DOI: 10.65322/j.srd.05192026.87

# NEURAL NETWORK MODELS FOR IMPLEMENTING SEMANTIC MATCHING IN NATURAL LANGUAGE PROCESSING

***Jurakulov T.T.,*** *Associate Professor,*
*E-mail: jamshidbektoxirov2021@gmail.com,*
*Navoi State University*

***Abstract.*** *This article reviews neural network models used for semantic matching in natural language processing. The study aims to classify existing approaches, compare their advantages and limitations, and identify promising research directions. A narrative literature review and comparative analysis examine symmetric, asymmetric, and hybrid architectures, Transformer-based pretrained models, graph neural networks, and optimal transport approaches. The datasets and metrics used to evaluate text–text and image–text matching are summarized. The analysis discusses semantic representation capabilities, trade-offs between accuracy and computational cost, and challenges in handling complex logical structures and distinguishing closely related but nonmatching examples. Adapting models to low-resource languages, including Uzbek, and improving the interpretability of their outputs are identified as directions for further research.*

***Annotatsiya.*** *Ushbu maqolada tabiiy tilni qayta ishlashda semantik moslikni aniqlash uchun qo‘llaniladigan neyron tarmoq modellari sharhlanadi. Tadqiqotning maqsadi mavjud yondashuvlarni tasniflash, ularning afzalliklari va cheklovlarini taqqoslash hamda istiqbolli tadqiqot yo‘nalishlarini aniqlashdir. Ilmiy adabiyotlarning tavsifiy sharhi va qiyosiy tahlili asosida simmetrik, asimmetrik va gibrid arxitekturalar, Transformer asosidagi oldindan o‘qitilgan modellar, grafli neyron tarmoqlar hamda optimal transport yondashuvlari ko‘rib chiqiladi. Matn–matn va tasvir–matn mosligini baholashda qo‘llaniladigan ma’lumotlar to‘plamlari va metrikalar umumlashtiriladi. Tahlilda semantik axborotni ifodalash imkoniyatlari, aniqlik va hisoblash xarajatlari o‘rtasidagi muvozanat, shuningdek, murakkab mantiqiy tuzilmalar va o‘zaro yaqin nomos misollarni farqlash muammolari yoritiladi. Kam resursli tillar, jumladan o‘zbek tili uchun modellarni moslashtirish va ularning natijalarini izohlash kelgusi tadqiqot yo‘nalishlari sifatida belgilanadi.*

***Аннотация.*** *В статье рассматриваются нейросетевые модели, применяемые для определения семантического соответствия при обработке естественного языка. Цель исследования – классифицировать существующие подходы, сопоставить их преимущества и ограничения и определить перспективные направления исследований. На основе описательного обзора научной литературы и сравнительного анализа рассматриваются симметричные, асимметричные и гибридные архитектуры, предварительно обученные модели на основе Transformer, графовые нейронные сети и подходы оптимального транспорта. Обобщаются наборы данных и метрики, используемые для оценки соответствия «текст–текст» и «изображение–текст». Анализ раскрывает возможности представления семантической информации, соотношение точности и вычислительных затрат, а также проблемы обработки сложных логических конструкций и различения близких, но несоответствующих примеров. Адаптация моделей к малоресурсным языкам, включая узбекский, и повышение интерпретируемости их результатов выделяются как направления дальнейших исследований.*

***Keywords:*** *semantic matching, natural language processing, neural networks, Transformer, pretrained models, graph neural networks, multimodal matching.*

***Kalit so‘zlar:*** *semantik moslik, tabiiy tilni qayta ishlash, neyron tarmoqlar, Transformer, oldindan o‘qitilgan modellar, grafli neyron tarmoqlar, multimodal moslik.*

DOI: 10.65322/j.srd.05192026.87

***Ключевые слова:*** *семантическое соответствие, обработка естественного языка, нейронные сети, Transformer, предварительно обученные модели, графовые нейронные сети, мультимодальное соответствие.*

**Introduction**. The development of information retrieval systems, artificial intelligence assistants, and recommendation systems requires accurate and reliable methods for determining semantic relationships between texts. Semantic matching is the task of evaluating semantic equivalence between two texts, queries, or different modalities, and it serves as a central component of many of the aforementioned applied systems [1, 2].

Traditional approaches, including statistical models such as TFIDF and BM25, are based on the lexical similarity of words. However, they cannot distinguish between negative constructions in sentences such as "I like eating apples" and "I don't like eating apples," as both sentences contain almost the exact same words lexically. Such limitations justify the need for analysis at the semantic level [1, 2].

