications," the terms "how" and "to" would be often used in the English language, thus the search engine would return a lot of pages containing them. The search engine may locate pages that contain keywords like "develop," "information," "retrieval," and "applications" by removing these frequently used terms, which will provide more relevant results. This explains the rationale for the use of stop words.

Exploratory Data Analysis (EDA): EDA techniques are employed to gain initial insights into the data, identify patterns, and detect outliers. Exploratory visualizations, statistical analyses, and data profiling provide a deeper understanding of the dataset's characteristics, informing subsequent integration and distillation strategies.



Feature Engineering: Feature engineering encompasses the creation of new features or the transformation of existing ones with the aim of enhancing model performance. This process may involve employing dimensionality reduction methods like principal component analysis

(PCA), utilizing feature selection algorithms, and crafting domain-specific features to encapsulate pertinent information.

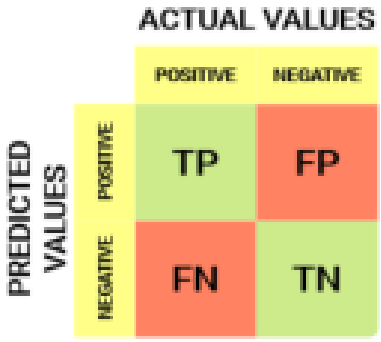
Model Selection and Training: (For distillation tasks involving machine learning or statistical modeling, careful selection of appropriate models is essential. This includes evaluating different algorithms, tuning hyperparameters, and training models on relevant subsets of the data. Ensemble methods and meta-learning approaches can also be considered to enhance model robustness and generalization.

Libraries Used:

- **Numpy:** Numpy library provides a rich set of mathematical functions and efficiently handles multidimensional arrays and matrices.
- **Pandas:** It offers high-performance, easy-to-use data structures for tabular data analysis.
- **Nltk:** It is used for text processing and natural language analysis that enables tasks like tokenization, stemming, lemmatization, named entity recognition, and sentiment analysis.
- **Sklearn:** Scikit-learn library features various classification, regression, and clustering algorithms, including support vector machines, random forests and k-means.
- **Pickle:** Pickle library is used for object serialization ,deserialization and limited data type support.

Distillation Techniques: Once the integrated knowledge framework is established, attention shifts to distillation techniques aimed at extracting essential insights and trends. Various analytical methods, such as statistical analysis, machine learning algorithms (e.gdecision trees, random forests, neural networks), natural language processing techniques, and visualization tools, are employed to identify patterns, correlations, and outliers within the integrated knowledge pool.

Confusion Matrix: A Confusion matrix, which consists of an N x N grid, acts as a crucial tool for evaluating the performance of a classification model by comparing actual target values with its predictions. It offers a detailed examination of the model's efficacy and discerns the various types of errors it generates.



Using Confusion Matrix we can calculate:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Count Vectorizer: CountVectorizer is a method used in natural language processing (NLP) for converting a collection of text documents into a matrix of token counts. It's a technique commonly employed in the preprocessing step before feeding text data into machine learning algorithms.

TF- IDF: TF-IDF is a numerical statistic that reflects the importance of a word in a document.The TF-IDF score is

calculated by multiplying the term frequency and the inverse document frequency. The higher the TF-IDF score, the more important the term is in the

document.

TF-IDF = TF * IDF

TF-IDF = TF * log(N/DF)

Validation and Evaluation: To ensure the reliability of the integrated models, thorough validation procedures are employed, incorporating techniques like cross-validation, holdout validation, and performance metrics such as accuracy, precision, recall and F1-score. Expert reviews, user feedback, and A/B testing further complement these evaluations, providing insight into the models' effectiveness in real-world applications.

Iterative Refinement: Continuous refinement and optimization of the integrated models are performed based on validation results and stakeholder feedback. This iterative process may involve incorporating additional domain knowledge, adjusting model parameters, or updating the integration and distillation pipelines to further improve recommendation performance and usability.

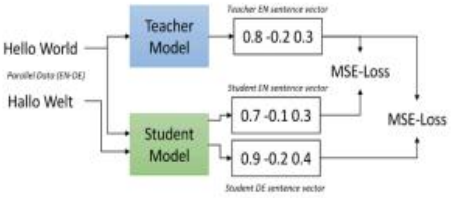
Deployment and Monitoring: Finally, the integrated models are deployed into production environments, and performance is monitored over time to ensure ongoing effectiveness and relevance. Regular updates and maintenance of the models are carried out to adapt to changing data distributions, evolving user preferences, and emerging trends in the problem domain.

