

ИССЛЕДОВАНИЕ АЛГОРИТМОВ И МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Драган Д.Д.

Содержание

- Интеллектуальная система поддержки принятия решений
- Кластеризация
- Self-Organizing Map
- Результаты моделирования алгоритма кластеризации
- Выводы

- **Интеллектуальная система поддержки принятия решений**
- Кластеризация
- Self-Organizing Map
- Результаты моделирования алгоритма кластеризации
- Выводы

Интеллектуальная система поддержки принятия решений

Интеллектуальная система поддержки принятия решений (ИСППР) (англ. *Intellect Decision Support System, DSS*) —

компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности.

Интеллектуальная система поддержки принятия решений

Data Mining (рус. Интеллектуальный анализ данных) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Интеллектуальная система поддержки принятия решений

Задачи, решаемые методами Интеллектуальный анализ данных в системах поддержки принятия решений, принято разделять на

- описательные (англ. descriptive);
- предсказательные (англ. predictive).

В описательных задачах самое главное — это дать наглядное описание имеющихся скрытых закономерностей, в то время как в предсказательных задачах на первом плане стоит вопрос о предсказании для тех случаев, для которых данных ещё нет.

К описательным задачам относятся:

- Поиск ассоциативных правил или паттернов (образцов).
- Группировка объектов или кластеризация.
- Построение регрессионной модели.

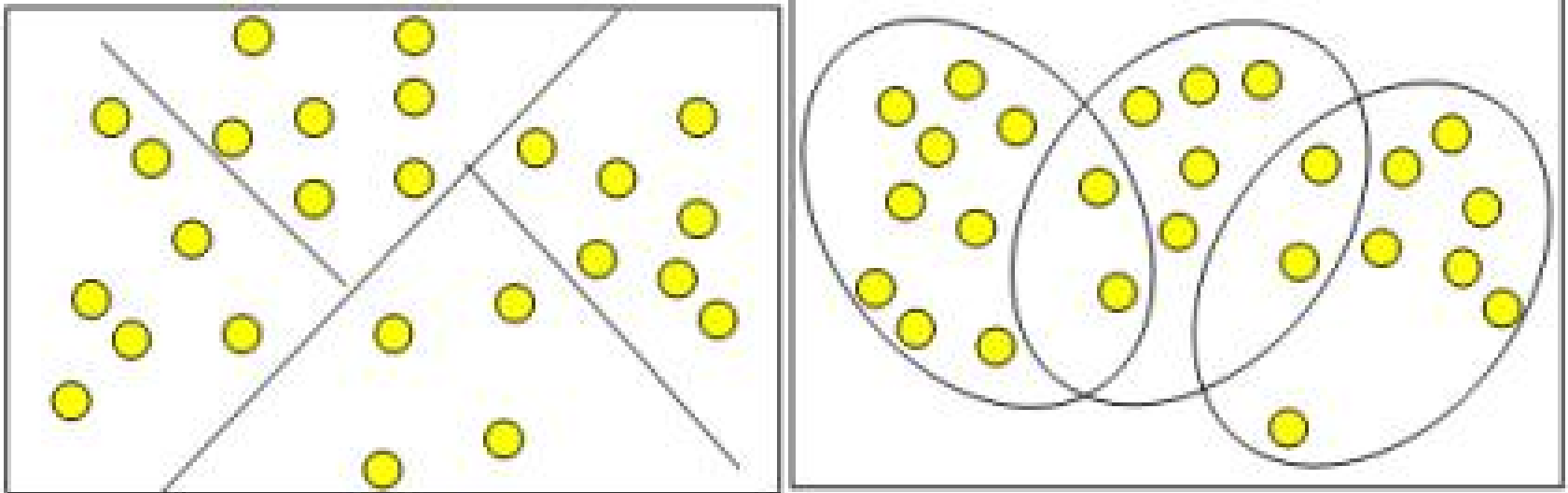
К предсказательным задачам относятся:

- Классификация объектов (для заранее заданных классов).
- Построение регрессионной модели.

