ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ ИЕРАРХИЧЕСКОЙ КЛАССИФИКАЦИИ С ПОДДЕРЖКОЙ ДООБУЧЕНИЯ

Выполнил: студент 427 группы Бабаков А.В

Научный руководитель: Недумов Ярослав Ростиславович

Проблема иерархической классификации

Асимметричность $\forall c_i, c_j \in C$, *if* $c_i \prec c_j$ *then* c_j not $\prec c_i$.

Антирефлексивность: $\forall c_i \in C$, c_i not $\prec c_i$.

Транзитивность: $\forall c_i, c_j, c_k \in C$, $c_i \prec c_j$ *and* $c_j \prec c_k$ *imply* $c_i \prec c_k$.

Методы иерархической классификации

- “Плоская” иерархическая классификация.
- Методы иерархической классификации, основанные на локальных классификаторах:
  - LCPN (Local Classifier per Parent Node).
  - LCN (Local Classifier per Node).
  - LCL (Local Classifier per Level).
- Глобальные иерархические классификаторы.

Дообучение классификаторов

- Проблема:
  - Нет обучающий данных.
  - Гипотеза классификации должна изменяться с течением времени.
- Решение:
  - Дообучение классификатора по мере поступления данных.

Постановка задачи

- Исследовать существующие методы иерархической классификации и дообучения.
- Найти (или создать) тестовые данные для оценивания иерархического классификатора.
- Разработать иерархический классификатор с дообучением.
- Провести тестирование иерархического классификатора (провести измерение иерархической F-меры по мере дообучения).

Иерархическая F-мера

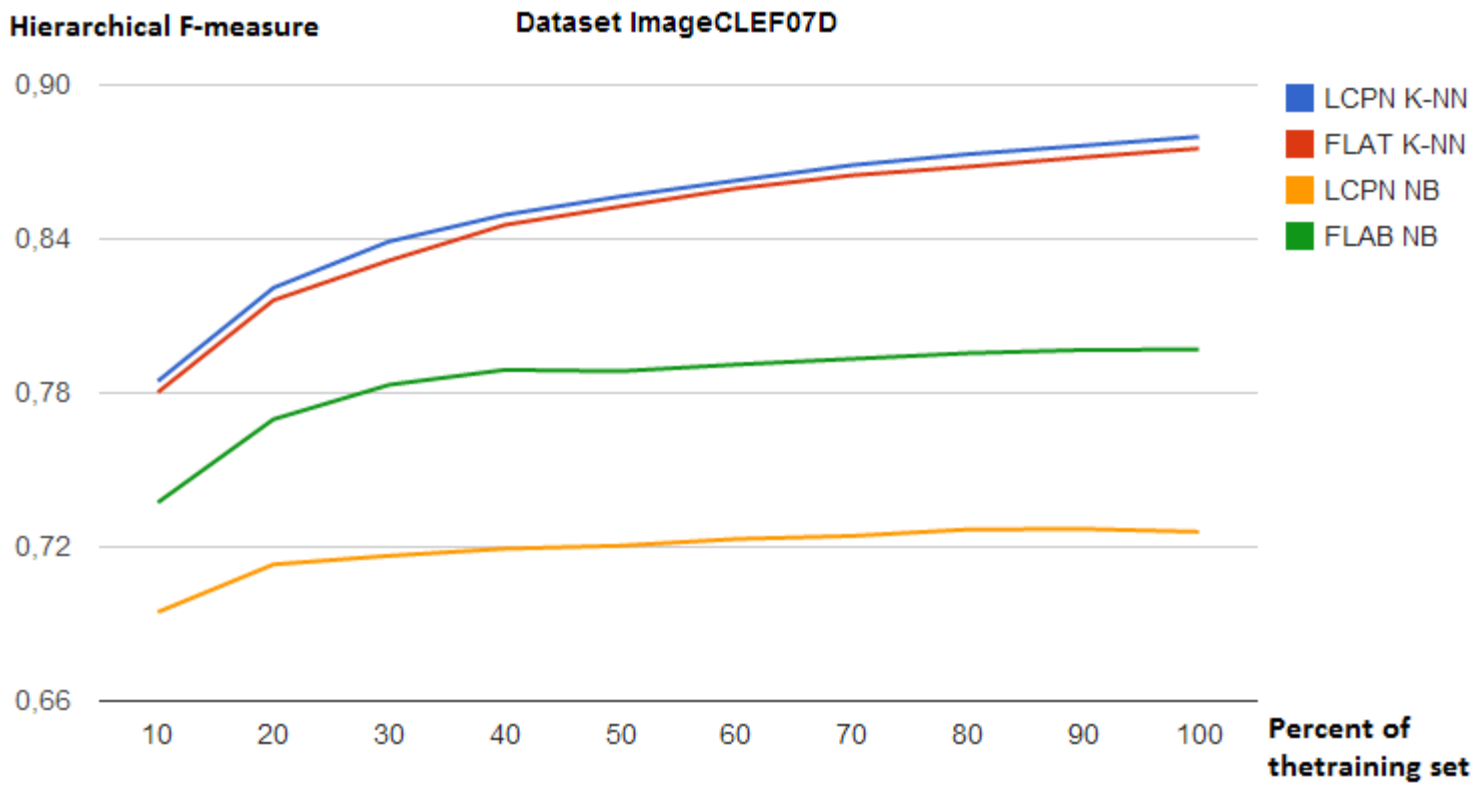
$$hP = \frac{\sum_i |\hat{P}_i \cap \hat{T}_i|}{\sum_i |\hat{P}_i|}, hR = \frac{\sum_i |\hat{P}_i \cap \hat{T}_i|}{\sum_i |\hat{T}_i|}, hF = \frac{2 * hP * hR}{hP + hR}$$