By following this systematic methodology, organizations can effectively harness domain knowledge integration and distillation techniques to derive valuable insights,decision-making , and achieve tangible business outcomes across a wide range of applications and industries

4. Implementation

Domain knowledge integration and distillation involve incorporating expert knowledge into machine learning models to improve their performance and interpretability. Here are some implementations across various domains:

- **Healthcare:**
    - Integrating medical guidelines: Rule-based systems are integrated with machine learning models for tasks like patient diagnosis and treatment recommendation. This ensures adherence to best practices while leveraging the power of data-driven learning.
    - ○ Distilling physician expertise: Machine learning models are trained on data containing physician annotations or diagnoses, capturing their implicit knowledge and improving model accuracy on specific tasks.
- **Natural Language Processing (NLP):**
    - Utilizing ontologies and knowledge graphs, NLP models integrate information regarding entities, relationships, and events, thereby enhancing the accuracy and subtlety of text comprehension.
    - Distilling expert annotations: NLP models are trained on data annotated by domain experts, such as sentiment analysis in finance or medical text analysis, transferring their knowledge to the model.
    - As an illustration, consider a proposed method based on Knowledge Distillation to expand current sentence embedding models into new languages, consequently generating multilingual versions from previously monolingual models. In this instance, an English Sentence-BERT serves as the teacher model, while XLM-R functions as the student model. The training process is detailed below.



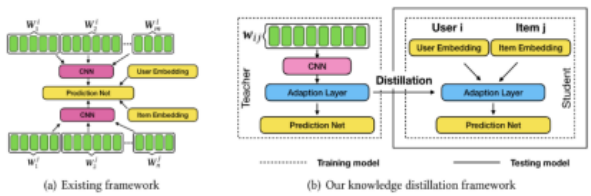
Given parallel data (e.g. English and German), train the student model such that the produced vectors for the English and German sentences are close to the teacher English sentence vector.

    - Fake News Detection System: The integration and distillation of domain knowledge through the use of machine learning and natural language processing (NLP) techniques is best demonstrated by a false news detection system. Through the integration of domain-specific knowledge into preprocessing techniques like stemming and lemmatization, the system simplifies textual input to its fundamental elements. This data is further reduced by feature extraction techniques like TF-IDF and Countvectorizer, which convert text into quantitative metrics that indicate the importance of terms in the context of the news. After being trained on these reduced features, the Random Forest classifier uses domain expertise to identify patterns suggestive of false news. This method not only improves the system's accuracy but also makes sure it can continue to adjust to the subtle changes in misleading data.
- **Robotics:**
    - **Integrating physical laws and robot kinematics:** Robot control algorithms consider domain knowledge about physics and robot mechanics, enabling safer and more efficient movements.

- **Learning from human demonstrations:** Reinforcement learning models are trained by observing human demonstrations of desired tasks, such as object manipulation or navigation, acquiring domain-specific skills.

- **Enhancing Image Recognition Models:**
  - Knowledge distillation techniques have been instrumental in optimizing image recognition models for deployment on edge devices and IoT platforms. By distilling the intricate features and visual representations from high-capacity convolutional neural networks into smaller counterparts, these techniques empower lightweight models to exhibit robust image recognition capabilities while conserving computational resources.

- **Recommendation Systems:**
  - Modern recommender systems have widely integrated external knowledge, including product photos or user reviews. This kind of information can provide detailed attributes about the individual or item, improve recommendation performance, and help explain recommendations better.
  - To illustrate the practical application of external knowledge, a recent study proposed a generalized distillation framework specifically designed for recommender systems. The framework's primary objective was to optimize model efficiency during the testing phase, ultimately leading to more responsive recommender systems. In this approach, a complex Convolutional Neural Network (CNN) was employed as the teacher model solely during the training phase, ensuring streamlined and effective knowledge transfer.



(a) Existing framework (b) Our knowledge distillation framework

(a) The core architecture of many existing review-based recommendation models. $\{W_1^i, W_2^i, \ldots W_m^i\}$ and $\{W_1^j, W_2^j, \ldots W_n^j\}$ are the reviews of user i and item j in the training set, which are respectively concatenated before inputting into the CNN networks to make the final rating prediction. (b) Our idea of decomposing user review modeling into a teacher model, which only exists in the training phase, and will not be leveraged when making predictions. w_{ij} is the review from user i to item j.