- Интеллектуальная система поддержки принятия решений
- **Кластеризация**
- Self-Organizing Map
- Результаты моделирования алгоритма кластеризации
- Выводы

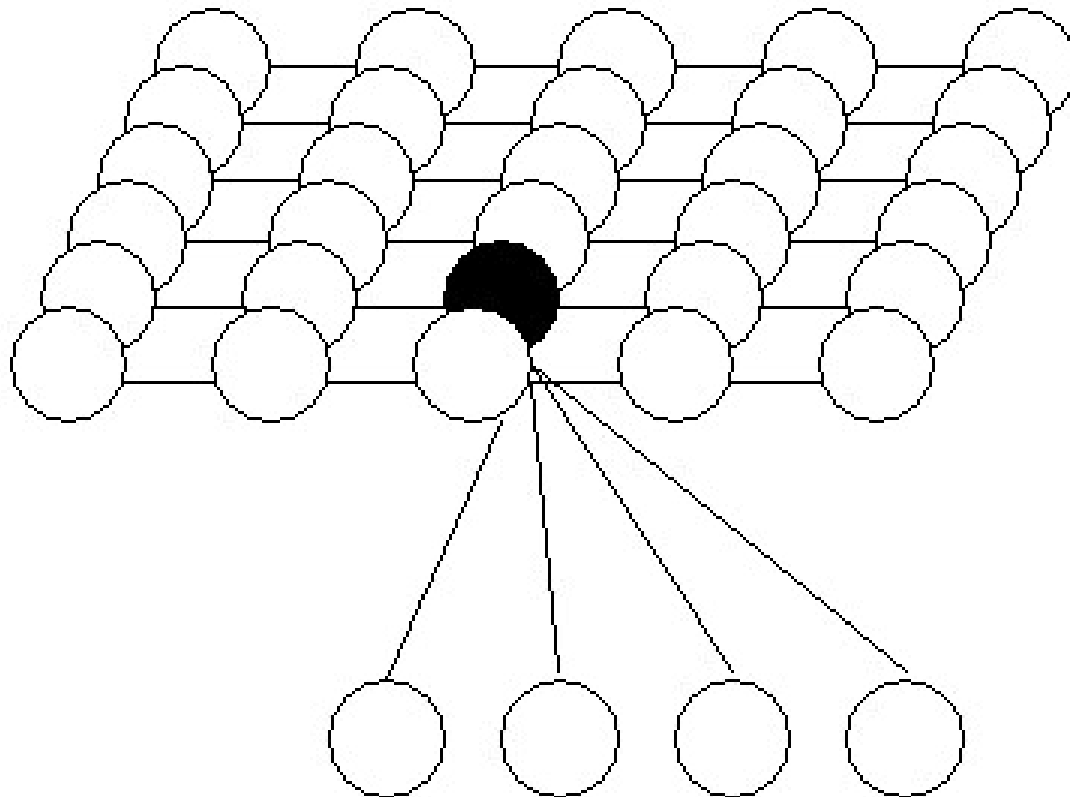
Кластеризация

Кластеризация (или кластерный анализ) – это задание автоматического разбития множества объектов на группы (кластеры).



- Интеллектуальная система поддержки принятия решений
- Кластеризация
- **Self-Organizing Map**
- Результаты моделирования алгоритма кластеризации
- Выводы

Self-Organizing Map



Kohonen-Neuronen
 $1, \dots, n$

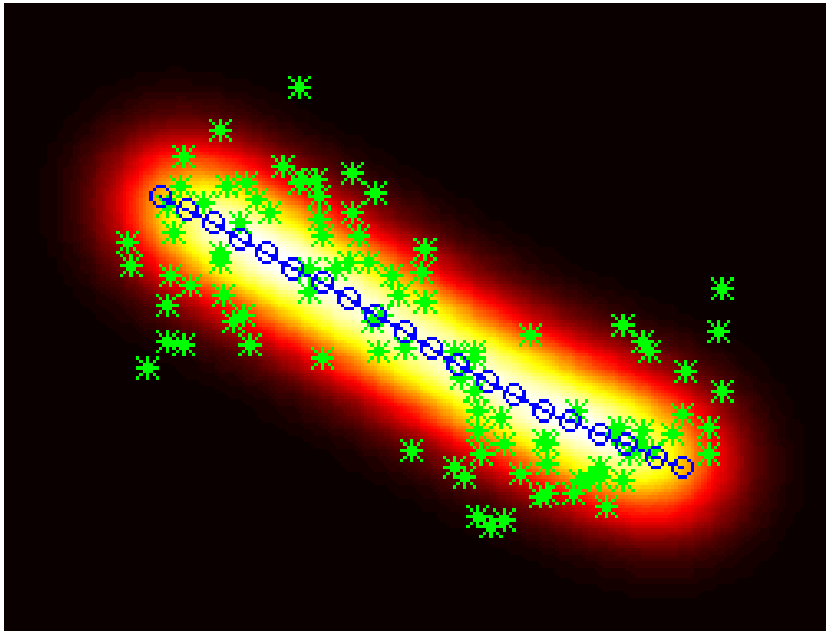
$$\mathbf{w}_j = (w_{j1}, w_{j2}, \dots, w_{jn})$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

