Geometric Interpretation of Semantic Relationships: Filtering and Signature Formation for Neural Network Interoperability

V. G. Evdokimov
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
vigandvdk@gmail.com

A. A. Navrotsky
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
navrotsky@bsuir.by

***Abstract*—This paper proposes a method for filtering vector representations and forming universal signatures based on the geometric interpretation of semantic space as points on a hypersphere. The algorithm combines differential-geometric characteristics with statistical metrics to create a unified representation ensuring compatibility between various neural networks. The method reduces data dimensionality by selecting the most informative connections and forming compact signatures that are invariant to architectural features of models. Experiments confirm a reduction in computational complexity while simultaneously improving analysis quality. This approach potentially establishes a foundation for a universal knowledge exchange interface between heterogeneous neural network systems.**

***Keywords*—Semantic technologies, geometric feature space, data signatures, neural networks, unified data representation, radius of curvature, angular deficit, filtering and interneural network exchange.**

I. Introduction

In modern neural networks, significant advances are observed in the fields of computer vision [1], natural language processing [2], and complex time series analysis [3]. Progress in these areas has substantially improved the quality of solutions for many practical artificial intelligence tasks. However, the fundamental problem of semantic compatibility and efficient knowledge exchange between heterogeneous architectures remains unresolved, creating a significant barrier to the further development of multimodal intelligent systems and collaborative artificial intelligence, where heterogeneous neural network models must function within a unified semantic space, ensuring coherent interaction.

The main difficulties in this field arise from the absence of mathematically sound and computationally efficient methods for optimal knowledge transfer between different neural network architectures. Currently existing methods of transfer learning [4] and knowledge distillation [5] require significant computational resources, structural compatibility of source models, and long training times, substantially limiting their practical application in a number of important scenarios. Moreover, these approaches are often unable to preserve conceptual and semantic integrity of representations when working with complex structured data, which is critically important for modern artificial intelligence applications.

As Hinton et al. (2015) note, "knowledge distillation provides a mechanism for information transfer between models, however, it remains problematic when substantial architectural differences exist" [5]. This limitation becomes especially significant in the context of modern trends in deep neural network development, where there is considerable diversification of architectural solutions optimized for specific data types and tasks. Bronstein et al. (2017) emphasize that "non-Euclidean geometry provides an effective mathematical apparatus for processing complex-structured data" [7], which opens new perspectives for developing unified methods of knowledge representation in heterogeneous neural network systems.

This paper proposes an innovative method for forming unified signatures, based on the geometric interpretation of feature space as a multidimensional hypersphere. Such a geometric concept potentially allows overcoming limitations of existing approaches and likely can provide stable knowledge translation between different neural network architectures. The approach considered in this paper allows:

- reduce data dimensionality by selecting the most informative features;
- create compact signatures that can be used as normalized data representations;
- enable knowledge transfer between neural networks with different architectures.

The research objectives include attempting to develop a mathematically rigorous approach to feature selection, describing a possible algorithm for data signature formation, and verifying the method's effectiveness on synthetic data.

Research object: methods for processing multidimensional data to create a unified representation.

Research subject: geometric approach to feature filtering and data signature formation to enable knowledge transfer between neural networks.

Research method: mathematical modeling and geometric analysis of the feature space, as well as experimental verification of the developed method on synthetic data.

The proposed approach aims to solve the problem of data unification for various neural network architectures, which will increase the efficiency of their joint use in multimodal systems.

II. Analysis of Existing Approaches to Organizing Interfaces for Inter-Neural Network Information Exchange

A. Problem of Knowledge Transfer Between Neural Networks

Modern machine learning methods allow for effective training of neural networks on specialized tasks. However, when there is a need to transfer knowledge between neural networks, for example, between models with different architectures or tasks, significant obstacles arise. In particular, the lack of unified interfaces for representing data and knowledge complicates the implementation of collaborative learning or multimodal systems.

B. Main Approaches to Information Exchange Between Neural Networks

Transfer learning allows the use of a pre-trained model to accelerate the training of a new model on a similar task [4]. This is achieved by "transferring" weights or network layers to a new task. Examples include using pre-trained models such as ResNet [6] or BERT [2] for fine-tuning on specific tasks. Limitations:

- architectural similarity between models is required;
- it is impossible to use data with completely different structures;
- difficult to adapt to multimodal data (e.g., video, text, and numerical data).

