

## Лабораторная работа № 6

### «Рассуждения на основе прецедентов. Решение задач классификации с помощью прецедентной системы»

#### 1. Цель работы

Изучить базовые понятия прецедентного подхода, ознакомиться с возможностями веб-приложения, реализующего классификацию данных с помощью вывода на основе прецедентов, и получить навыки работы с ним.

#### 2. Подготовка к работе

- Изучить основные определения и понятия метода рассуждения на основе прецедентов.
- Подготовить данные для классификации. Данные можно взять из хранилища UCI Machine Learning Repository (Калифорнийский университет, США). Для этого нужно перейти по следующей ссылке: <https://archive.ics.uci.edu/ml/datasets.php> и выбрать набор данных для классификации (желательно выбирать наборы данных с числовыми значениями атрибутов, в случае наборов данных с качественными признаками нужно самостоятельно до загрузки в систему перевести их в количественные).
- Зарегистрироваться на сайте: <https://nir.pythonanywhere.com/>

#### 3. Лабораторное задание

- Решить задачу классификации с использованием веб-приложения.
- Определить лучшее число соседей и метрику для выбранного набора данных.
- Рассмотреть случай работы с учетом коэффициентов важности параметров.
- Подготовить отчет для защиты лабораторной работы.
  - В отчет по лабораторной работе включить информацию об основных понятиях в прецедентном подходе, выбранном наборе данных, описание используемого алгоритма для классификации, результаты классификации с различными метриками и различным числом соседей.
  - Написать вывод по полученным результатам.

#### 4. Требования

- Лабораторная работа должна быть выполнена с использованием веб приложения <https://nir.pythonanywhere.com/>

#### 5. Методические указания

- Перед использованием приложения необходимо подготовить набор данных для последующей работы.

➤ Данные должны храниться в файле формата “.txt” или “.csv” и быть представлены в виде строки, содержащей набор параметров, разделенных через знак “;”. В конце строки содержится решение.

➤ Набор данных должен состоять из числовых параметров, при этом решение может быть представлено, как числовым, так и строковым типом.

➤ Примеры наборов данных: Ирисы - <https://archive.ics.uci.edu/ml/datasets/Iris>

<https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>

*крестики-нолики* - <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame> (заменить качественные значения атрибутов на количественные)

**Указания по выполнению работы см. в разделе “Описание системы”.**

## **6. Содержание отчета**

1. Цель работы
2. Задание
3. Краткое описание действий по пунктам
4. Таблицы
5. Выводы по каждой части задания

## **7. Литература**

1) Варшавский П.Р., Еремеев А.П. Моделирование рассуждений на основе прецедентов в интеллектуальных системах поддержки принятия решений // Искусственный интеллект и принятие решений. – №2. – 2009. – С. 45–57.

## **8. Контрольные вопросы**

- 1) Что такое прецедент?
- 2) Что такое вывод на основе прецедентов?
- 3) Общий вид параметрического представление прецедентов?
- 4) Какие методы извлечения прецедентов были рассмотрены в лабораторной работе? В чем суть данных методов?
- 5) Какие метрики были рассмотрены в лабораторной работе?

## Описание системы

Веб-приложение позволяет выполнять редактирование и анализ данных, сравнивать различные методы их обработки, с целью определения наилучшего для решения поставленной задачи.

Для работы приложения необходимо наличие БП. Для этого нужно создать структуру будущей БП и внести в нее прецеденты, а также можно **загрузить данные из файла** формата Comma-Separated Values (.csv) или Text (.txt), содержащего набор данных. В этом случае программа автоматически сгенерирует структуру БП и заполнит ее данными из файла (рис. 1). Также при создании БП необходимо распределить набор данных на обучающую и тестовую выборки (распределение в %). По умолчанию обучающая и тестовая выборки распределяются в соотношении 80 на 20 от исходного набора данных.

Параметр *	Тип *	От (Диапазон)*	До (Диапазон)*	Описание	Единицы измерения
	Вещественный	4.3	7.9		
	Вещественный	2	4.4		
	Вещественный	1	6.9		
	Вещественный	0.1	2.5		
	Текст				

Рис. 1. Создание БП

Рассмотрим работу приложения на примере решения задачи классификации.

1. Для этого необходимо на странице «Подбор числа  $k$  для метода  $kNN$ » (пункт меню «Извлечение») нужно выбрать созданную БП и провести эксперименты по подбору оптимального числа соседей для алгоритма  $kNN$ . *Изначально веса не учитывать.*
2. Провести те же операции с учетом весов параметров, результаты внести в таблицу.
3. На вкладке «с использованием базы “удачных” и “неудачных”» добавить новый прецедент для классификации, при этом использовать параметры алгоритма  $kNN$ , полученные на предыдущем шаге.
4. Добавить в БП “удачный” прецедент и проанализировать как изменится результат классификации на вкладке “Подбор числа  $k$  для метода  $kNN$ ”. Отразить это в отчете.

ирисы Тестовая выборка Показать содержимое

id	one	two	three	Col3	Col4
121	4.6	3.1	1.5	0.2	Iris-setosa
122	5.4	3.7	1.5	0.2	Iris-setosa
123	5.7	4.4	1.5	0.4	Iris-setosa
124	5.4	3.9	1.3	0.4	Iris-setosa
125	5.1	3.5	1.4	0.3	Iris-setosa
126	4.8	3.4	1.9	0.2	Iris-setosa
127	4.8	3.1	1.6	0.2	Iris-setosa

Подбор оптимального числа соседей (k) для метода kNN:

Метод кросс-валидации:

Максимальное число k:

Метрика:

Учитывать веса параметров:

Рис. 2. Подбор оптимального числа соседей для алгоритма

ирисы Обучающая выборка Показать содержимое

4	4.6	3.2	1.4	0.2	Iris-setosa
5	4.9	2.5	4.5	1.7	Iris-virginica
6	6.7	3.3	5.7	2.5	Iris-virginica
7	5.7	2.5	5.0	2.0	Iris-virginica
8	6.0	2.2	5.0	1.5	Iris-virginica
9	5.6	3.0	4.5	1.5	Iris-versicolor
10	5.5	3.5	1.3	0.2	Iris-setosa
11	6.2	2.9	4.3	1.3	Iris-versicolor

Выбрать прецеденты из таблицы  
 Выбрать классифицируемые данные из таблицы  
 Добавить классифицируемые данные

id	one (double [4.3, 7.9])	two (double [2.0, 4.4])	three (double [1.0, 6.9])	Col3 (double [0.1, 2.5])
1	0	0	0	0

Выбор алгоритма: 
 Количество соседей (k): 
 Выбор метрики: 
 Учитывать веса параметров:

Рис. 3. Классификация нового прецедента