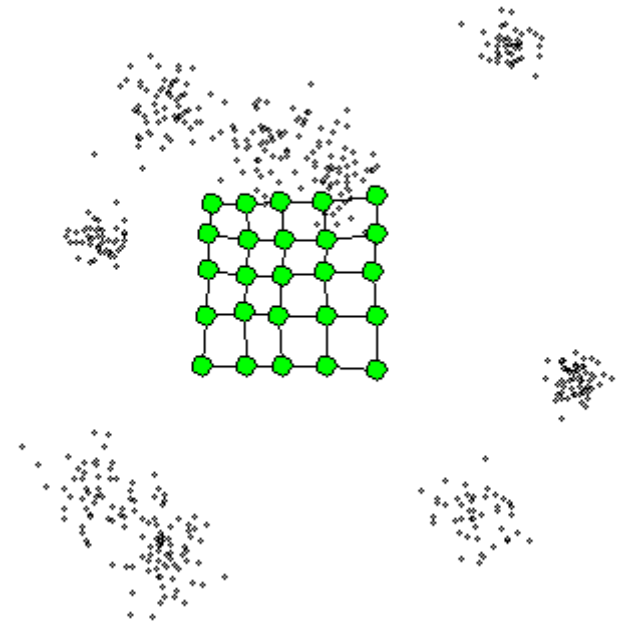
Self-Organizing Map

Виды представления

Одномерная структура
(цепочка)



Двумерная структура
(решетка)



Self-Organizing Map

После инициализации карты проходят три основных процесса, связанные с формированием самоорганизующей карты:

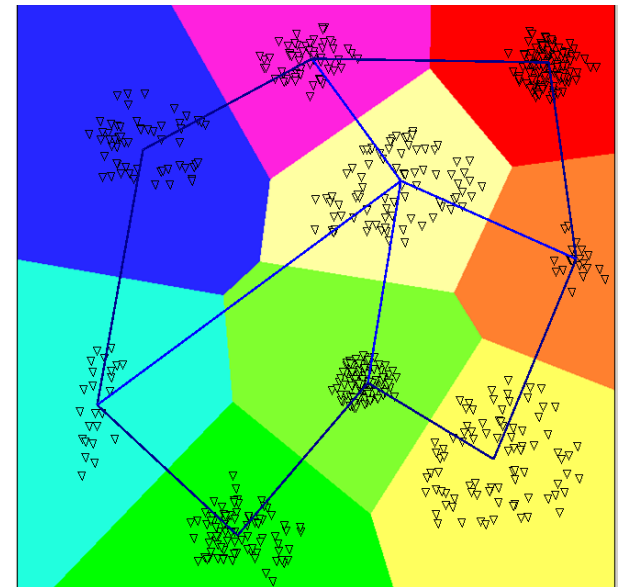
- ***Состязание***
- ***Сотрудничество***
- ***Синаптическая адаптация***

Self-Organizing Map

Состязание

Для каждого входного вектора в нейронах на карте высчитываются его соответствующие значения дискриминантной функции. За счет дискриминантной функции обеспечивается состязания между нейронами. Кроме этого нейрон с наибольшим значениям дискриминантной функции оглашается «победителем» соревнования.

Наиболее распространенный выбор дискриминантной функции основан на максимальном скалярном произведении $\overline{w}_j^T \overline{x}$, математически эквивалентен к минимизации Эвклидова расстояния.