Knowledge distillation allows transferring knowledge from a larger (teacher) model to a smaller (student) model, using the outputs of the teacher model as additional information when training the student model [5]. Limitations:

- dependence on the architectural features of the teacher model;
- complexity of training, especially in the case of significantly different architectures;
- high computational complexity for large models.

Training Multimodal Models Multimodal models, such as CLIP [8] or T5 [10], combine information from various sources (e. g., text and images) to solve tasks. These models are created using complex pretraining on enormous datasets. Limitations:

- enormous computational resources for training;
- lack of universality when adding new data types;
- inability to transfer knowledge between individual components of the model.

Geometric approaches Geometric methods, such as data representation on manifolds or in hyperbolic spaces [7], allow for considering complex dependencies between features. For example, graph neural networks (GNNs) use graph structures to transmit information. Limitations:

- high implementation complexity;
- limited applicability for tasks with time series or tabular data;
- necessity of prior knowledge of data structure.

Common Limitations of Existing Approaches

- architectural dependence: most methods require similar architectures for information transfer, making them unsuitable for heterogeneous systems;
- high computational complexity: methods such as knowledge distillation or transfer learning require significant resources for training;
- lack of normalized data representation: existing approaches do not offer a universal method for data representation suitable for transfer between neural networks;
- complexity of processing multimodal data: multimodal models are limited to predefined data types, making their extension difficult.

C. Potential Solution to Limitations Through the Proposed Method

Universality of data representation: potential data signatures can be compact, normalized vectors that can be transmitted to any neural network regardless of its architecture. This eliminates the dependency on architectural similarity.

Effective dimensionality reduction: using radii of curvature and angular deficits can allow selection of only the most important features, reducing data dimensionality and computational complexity.

Adaptability to data structure: the method can potentially automatically adapt to the data structure through analysis of the geometric properties of features. This allows for accounting for complex non-linear dependencies.

Noise resistance: noisy features can potentially be automatically excluded during filtering based on geometric criteria, increasing the quality of data representation. It should be noted that this research did not consider cases where noise contains useful information. This would require additional signature analysis prior to parameter filtering.

Multimodality: data signatures can potentially be formed for various data types (e. g., numerical, time series, images) and combined into a unified representation.

Thus, spherical filtering and data signature formation can potentially provide a unified interface for knowledge transfer between neural networks, making it suitable for building collaborative and multimodal systems.

III. Mathematical Formalization of the Proposed Method

A. Mapping Characteristics into a Multi-level Hyperspherical Structure

The vector space of characteristics $X \in \mathbb{R}^{m \times n}$, where m represents the number of instances and n is the dimension of the characteristic vector, undergoes transformation to form a multilevel hyperspherical structure. Within the simplified model considered in this section, we focus on the basic mapping to a unit hypersphere S^{n-1}, which is accomplished through centering followed by normalization (1):

$$\tilde{x_i} = \frac{x_i - \mu}{|x_i - \mu|}, \quad \mu = \frac{1}{m}\sum_{i=1}^{m} x_i, \tag{1}$$

where μ denotes the centroid of the characteristic distribution.

The fundamental advantage of this approach lies in the ability to organize different contextual feature sets into separate unit spheres, which collectively form a multidimensional hyperspherical structure. This allows modeling complex semantic relationships, where features from different modalities or architectures can exist in their own spherical subspaces while maintaining consistency within the overall hyperspherical topology. Such an organization provides a natural mechanism for encapsulating context-dependent semantic spaces, simultaneously creating prerequisites for constructing a metaspace in which coherent interaction of heterogeneous neural network representations becomes possible.

In the simplified model described by Equation (1), we consider projection onto a unit hypersphere as a basic case that demonstrates the key properties of the proposed method; however, the complete concept envisions a more complex structure of nested and interconnected spherical spaces.

B. Correlation Matrix

To analyze the relationships between features, an adjusted correlation matrix C is calculated, representing the cosine similarity between normalized features (2):

$$C_{ij} = \frac{\langle \tilde{x}_i, \tilde{x}_j \rangle}{|\tilde{x}_i||\tilde{x}_j|}, \quad \forall i, j \in [1, n], \tag{2}$$

where $\langle \cdot, \cdot \rangle$ is the dot product. Cosine similarity allows to evaluate the degree of dependence between features, considering their direction in space. Values close to 1 indicate a strong correlation, while values close to 0 indicate independence.