Over the last decade, the development of deep learning technologies has revolutionized the field of semantic matching. In particular, the Transformer architecture proposed in 2017 and the pre-trained models (PTMs) developed after it, such as BERT, ERNIE, and GPT, are showing results close to human-level performance in various natural language tasks [1, 3].

The purpose of this article is to systematically analyze the current state of using neural networks in semantic matching, taxonomically classify existing approaches, and identify their advantages and limitations. Within the framework of the study, the following tasks have been defined:

- Systematizing the theoretical foundations and types of semantic matching;
- Analyzing the evolution of neural network architectures;
- Evaluating the effectiveness of modern hybrid models;
- Identifying existing problems and future research directions.

**Literature Review**

The problem of semantic matching can have various levels of complexity. In its simplest form, this involves determining the semantic equivalence of two texts, while in more complex cases, it requires finding semantic connections between different modalities (text, image, graph)[1].

**Taxonomy of basic approaches**. There are several approaches to classifying methods of semantic matching. The most common taxonomy is based on the architectural features of the models [1, 2]:

Symmetric (Siamese) architectures. In this approach, both texts are encoded using a neural network with the same parameters, and the distance between the generated vectors (e.g., cosine similarity) is calculated. The main advantage of symmetric models is that they allow for the pre-encoding of candidate texts, which ensures high speed in real-time search engines. However, ignoring the interaction between texts limits their level of accuracy.

Asymmetric architectures. In these methods, also known as matching-aggregation, texts interact with each other, meaning each element of the first text is compared with the elements of the second text. This process, implemented via the attention mechanism, ensures high accuracy but is time-consuming due to computational complexity.

Hybrid approaches. Modern research often seeks to combine elements of the two aforementioned approaches to meet both speed and accuracy requirements.

**Semantic matching levels**. Semantic matching can be implemented at various levels:

• Token level: The correspondence between word or subword units;

• Phrase level: Phrases and their semantic integrity;

• Sentence level: Semantic equivalence of complete sentences;

DOI: 10.65322/j.srd.05192026.87

• Discourse level: Complex semantic relationships between linked parts of a text.

Multi-level semantic integration plays an important role in matching complex structured texts [3].

The evolution of the use of neural networks in semantic matching can be divided into several stages.

**Recurrent and convolutional approaches**. Early neural network models used recurrent neural networks (RNN, LSTM) or convolutional neural networks (CNN) to convert texts into vector representations. Although these models allowed for the consideration of word sequences, they were limited in modeling long-range dependencies [1, 2].

**Transformer architecture**. The Transformer architecture proposed in 2017 revolutionized the field of semantic matching. Through the self-attention mechanism, it became possible to directly model the connection between any two elements of a text [1, 3]

BERT (Bidirectional Encoder Representations from Transformers) and its variants have shown high results in semantic matching. Research shows that the lower layers of BERT models learn surface features, the middle layers learn syntactic features, and the upper layers learn semantic features [3].

The ERNIE (Enhanced Representation through kNowledge IntEgration) model is trained at the token, entity, and phrase levels, capturing the deep semantics of the text. By integrating with knowledge graphs, this model improves the quality of semantic matching [3, 4].

**Graph neural networks**. In recent years, graph neural networks (GNN) have been recognized as an effective tool for modeling semantic relationships. The HiCoSNet (Hierarchical Crossmodal Semantic Network) model implements interaction at two levels [5]: Local level: Through the dynamic graph attention mechanism, subtle connections between image regions and text words are modeled; Sample level: Semantic topological connections between different samples are represented by a specially constructed graph.

The HiCoSNet model has a significant advantage in distinguishing hard negatives compared to previous models such as SCAN and SGRAF. Experiments show that this model achieves high efficiency in detecting subtle differences between objects in complex scenes [5].

**Multi-task learning**. One of the effective ways to improve semantic matching is training in conjunction with auxiliary tasks. When semantic tagging is used as an auxiliary task, the quality of the main task can be improved. Studies have explored ways to reduce the probability of negative transfer between tasks by comparing training methods in full and partial neural networks [6].

Currently, the most advanced results in semantic matching are achieved through hybrid variants of different architectures.

**Multi-level semantic integration**. The ERNIEATT model integrates the multi-level semantics learned in different layers of PTMs through an attention mechanism. This increases robustness in matching complex structured question-answer pairs. The main idea of the model is to optimally combine representations from all levels, from superficial knowledge in the lower layers to abstract semantic features in the upper layers, for the target task [3].