Self-Organizing Map

Сотрудничество

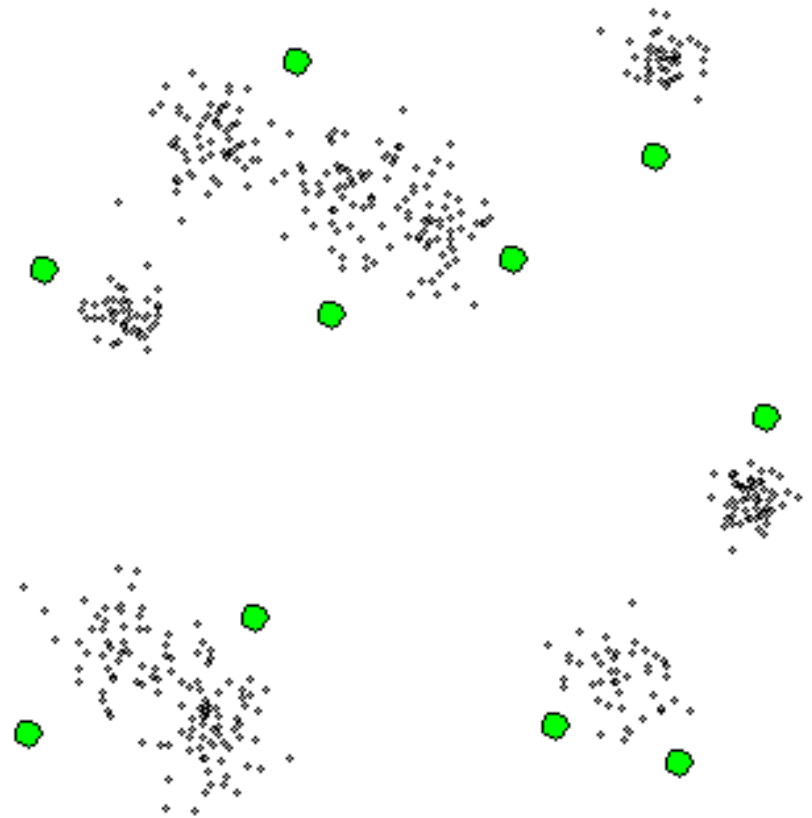
Нейрон-победитель определяет пространственное размещение топологического окружения нарушенных нейронов, тем самым обеспечивается основа для сотрудничества между соседними нейронами.



Self-Organizing Map

Синаптическая адаптация

Последний механизм позволяет смещенным нейронам увеличить их значения дискриминантной функции по отношению к входному образцу .
Внесение таких корректив обеспечивает приближения нейронов–победителей к входного набору.



- Интеллектуальная система поддержки принятия решений
- Кластеризация
- Self-Organizing Map
- **Результаты моделирования алгоритма кластеризации**
- Выводы

Результаты моделирования алгоритма кластеризации

Тест №1

Ирисы Фишера состоят из данных о 150 экземплярах ириса, по 50 экземпляров из трёх видов — Ирис щетинистый (*Iris setosa*), Ирис виргинский (*Iris virginica*) и Ирис разноцветный (*Iris versicolor*).

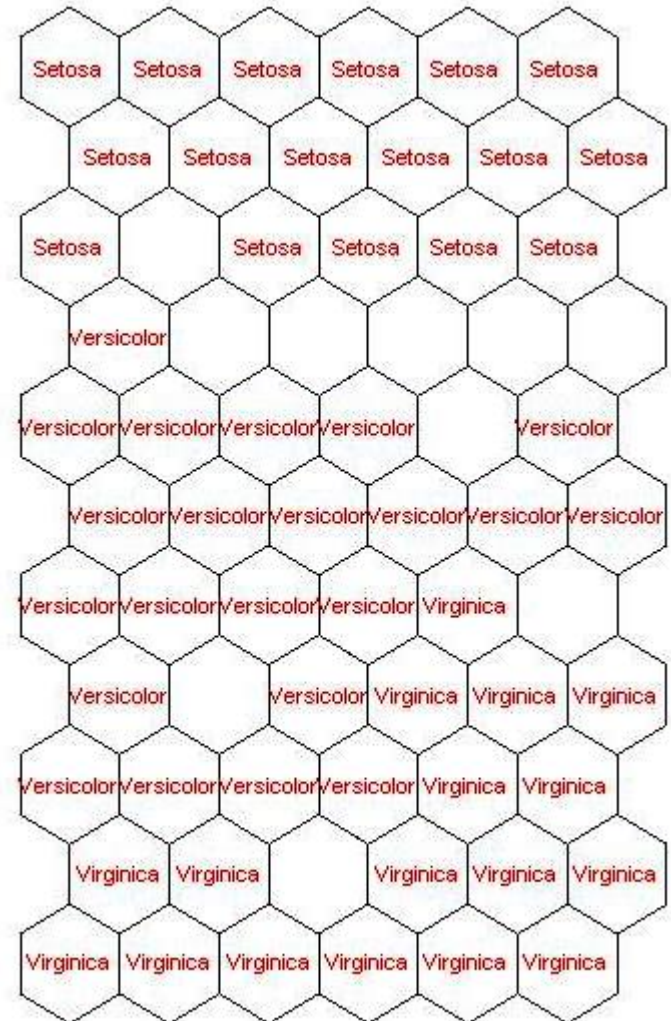
Для каждого экземпляра измерялись четыре характеристики (в сантиметрах):

- Длина чашелистика (англ. *sepal length*);
- Ширина чашелистика (англ. *sepal width*);
- Длина лепестка (англ. *petal length*);
- Ширина лепестка (англ. *petal width*).