It is important to note that in the context of normalized features on a hypersphere, cosine similarity has a natural geometric interpretation as a measure of the angle between vectors. Since all vectors $\tilde{x}_i$ have unit length, the expression simplifies to the dot product of normalized vectors. For vectors located on the same hypersphere, this measure reflects their geodesic proximity and is more informative than traditional correlation measures in Euclidean space, especially when identifying non-linear dependencies in high-dimensional data.

The matrix C provides a complete topological map of relationships in the feature space, which is particularly important when integrating heterogeneous neural network architectures, where semantic consistency must be maintained not only at the level of individual characteristics but also at the level of their mutual relationships.

C. Radius of Curvature

The radius of curvature r_j determines the degree of independence of the feature j from other features (3):

$$r_j = 1 - \frac{1}{k} \sum_{i \in k\text{-neighbors}} |C_{ij}|, \tag{3}$$

where k is the number of closest neighbors, determined by the correlations $|C_{ij}|$. The larger the radius of curvature, the less dependent the feature is on others. This indicator is used to select features that contribute unique information to the model. Features with a low r_j can be considered redundant, as they have a strong correlation with other characteristics.

The concept of radius of curvature has a deep connection with the differential geometry of manifolds. In this context, it characterizes the local curvature of the information space around a specific feature. Mathematically, the value of r_j reflects the degree of "distinctiveness" of a feature in the overall data structure: high values correspond to features that form relatively isolated information clusters, while low values indicate features that are part of denser information structures.

The choice of parameter k in this context is critical and should be determined taking into account the dimensionality of the data and the expected degree of sparsity of the information space. Too small values of k can lead to noisy estimates of the radius of curvature, while too large values can level out local features of the data structure.

Applying the concept of radius of curvature allows effective solving the feature selection problem within the proposed hyperspherical representation, providing a balance between the informativeness of the model and its computational complexity. Furthermore, it creates a theoretical foundation for the subsequent development of mechanisms for combining heterogeneous neural network models through selective matching of their most informative components.

D. Angular Deficit

To evaluate the non-linearity of features, the angular deficit D_j is used, which is calculated as the deviation of the sum of angles from 2π (4):

$$D_j = 2\pi - \sum_{i=1}^{k} \arccos(C_{ij}), \quad (4)$$

where the angles $\arccos(C_{ij})$ are calculated for the nearest neighbors of feature j. The angular deficit allows for the identification of features with non-linear dependencies, since the sum of angles for linearly dependent features will be close to 2π.

The concept of angular deficit has roots in differential geometry, representing a measure of deviation of the local topology from Euclidean geometry. Positive values of D_j indicate positive curvature in the vicinity to the feature j, indicating non-linear relationships within the data structure.

The angular deficit and radius of curvature complement each other: while r_j characterizes the degree of independence of a feature, D_j evaluates the nature of this dependence. Two features may have similar r_j values but differ in D_j, indicating different types of relationships — linear or non-linear.

For reliable estimation of the angular deficit, it is recommended to use a sufficiently large sample and perform data normalization. The combined use of metrics r_j and D_j enables multi-criteria feature selection, accounting for both informational uniqueness and ability to model nonlinear relationships, which is particularly valuable when integrating heterogeneous neural network architectures.

E. Feature Filtering

Feature selection is carried out the basis of the radius of curvature and angular deficit. Feature j is included in the final set if the following conditions are met:

- The radius of curvature r_j exceeds a specified threshold ϵ_r, indicating its independence.
- The angular deficit D_j has a significant positive value, indicating the feature's ability to capture nonlinear relationships in the data.

This approach allows reducing the dimensionality of the feature space while preserving the most important features that contribute unique information to the model, thus reducing data redundancy and improving model interpretability.

The feature filtering process represents a multi-criteria optimization problem, balancing feature independence and their ability to describe nonlinear relationships. High values of r_j (closer to 1) indicate that the feature j is weakly correlated with other features and carries unique information. For the angular deficit D_j, minimal values indicate linear dependencies, while higher positive values correspond to nonlinear relationships. The selection of the optimal threshold ϵ_r depends on the specific domain and data characteristics, with adaptive thresholds sometimes preferable.

The methodology can be enhanced with a feature ranking procedure, where each feature is assigned a composite rating based on a weighted combination of r_j and D_j, allowing for more flexible control of the feature selection process.