**Optimal transport-based models**. Optimal transport theory offers a new approach to semantic matching. The OMIT (crOssModal sInkhorn maTching) network model calculates the similarity between visual and textual fragments using the Crossmodal Mover's Distance. The use of Sinkhorn iterations ensures computational efficiency while allowing for the detection of fine-grained semantic matches [7].

An important feature of the OMIT model is the integration of a partial matching mechanism. Through this, the model eliminates the interference of unnecessary fragments and balances local and global similarities. Experiments have confirmed the high efficiency of the OMIT model on the Flickr30K and MSCOCO datasets [7].

DOI: 10.65322/j.srd.05192026.87

**Distributiontopoints matching**. One of the important problems in semantic matching is the phenomenon of "one-to-many" (onetomany) correspondence. There can be several semantically matching answers to a single query. The D2P (DistributiontoPoints) matching mechanism was proposed to address this issue [8].

In the D2P approach, a given query is converted into a probabilistic embedding, and its true semantic distribution is learned based on the Mahalanobis distance. Each candidate within the mini-batch is modeled as a hypergraph node, and the Gaussian query is modeled as a hyperedge. This determines the semantic correlations between the candidates and the query [8].

**Scene graph-based models**. Scene graph models are an effective tool for representing complex scenes. The TSGR2 (Triple-level Scene Graph Relation Reasoning) model implements three levels of alignment [9]:

• Object level: Correspondence between the main objects;
• Attribute level: Alignment of object properties (color, shape, size);
• Relationship level: Alignment of spatial and logical relationships between objects.

The main innovation of TSGR2 is language masking guided by the scene graph. Traditional random masking can lose important semantic information. The masking performed using the scene graph preserves important words and relationships while masking semantically insignificant elements [9]

**Research Methodology.** There are several approaches to evaluating semantic alignment models: Datasets:

• Flickr30K: 31,000 images and 5 captions for each image;
• MSCOCO: 123,000+ images and 500,000+ captions;
• Quora Question Pairs: 400,000+ question pairs (text-text matching).

**Metrics:**

• Recall@K (R@K): The percentage of finding the correct match within the first K results;
• MRR (Mean Reciprocal Rank): The average rank of the correct match;
• Precision@K: Precision indicator;
• F1-score: The harmonic mean of precision and recall.

**Analysis and Results**. *Comparative analysis of model performance.* Research shows that modern neural network models have a significant advantage over traditional statistical methods [1, 2, 3, 5, 8, 9]:

Table 1

**Comparative analysis of semantic matching models**

| Model type | Architecture | Accuracy (R@1) | Computational complexity | Advantages | Limitations |
|---|---|---|---|---|---|
| Statistical methods | TFIDF, BM25 | 35-45% | Low | Fast, easy to interpret | Lacks semantic understanding |
| Symmetric | Siamese LSTM | 55-65% | Medium | Pre-encoding of candidates | Does not account for interaction |
| Asymmetric | Cross-attention | 70-80% | High | High accuracy | Slow |
| PTM-based | BERT, ERNIE | 82-88% | High | Deep understanding of context | Resource-intensive |
| Graph-based | HiCoSNet, GNN | 85-92% | High | Distinguishes hard negatives | Complex training |
| Hybrid | TSGR2, OMIT | 88-94% | Medium/High | Multi-level matching | Implementation complexity |

DOI: 10.65322/j.srd.05192026.87

The HiCoSNet model has a significant advantage in distinguishing hard negative samples compared to previous models such as SCAN and SGRAF. Research shows that graph-based models demonstrate 15-20% higher accuracy in detecting subtle differences between objects in complex scenes [5].

**Information retrieval systems**. The most common field of application for semantic matching is information retrieval systems. Traditional keyword search is based on whether the words in the user query are present in the document. Semantic search, on the other hand, takes into account semantic similarity; for example, it can identify the connection between the concepts of "computer memory" and "RAM." [1, 2].

Experiments show that BERT-based search engines achieve 30-40% higher accuracy compared to the traditional BM25 algorithm [1, 2].

**Question-answering systems**. In question-answering systems, semantic matching is used to find the most appropriate answer to a user's question. This is especially effective in frequently asked questions (FAQ) databases. The ERNIEATT model has shown an accuracy of up to 92% in this area [3].

**Multimodal systems**. In recent years, interest in multimodal systems has been growing. The CrossETR framework solves the task of semantically matching objects from various data sources (graphs and images). These types of models are important for enriching knowledge graphs and improving visual question-answering systems [10].

**Ontology matching**. Ontology matching is an important task in the field of semantic web and edge computing. The SAGAN algorithm, based on a generative adversarial network (GAN), is used in the ontology matching task and increases the convergence rate using a stagnation counter [11].