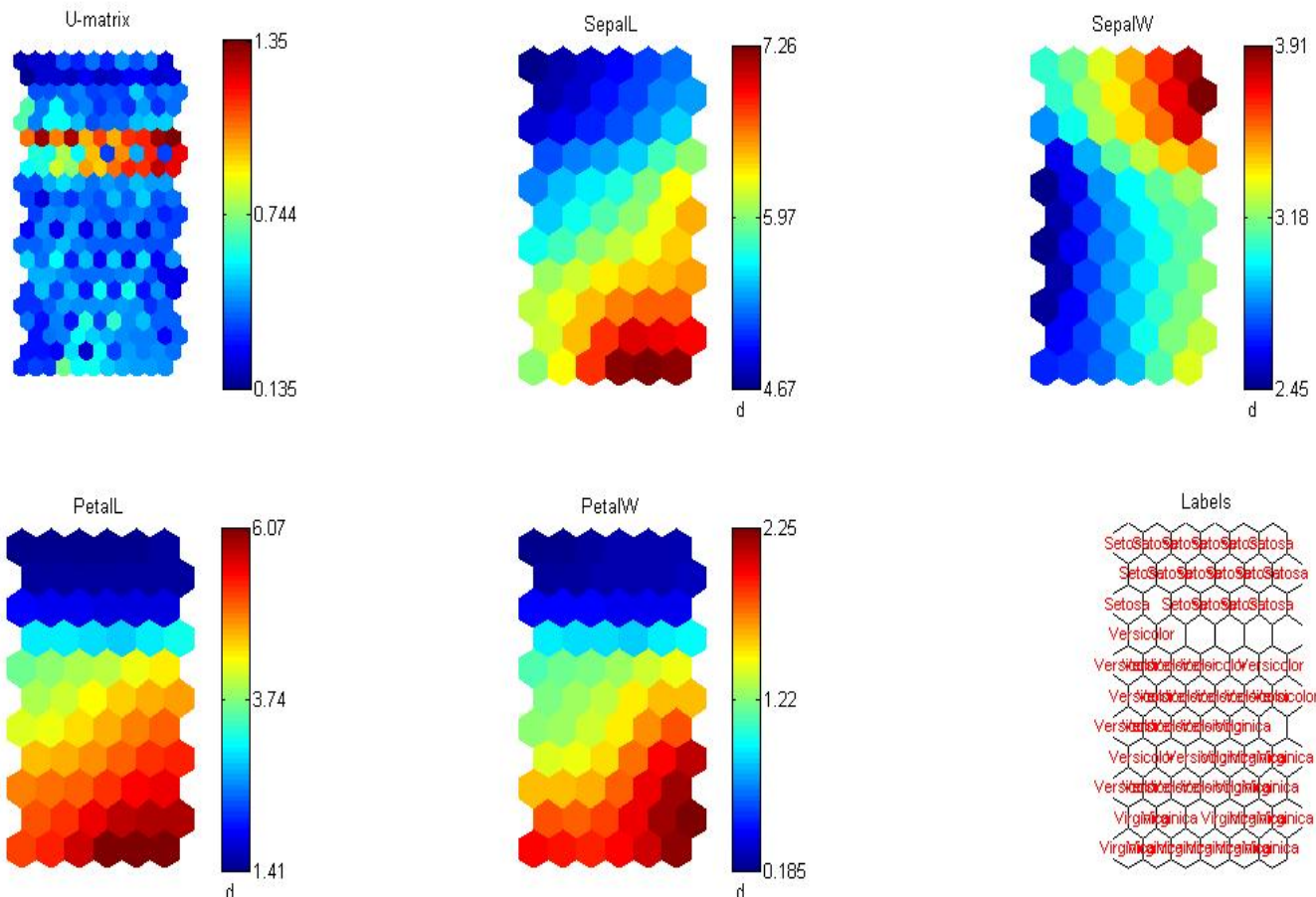
Результаты моделирования алгоритма кластеризации

Унифицированная матрица (англ. U-matrix) расстояний применяется для визуализации структуры кластеров, полученных в результате обучения карты. Ее элементы определяют расстояние между весовыми коэффициентами нейрона и его ближайшими соседями. Большое значение говорит о том, что данный нейрон сильно отличается от окружающих и относится к другому классу.