F. Formation of Semantic Signature as a Unified Interface

For each time window X_t, a semantic signature is formed, representing a unified representation of aggregated characteristics of the selected features:

$$\begin{aligned}\text{Signature}(X_t) = [&\mu_1, \sigma_1, \text{kurt}_1, \text{skew}_1, \ldots, \\ &\mu_p, \sigma_p, \text{kurt}_p, \text{skew}_p], \end{aligned} \quad (5)$$

where μ, σ, kurt, skew are the mean value, standard deviation, kurtosis, and skewness respectively, calculated for each selected feature.

The semantic signature serves as a unified interface enabling effective interaction between heterogeneous neural network architectures. This approach creates an intermediate abstraction layer allowing different types of neural networks to exchange information in a standardized format, functioning as a semantic data descriptor invariant to the internal architectural specifics of particular models.

The proposed interface addresses the problem of integrating heterogeneous neural components by forming a common semantic space for data interpretation. For example, a transformer-based model can utilize signatures from a convolutional neural network without transforming internal data representations. This is crucial in ensemble systems and multimodal architectures where subsystems specialize in different aspects of analysis.

From an ontological perspective, the semantic signature maps multidimensional feature space into a structured representation preserving key distributional characteristics of the original data. The semantic interoperability provided by this mechanism enables scalable hybrid architectures where different neural networks can be dynamically combined depending on the task context and input data characteristics.

IV. Experimental Results

To conduct experimental verification of the method, a synthetic dataset was used, containing 14 descriptive features of various types: independent basic features, linearly dependent derivatives, nonlinearly dependent derivatives, and stochastic noise components.

A. Description of the Experiment

Generation of Synthetic Data:

- 1500 samples were created using sinusoidal signals, noise, and nonlinear dependencies;
- the data includes independent features, derivative features, and noise components;
- the target variable (y) is constructed as a nonlinear combination of features with added random noise.

Feature Filtering:

- spherical filtering was performed using radii of curvature and angular deficits;
- 12 of the most informative features out of 14 were selected;
- filtering reduced the dimensionality of the space while preserving useful information.

Data Preparation:

- without using signatures: feature windows were used in their "raw" form;
- with signatures: statistical characteristics (mean, standard deviation, seasonality, and trend) were calculated for each time window.

Model Training and Evaluation:

- 6 models were used: MLP, Random Forest, Linear Regression, SVR, XGBoost, LightGBM, as well as the VotingRegressor ensemble;
- evaluation was conducted using cross-validation (TimeSeriesSplit, 5 folds);
- metric: MSE (Mean Squared Error).

Comparison of Results:

- MSE of models was compared when using "raw" data and data with signatures;
- For each model, it was determined whether the signature improved the forecasts.

Final Analysis:

- a table with results was constructed and conclusions were drawn about the impact of signatures on the performance of each model;
- it was noted that signatures are particularly useful for linear models and ensembles, but require refinement for nonlinear models.

B. Analysis of Results

Feature Selection Using Spherical Filtering:

- 12 features out of 14 were selected based on radii of curvature and angular deficits. The dimensionality of the space was reduced while preserving key information. This confirms the effectiveness of the selection method based on geometric characteristics of the feature space;
- The dimensionality of the hypersphere after feature selection: S^{13}, which corresponds to 14 features in the original space

Results Without Using Signatures:

- Linear Regression showed the worst MSE value (2409.5115), which is expected since it performs poorly with nonlinear relationships;
- SVR showed the best MSE value (12.0612), which is related to its ability to process nonlinear relationships;
- Ensemble methods (Random Forest, XGBoost, LightGBM, VotingRegressor) demonstrated MSE in the range of 15-79, which also shows their robustness.

Results Without Using Signatures:

- Significant improvement is observed for Linear Regression: MSE decreased from 2409.5115 to 117.0779 (95.1% reduction). This suggests that signatures help linear models better account for complex dependencies;
- Ensemble models (Random Forest, XGBoost, LightGBM) also showed a decrease in MSE, confirming the value of signatures for models working with nonlinear data;
- For MLP and SVR, MSE increased, which may be related to signatures distorting the original data that these models could optimally process.

VotingRegressor: The average MSE of VotingRegressor with signatures improved: decreasing from 78.9186 to 20.9154. This indicates that combining multiple models benefits from using signatures.

Application of the developed filtering algorithm, based on evaluating the radii of curvature of spherical polygons and calculating angular deficits, made it possible to reduce the dimensionality of the feature space by 14.3%, reducing the number of features to 12.