- **Visual Question Answering:**
  - VQA, or Visual Question Answering, addresses the scenario in which a model must determine the answer to a query regarding an input image. This is an extremely difficult challenge since VQA models handle several recognition tasks concurrently inside a single framework, requiring a grasp of both the natural text-based query and the local and global context of a picture.Therefore, a VQA model needs to be capable of a variety of reasoning tasks in order to extract relevant information from input images and questions.
  - Multiple Choice Learning, a framework used in VQA, assigns instances to a subset of models based on their best accuracy. It follows that every model should be tailored to particular kinds of cases. A Multiple Choice Learning with Knowledge Distillation framework was suggested by Mun et al. to learn models that are customized to a subset of tasks. Since the models trained via MCL have better oracle accuracy, that is, at least one of the models makes a true prediction for every example, the authors contend that the specialized models have the ability to beat the generalist models on all tasks.

5. Result

This study looks at two methods that are useful for managing the massive volumes of data we deal with on a daily basis: Domain Knowledge Integration (DKI) and Distillation (KD). Knowledge from multiple sources is merged by DKI, and essential points are extracted by KD from this combined knowledge. The procedure for putting these strategies into practice is described in the study, along with phases like data collecting and analysis. Applications in the real world are investigated, including the Fake News Detection and healthcare. Organizations can enhance operations, improve decision-making, and obtain a competitive edge by utilizing DKI and KD. All things considered, this study emphasizes how critical these methods are to making the most out of knowledge in the modern environment.

6. Challenges

Domain knowledge integration and distillation confront a myriad of challenges that impede their seamless implementation in various domains. Foremost among these hurdles is the inherent heterogeneity of data, stemming from disparate sources with differing formats, structures, and semantics. Harmonizing this diversity demands meticulous efforts to align representations and resolve semantic disparities. Additionally, the representation of domain knowledge, spanning unstructured text to structured databases and ontologies, poses a significant challenge, requiring cohesive frameworks to bridge these diverse forms. Ensuring the quality and reliability of integrated knowledge presents another formidable challenge, given the potential for data incompleteness, inconsistency, and noise. The scalability and efficiency of integration and distillation processes further complicate matters, particularly when dealing with large datasets or real-time data streams, necessitating robust computational frameworks and algorithms.

One significant challenge is the complexity of distilling actionable insights from vast and diverse knowledge sources. Integrating and distilling domain knowledge often involves

dealing with large volumes of data, spanning multiple sources and formats, which can be challenging to process efficiently. Additionally, ensuring the quality and reliability of distilled insights poses a formidable challenge, as data may contain errors, biases, or inconsistencies that can impact the accuracy of the distilled knowledge. Scalability is another pressing concern, particularly when distilling knowledge from massive datasets or in real-time scenarios, necessitating scalable algorithms and processing frameworks to handle the workload effectively.

Finally, the need for domain-specific expertise to interpret, validate, and contextualize distilled insights adds complexity, particularly in specialized or emerging domains. Addressing these challenges requires interdisciplinary collaboration, innovative methodologies, and a commitment to ethical practices to harness the full potential of domain knowledge distillation and integration while ensuring its reliability and relevance.

7. Future Perspective

7.1. The Future aspect of Domain Knowledge Integration

Machine learning complements domain knowledge integration by unraveling hidden patterns and relationships in data. By leveraging machine learning algorithms, data scientists can enhance the understanding derived from domain knowledge. This convergence enables the development of more accurate and interpretable models, ultimately improving decision-making processes and outcomes.

The Role of Artificial Intelligence in Domain Knowledge Integration . Artificial intelligence (AI) plays a pivotal role in automating the integration of domain knowledge, offering increased efficiency and scalability. By harnessing AI technologies, data scientists can streamline the process of combining specialized domain knowledge with data analysis. This integration empowers organizations to acquire valuable insights and make well-informed decisions.

Recognizing the pivotal role of domain knowledge in data science, greater collaboration between data scientists and domain experts is anticipated. This collaboration fosters a holistic approach to data analysis, where domain expertise and data science techniques intertwine seamlessly. The synergy between these professionals leads to the creation of reliable models and facilitates better decision-making across industries.