Результаты моделирования алгоритма кластеризации

Представление в виде компонентных плоскостей



Результаты моделирования алгоритма кластеризации

Индекс Дэвиса-Болдина

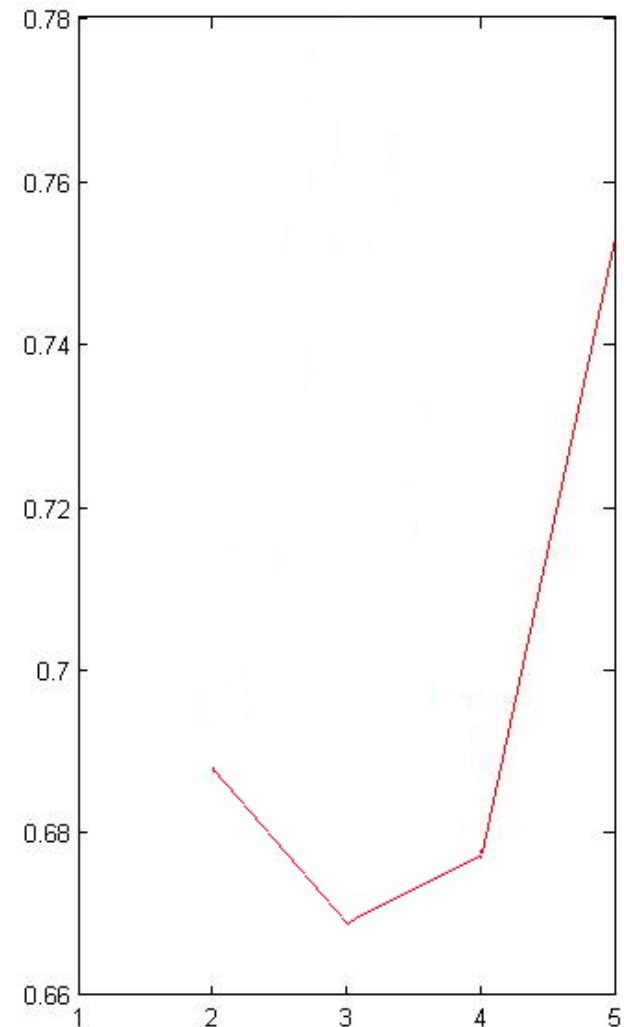
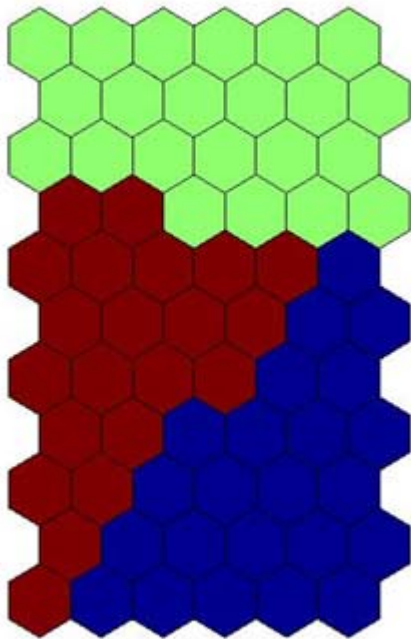
$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left\{ \frac{S_n(Q_i) + S_n(Q_j)}{S(Q_i, Q_j)} \right\}$$

где n – количество кластеров, S_n – среднее расстояние от объектов к центру кластера, $S(Q_i, Q_j)$ – расстояние между центрами кластеров.

Следовательно чем меньше значение индекса, тем кластеры компактнее и отдаление друг от друга.

Результаты моделирования алгоритма кластеризации

График зависимости значений индекса Дэвиса-Болдина от количества кластеров и конечное разбиение U-матрицы