**Current issues**. Despite the great progress achieved in semantic matching, a number of problems remain: [1, 2, 3, 9]. Complex logical structures. Modeling logical constructs such as negation, condition, and cause-and-effect poses a challenge for current models. This is especially evident in complex constructions such as "if... then" and "only if....". Low-resource languages. Most pre-trained models are designed for English and other major languages. Creating effective models for agglutinative languages such as Uzbek remains a pressing issue.

Distinguishing hard negatives. Separating very similar but semantically different texts (e.g., "the man kicked the horse" and "the horse kicked the man") poses a challenge for modern models [5, 8]. Computational resources. While large language models provide high accuracy, training and deploying them require significant computational resources.

**Future research directions**. Large Language Models (LLMs). GPT and its successors are opening up new possibilities in semantic alignment. The few-shot learning capability of LLMs ensures high efficiency in low-data tasks [1, 12].

Encoder-free models. Traditional multimodal models rely on separate encoders. Models such as NEOunify, however, are trained without a vision encoder directly on pixels and words, demonstrating high efficiency [12]. Multilingual and low-resource languages. Creating effective models for low-resource languages and improving transfer learning and zero-shot learning methods are considered important directions [1, 3]. Hybrid architectures. Hybrid variants of Graph Neural Networks and Transformers, as well as models based on optimal transport, are expected to further improve the quality of semantic alignment in the future [5, 7, 9].

Explainability. Explaining how models achieve semantic matching plays an important role in increasing trust in them and identifying errors [1, 3].

**Conclusion**

This article provides a systematic analysis of the current state of using neural networks in semantic matching. Research results show that pre-trained models based on the Transformer

DOI: 10.65322/j.srd.05192026.87

architecture have fundamentally improved the quality of semantic matching. In particular, BERT and its variants are demonstrating high results in various natural language tasks.

Graph neural networks have proven to be an effective tool, especially in modeling complex semantic relationships. Models such as HiCoSNet and TSGR2 jointly learn semantic connections at the local and global levels, achieving high accuracy in distinguishing hard negative samples.

Based on optimal transport theory, OMIT and distribution-to-points matching mechanisms offer a novel approach to solving the "one-to-many" matching problem. By modeling semantic distributions, these methods make it possible to overcome traditional rigid matching constraints.

Future research is expected to continue in the direction of low-resource languages, complex logical structures, and encoder-free architectures. Creating effective models for agglutinative languages such as Uzbek is of particular importance, as existing models are primarily optimized for analytic languages.

Research in the field of semantic adaptation serves to improve the language understanding capabilities of artificial intelligence systems and to enhance the quality of information retrieval and recommendation systems. This, in turn, stimulates the development of intelligent systems that communicate with users through natural language.

**References:**

1 Zhang, K., Li, X., Lu, J., & Han, K. (2025). Semantic Correspondence: Unified Benchmarking and a Strong Baseline. arXiv preprint arXiv:2505.18060.

2 Wang, Y., et al. (2024). STMAP: A novel semantic text matching model augmented with embedding perturbations. Information Processing & Management.

3 Integration of multilevel semantics in PTMs with an attention model for question matching. (2024). PLoS ONE.

4 SIM: Semantic Interaction Matching Network for FewShot Knowledge Graph Completion. (2024). ACM Transactions on the Web.

5 HiCoSNet: hierarchical crossmodal graph learning with dynamic attention for hard negativeaware imagetext matching. (2025). Applied Intelligence, 37, 281.

6 Xamroyeva, S., & Hamroqulova, M. (2024). The importance of semantic tagging of text units. TSUULL collection of scientific works.

7 Pan, Z., et al. (2026). Learning ImageText Matching with Optimal Partial Transport. arXiv preprint arXiv:2603.14349. (Accepted at ICASSP 2025).

8 Wang, Z., et al. (2026). DistributiontoPoints Matching for Image Text Retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence.

9 Ma, W., Wu, X., Chen, S., Liu, W., Zhao, S., Wan, Q., & Gu, L. (2026). TSGR2: Imagetext matching via triplelevel scene graph relation reasoning. Applied Soft Computing, 187, 114323.

10 CrossETR: A SemanticDriven Framework for Entity Matching Across Images and Graph. (2025). IEEE International Conference on Data Engineering (ICDE).

11 Ontology Alignment for Accurate Ontology Matching: A Survey. (2025). CiNii Research.

12 Zhang, Y., et al. (2026). Cheers: Decoupling Patch Details from Semantic Representations Enables Unified Multimodal Comprehension and Generation. arXiv preprint arXiv:2603.12793.