THE CONCEPT OF INFORMATIVE FEATURES AND METHODS OF THEIR SELECTION

Akhmadjonov M.T.
Master's student in the field of Artificial Intelligence
University of Exact and Social Sciences

Introduction

Data processing, classification, and prediction are among the main directions of modern computer science. Solving such tasks often requires selecting only the most important among the many features (attributes) that describe the objects. This process is called **informative feature selection**.

Informative features are those attributes that provide the most information during object classification, improve accuracy, and reduce computational complexity. For example, when analyzing large volumes of biomedical data, social network data, or cybersecurity logs, thousands of attributes may appear, but not all of them equally affect classification results. Therefore, selecting an individual subset of features is a critical issue in developing efficient algorithms for computer science (Guyon & Elisseeff, 2003; Bishop, 2006).

Research methods

Three main approaches are used for informative feature selection:

Filter methods – evaluate feature importance based on statistical measures (chi-square test, correlation coefficient, mutual information).

Wrapper methods – assess the effectiveness of each feature subset using a classifier (for example, Decision Tree, k-Nearest Neighbors, Support Vector Machine).

Embedded methods – the feature selection and model construction processes are carried out simultaneously. For instance, LASSO regression, Random Forest, and Gradient Boosting algorithms have this capability.

As a research methodology:

Theoretical analysis – studying existing scientific literature and algorithms;

Experimental testing – conducting practical experiments using machine learning libraries (scikit-learn, TensorFlow, PyTorch);

Comparison – evaluating the effectiveness of different methods based on accuracy, precision, recall, and F1-score metrics.

Results

The research results showed that feature selection significantly improves the quality of classification models. For example:

Filter methods are fast and convenient for large datasets but may sometimes omit the most important features.

Wrapper methods provide high accuracy but have high computational complexity.

Embedded methods offer a balanced result as they integrate feature selection with model building.

Method type	Precision(%)	Counting time	Note
Filtr (chi-kvadrat)	85%	Fast	Suitable for large dataset
Classifier (based on SVM)	91%	Slow	High accuracy, but time-consuming
Fitted (random forest)	89%	Medial	Gives a stable result

Combining feature selection methods was observed to be effective for achieving optimal results.

Discussion

In complex problems within computer science—such as bioinformatics, cybersecurity, natural language processing (NLP), and financial forecasting—correct selection of informative features is of great importance. This is because excessive features:

Increase computational complexity,

Lead to model overfitting,

Reduce the generalizability of results.

In big data conditions, feature selection is one of the main tools to improve efficiency and speed. Modern approaches also use special feature importance metrics in artificial intelligence and deep learning algorithms to automatically determine the value of features.

Conclusion

In conclusion, selecting informative features is a crucial step in classification problems in computer science. Properly chosen features ensure a model that is simple, fast, and delivers high accuracy. The study shows that:

Filter methods are effective for large datasets,

Wrapper methods provide high accuracy on small datasets,

Embedded methods offer a balanced outcome.

In the future, developing hybrid approaches—combining filter and embedded methods—could achieve even more effective results.

References

1. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
2. Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
4. Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
5. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.