Результаты моделирования алгоритма кластеризации

Тест №2

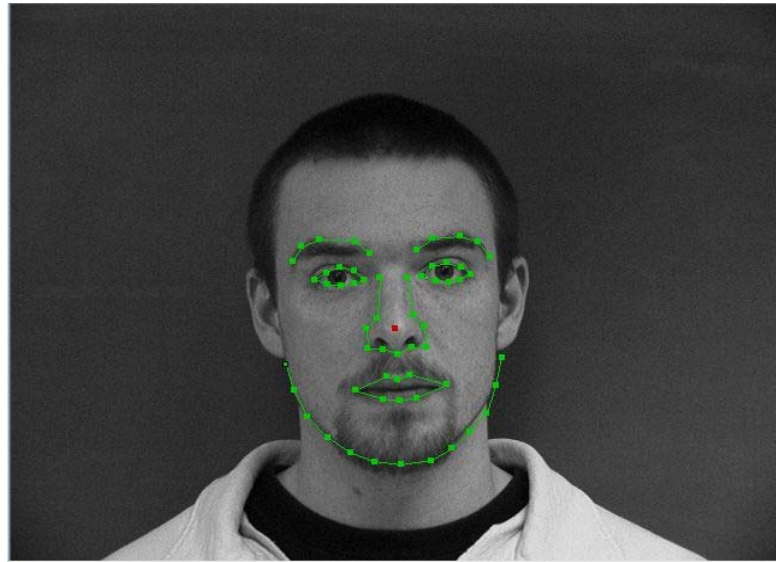
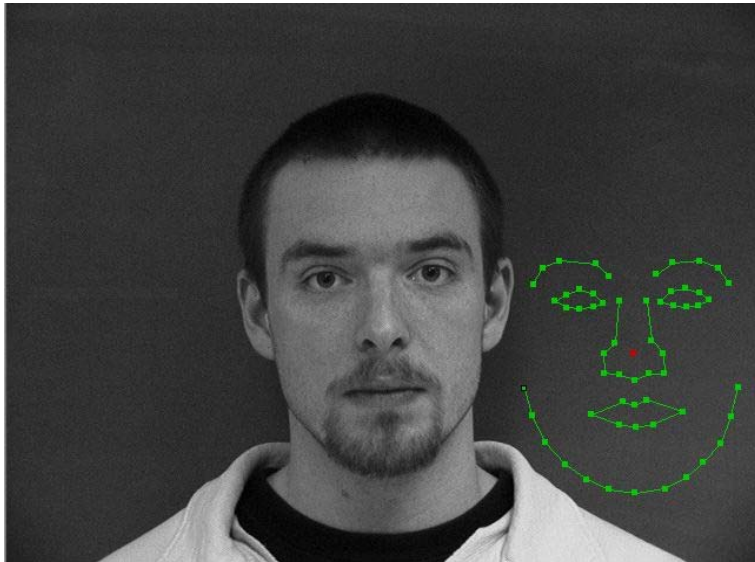
IMM Face Database

(Informatics and
Mathematical Modelling) -
база лиц, используемая
для задачи классификации
лиц.



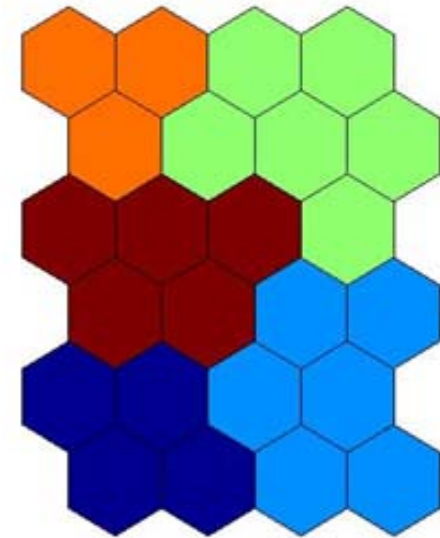
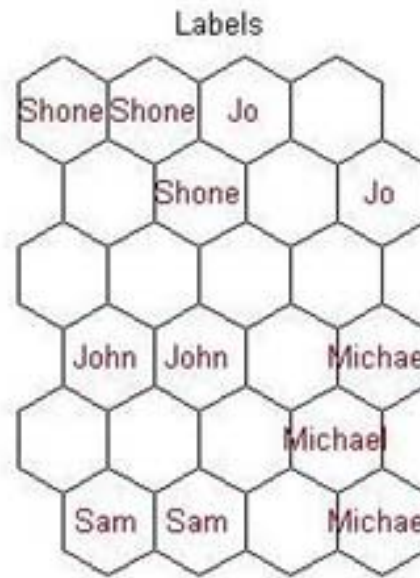
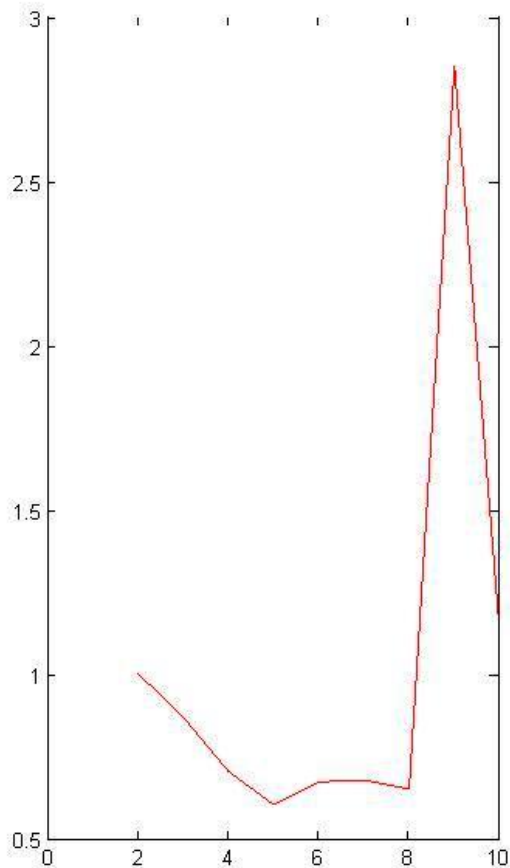
Результаты моделирования алгоритма кластеризации