Anticipating the pivotal role of domain knowledge in the realm of data science, future trends suggest a deepening collaboration between data scientists and domain experts. This partnership fosters a holistic approach to data analysis, where the nuances of domain expertise intertwine seamlessly with the analytical capabilities of data science. The synergy between these professionals is poised to redefine the landscape, leading to the creation of reliable models and facilitating better decision-making practices across industries. This collaborative endeavor acknowledges the unique strengths each discipline brings to the table, promising a future where the intersection of domain knowledge and data science yields

unparalleled insights and innovations.

7.2. The Future aspect of Domain Knowledge Distillation

The future of domain knowledge distillation promises to revolutionize AI, making it more powerful, interpretable, and adaptable. Automated knowledge extraction, explainable distillation methods, and personalization will unlock this potential. Integration with emerging AI techniques like reinforcement learning and generative models will expand possibilities further. We must address ethical concerns like fairness and bias while ensuring responsible usage. With advancements in toolkits, standardization, and open-source initiatives, this exciting field holds the opportunity to harness the full potential of AI for the betterment of society.

Some of the special areas which can be explored in future w.r.t knowledge distillation are:

- **Specific domains:** Domain knowledge distillation has the potential to solve real-world problems in diverse domains, including healthcare, finance, NLP, robotics, and more. By delving into the unique challenges and opportunities within specific domains, data scientists can unearth novel applications for knowledge distillation. The impact of these solutions on society could be profound, bringing about positive transformations.
- **Ethical considerations:** As domain knowledge distillation progresses, it is imperative to address ethical concerns such as fairness, bias, and explainability. Ongoing research and best practices are essential for overcoming these challenges, ensuring responsible and unbiased usage of distillation methods. By actively addressing ethical considerations, we can harness the true potential of domain knowledge distillation for the benefit of society.

The continuous evolution of knowledge distillation is accompanied by advancements in toolkits, standardization

efforts, and open-source initiatives. These developments contribute to a more accessible and collaborative landscape, fostering innovation and knowledge sharing within the AI community. As toolkits become more sophisticated and standardized, data scientists and researchers can work more efficiently, accelerating the development and application of knowledge distillation techniques. The open-source nature of initiatives ensures a collective effort toward unlocking the true potential of AI for the broader benefit of society.

8. Conclusion

The exploration of domain knowledge integration and distillation underscores their critical importance in contemporary organizational settings. These methodologies serve as

indispensable frameworks for synthesizing diverse knowledge sources, extracting actionable insights, and driving innovation. Through a comprehensive review of existing literature, methodologies, and real-world applications, we have gained a nuanced understanding of these concepts and their significance in navigating the complexities of modern decision-making.

Despite the myriad challenges posed by data heterogeneity, scalability, and the complexity of distilling actionable insights, the future holds tremendous promise for domain knowledge integration and distillation. The convergence of machine learning and artificial intelligence offers opportunities to automate and enhance these processes, ultimately improving decision-making outcomes. Moreover, by addressing ethical considerations and fostering greater collaboration between data scientists and domain experts, we can ensure responsible and unbiased usage of distillation methods, thereby unlocking the full potential of these techniques for the benefit of society.

As we continue to advance knowledge and contribute to the ongoing discourse surrounding domain knowledge integration and distillation, we remain committed to fostering a culture of continuous improvement and innovation. By embracing these methodologies, organizations can streamline operations, enhance strategic planning, and secure a competitive advantage in the ever-changing landscape of modern technology. Together, let us harness the power of domain knowledge integration and distillation to drive positive transformations and shape a more resilient and agile future.

References

[1] Xu Chen, Yongfeng Zhang, and Zheng Qin. 2018. Adversarial Distillation for Efficient Recommendation with External Knowledge. ACM Journals. 37:35-54.

[2] Yuzhu Wang and Lechao Cheng. 2023. Improving Knowledge Distillation via Regularizing Feature Norm and Direction

[3] Himel Das Gupta .2022. A RoadMap to Domain Knowledge Integration in Machine Learning.

[4] Mertins, K., & Jochem, R. (2006). Integrating knowledge management and enterprise resource planning

[5] Zhang, X., & Zhao, J. (2018). Knowledge graph embedding: A survey of approaches and applications.

[6] Bonomi, L., Herrera, P., & Micchelli, C. A. (2012). Distillation of knowledge from multiple information sources. Applied and Computational Harmonic Analysis

[7] Tsai, C. H., Chiang, J. C., & Lin, C. H. (2015). Extracting and visualizing the knowledge map of educational big data