

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ
БЕЛАРУСЬ**
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра интеллектуальных методов моделирования

Иванов
Дмитрий Сергеевич

Аннотация к дипломной работе

Нейросетевые алгоритмы кластеризации

Научный руководитель:
доктор физ.-мат. наук
К.В. Лыков

Минск, 2025

АННОТАЦИЯ

Дипломная работа содержит: 28 страниц, 8 иллюстраций, 1 таблицу, 4 использованных источника.

Ключевые слова: КЛАСТЕРИЗАЦИЯ, ОБУЧЕНИЕ БЕЗ УЧИТЕЛЯ, НЕЙРОННЫЕ СЕТИ, САМООРГАНИЗУЮЩИЕСЯ КАРТЫ КОХОНЕНА (SOM), U-MATRIX, МЕТОД К-СРЕДНИХ, ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ, АНАЛИЗ ДАННЫХ, ВИЗУАЛИЗАЦИЯ ДАННЫХ.

В дипломной работе исследуются нейросетевые алгоритмы кластеризации, с особым акцентом на самоорганизующиеся карты Кохонена (SOM). Проводится сравнительный анализ их эффективности с классическими методами, такими как k-means и иерархическая кластеризация, при работе с данными сложной структуры.

Целью работы является систематический анализ преимуществ и недостатков SOM, демонстрация их способности обнаруживать нелинейные зависимости и сохранять топологическую структуру данных, а также исследование методов визуализации и оценки качества кластеризации.

Методология исследования включала теоретический анализ математических основ алгоритмов кластеризации, проведение вычислительного эксперимента на языке Python с использованием библиотек Scikit-learn и MiniSom, генерацию синтетических наборов данных с различной топологией, а также оценку качества кластеризации с помощью внешних (ARI) и внутренних (Silhouette Score) метрик.

В результате эксперимента было установлено, что классические методы не справляются с кластеризацией данных со сложной нелинейной топологией. Самоорганизующиеся карты Кохонена продемонстрировали способность эффективно идентифицировать кластеры невыпуклой формы и сохранять их взаимное расположение, что делает их мощным инструментом для исследовательского анализа.

Полученные в работе результаты применимы для решения практических задач анализа данных в биоинформатике, финансовой аналитике и компьютерном зрении.

АННАТАЦЫЯ

Дыпломная работа змяшчае: 28 старонак, 8 ілюстрацый, 1 табліцу, 4 выкарыстаныя крыніцы.

Ключавыя словы: КЛАСТАРЫЗАЦЫЯ, НАВУЧАННЕ БЕЗ НАСТАЎНІКА, НЕЙРОННЫЯ СЕТКІ, САМААРГАНІЗУЕМЫЯ КАРТЫ КАХОНЕНА (SOM), U-MATRIX, МЕТАД К-СЯРЭДНІХ, ІЕРАРХІЧНАЯ КЛАСТАРЫЗАЦЫЯ, АНАЛІЗ ДАДЗЕНЫХ, ВІЗУАЛІЗАЦЫЯ ДАДЗЕНЫХ.

У дыпломнай рабоце даследуюцца нейрасеткавыя алгарытмы кластарызацыі, з асаблівым акцэнтам на самаарганізуемыя карты Кахонена (SOM). Праводзіцца параўнальны аналіз іх эфектыўнасці з класічнымі метадамі, такімі як k-means і іерархічная кластарызацыя, пры рабоце з дадзенымі складанай структуры.

Мэтай работы з'яўляецца сістэматычны аналіз пераваг і недахопаў SOM, дэманстрацыя іх здольнасці выяўляць нелінейныя залежнасці і захоўваць тапалагічную структуру дадзеных, а таксама даследаванне метадаў візуалізацыі і ацэнкі якасці кластарызацыі.

Метадалогія даследавання ўключае тэарэтычны аналіз матэматычных асноў алгарытмаў кластарызацыі, правядзенне вылічальнага эксперыменту на мове Python з выкарыстаннем бібліятэк Scikit-learn і MiniSom, генерацыю сінтэтычных набораў дадзеных з рознай тапалогіяй, а таксама ацэнку якасці кластарызацыі з дапамогай знешніх (ARI) і ўнутраных (Silhouette Score) метрык.

У выніку эксперыменту было ўстаноўлена, што класічныя метады не спраўляюцца з кластарызацыяй дадзеных са складанай нелінейнай тапалогіяй. Самаарганізуемыя карты Кахонена прадэманстравалі здольнасць эфектыўна ідэнтыфікаваць кластары невыпуклай формы і захоўваць іх узаемнае размяшчэнне, што робіць іх магутным інструментам для даследчага аналізу.

Атрыманыя ў рабоце вынікі могуць быць выкарыстаны для вырашэння практычных задач аналізу дадзеных у біяінфарматыцы, фінансавай аналітыцы і камп'ютарным зроку, дзе традыцыйныя падыходы з'яўляюцца неэфектыўнымі.

ANNOTATION

The thesis contains: 28 pages, 8 figures, 1 table, 4 references.

Keywords: CLUSTERING, UNSUPERVISED LEARNING, NEURAL NETWORKS, KOHONEN SELF-ORGANIZING MAPS (SOM), U-MATRIX, K-MEANS METHOD, HIERARCHICAL CLUSTERING, DATA ANALYSIS, DATA VISUALIZATION.

This thesis investigates neural network clustering algorithms, with a special focus on Kohonen Self-Organizing Maps (SOM). It provides a comparative analysis of their effectiveness against classical methods, such as k-means and hierarchical clustering, when applied to data with complex structures.

The aim of the work is a systematic analysis of the advantages and disadvantages of SOM, a demonstration of their ability to detect non-linear dependencies and preserve the topological structure of data, as well as an exploration of visualization and clustering quality assessment methods.

The research methodology included a theoretical analysis of the mathematical foundations of clustering algorithms, conducting a computational experiment in Python using the Scikit-learn and MiniSom libraries, generating synthetic datasets with various topologies, and evaluating clustering quality using external (ARI) and internal (Silhouette Score) metrics.

The experiment established that classical methods fail to cluster data with complex non-linear topologies. Self-Organizing Maps demonstrated the ability to effectively identify non-convex clusters and preserve their spatial relationships, making them a powerful tool for exploratory analysis.

The results obtained in this work are applicable to solving practical data analysis problems in fields such as bioinformatics, financial analytics, and computer vision, where traditional approaches prove to be ineffective.