С помощью метода Active Appearance Model (ААМ) строим модели лиц и получаем данные о каждом лице в виде координат 57 точек.



Результаты моделирования алгоритма кластеризации

График зависимости значений индекса Дэвиса-Болдина от количества кластеров и конечное разбиение U-матрицы



Результаты моделирования алгоритма кластеризации

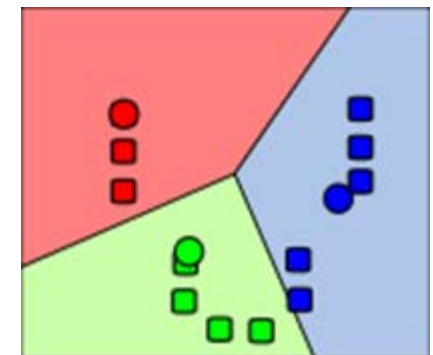
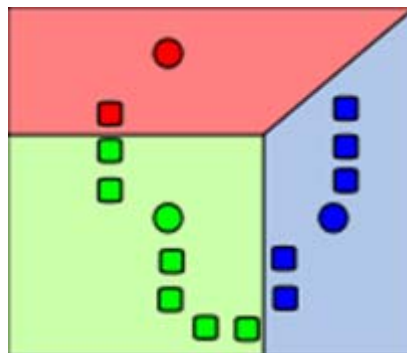
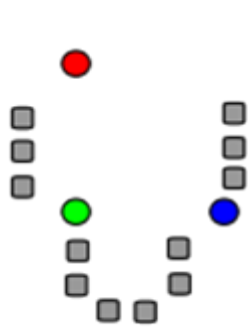
Метод К-средних

Исходные точки и случайно выбранные начальные центры.

Точки, отнесённые к начальным центрам. Разбиение на плоскости — диаграмма Вороного относительно начальных центров.

Вычисление новых центров кластеров

Предыдущие шаги повторяются, пока алгоритм не сойдётся.



Результаты моделирования алгоритма кластеризации

Сравнение методов SOM и K-средних на различных наборах данных, используя среднеквадратическую ошибку разбиения, ее среднеквадратическое отклонение, среднюю ошибку кластеризации и среднюю структурную ошибку для 100 независимых инициализаций

Набор данных	Метод	Среднеквадратическая ошибка разбиения	Среднеквадратическое отклонение	Средняя ошибка кластеризации	Средняя структурная ошибка
IRIS	SOM	86.67	0.33	9.22	0
	K-средних	91.35	25.76	15.23	18
IMM	SOM	27.11	0.21	1.39	0
	K-средних	27.80	7.63	2.51	9

- Интеллектуальная система поддержки принятия решений
- Кластеризация
- Self-Organizing Map
- Результаты моделирования алгоритма кластеризации
- **Выводы**

Выводы

□ **Преимущества**

- Простота реализации
- Проецирование многомерных пространств на двумерное
- Объекты с похожими характеристиками (по всем компонентам) располагаются рядом, образуя кластеры.
- Визуализация данных в виде двумерной карты позволяет проводить кластеризацию данных и значительно упрощает корреляционный анализ

Выводы

- Недостатки

- **Время тренировки**

- **Выбор параметров в зависимости от поставленной задачи**



Спасибо за внимание