When exceeding the critical volume of the feature set, which allows for generating new descriptors through algebraic or functional transformations of the basic subset, a significantly more pronounced dimensionality reduction is theoretically predicted. The prediction quality results are shown in Table I.

Table I
Comparison of Mean Squared Error (MSE) for Various Models:

№	**Results**		
	Model	***Without signature***	***With signature***
1	MLP	14.5487	14.9106
2	RF	15.8648	14.3694
3	LR	2409.5115	117.0779
4	SVR	12.0612	14.9699
4	XGB	15.0982	13.7349
4	LGBM	16.2109	13.8642
4	Voting	78.9186	20.9154

[a]Using synthetic data

The use of data signatures demonstrated a significant improvement in the quality of forecasts for most models, especially for Linear Regression (LR), where the reduction in Mean Squared Error (MSE) was 95%. This confirms the high efficiency of the method for models that are sensitive to input data quality and redundant features. However, for MLP (Multi-Layer Perceptron) and SVR (Support Vector Regression) models, the use of signatures led to a slight deterioration in forecast quality. Nevertheless, this deterioration is not significant and demonstrates that the method can be applicable to these models after additional refinement and optimization.

V. Conclusion

In this work, an innovative feature filtering method based on geometric formalization and data signature creation is presented. The method provides a unified information representation for compatibility between heterogeneous neural network architectures.

The proposed mathematical apparatus, utilizing geometric characteristics of the feature space (curvature radii, angular deficits), allows for efficient selection of the most significant descriptors. The formation of complex signatures achieves compression of semantic relationships while preserving contextual integrity.

Experimental verification confirmed the approach's effectiveness: on a heterogeneous dataset, a dimensionality reduction of 14.3% was achieved without loss of informativeness. The positive effect is particularly pronounced for linear models and ensemble algorithms (MSE reduction up to 95.1%).

Prospects for further research include adapting the method for nonlinear architectures (MLP, SVR) and expanding testing on complex multi-level data, approximating real-world semantic information processing tasks.

Method Advantages

- effective feature filtering based on geometric characteristics, providing a 14.3% dimensionality reduction without loss of informational significance;
- formation of informative signatures integrating statistical, frequency, and seasonal parameters, ensuring unified representation for various architectures;
- significant improvement in modeling quality for linear (95.1% MSE reduction) and ensemble methods;
- reduced computational complexity: decreasing the feature space dimensionality facilitates model training and reduces the risk of overfitting.

Method Limitations

- insufficient effectiveness for nonlinear architectures (MLP, SVR), requiring additional method adaptation;
- validation on synthetic data, necessitating verification on real-world multimodal datasets;
- sensitivity to filtering parameters, requiring careful tuning of geometric characteristics (curvature radii and angular deficits).

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105, 2012.

[2] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2019.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in International Conference on Artificial Neural Networks, pp. 270–279, 2018.

[5] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[7] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond Euclidean data," IEEE Signal Processing Magazine, vol. 34, no. 4, pp. 18–42, 2017.

[8] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, et al., "Learning transferable visual models from natural language supervision," in Proceedings of the International Conference on Machine Learning, pp. 8748–8763, 2021.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.

[10] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," The Journal of Machine Learning Research, vol. 21, no. 1, pp. 5485–5551, 2020.

ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ СЕМАНТИЧЕСКИХ СВЯЗЕЙ: ФИЛЬТРАЦИЯ И ФОРМИРОВАНИЕ СИГНАТУР ДЛЯ ИНТЕРОПЕРАБЕЛЬНОСТИ НЕЙРОСЕТЕЙ

Евдокимов В. Г., Навроцкий А. А.

В данной статье рассматривается метод фильтрации векторных представлений и формирования универсальных сигнатур на основе геометрической интерпретации семантического пространства как точек на гиперсфере. Алгоритм объединяет дифференциально-геометрические характеристики со статистическими метриками для создания единого представления, обеспечивающего совместимость между различными нейронными сетями. Метод уменьшает размерность данных путем выбора наиболее информативных связей и формирования компактных сигнатур, инвариантных к архитектурным особенностям моделей. Эксперимент подтверждает снижение вычислительной сложности при одновременном улучшении качества анализа. Такой подход потенциально создает основу для универсального интерфейса обмена знаниями между гетерогенными нейронными сетевыми системами.

Received 30.03.2025