где $\hat{P}_i$ - это множество, состоящее из наиболее специфичных классов, к которым был отнесен тестовый объект i и все их родительские классы, $\hat{T}_i$ - это множество, состоящее из всех истинных наиболее специфичных классов для тестового объекта i и всех их родительских классов. Суммирование ведется по всем тестовым объектам. hP – иерархическая точность, hR – иерархическая полнота, hF – иерархическая F мера.

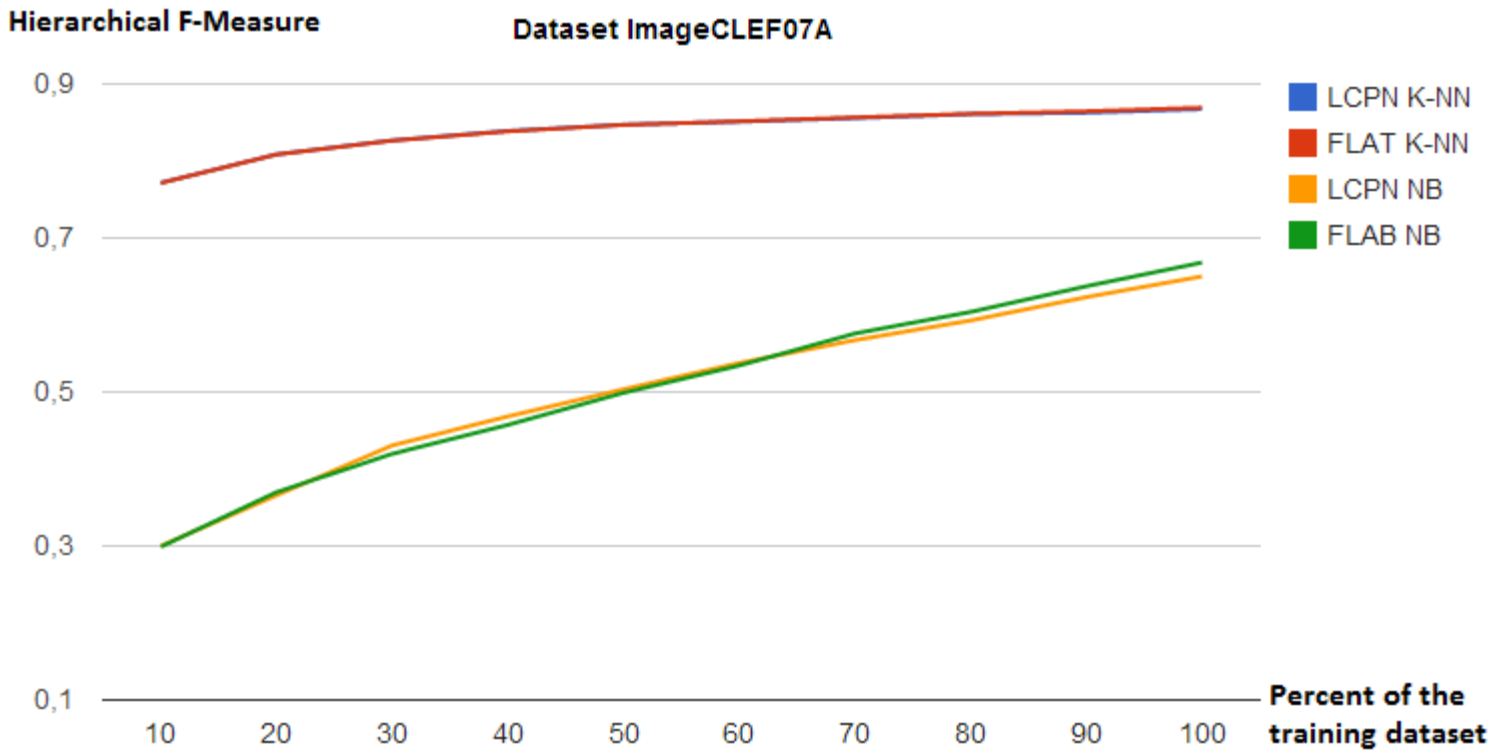
Описание проделанной работы

- Разработаны иерархические классификаторы с дообучением:
  - LCPN на основе KNN.
  - LCPN на основе NB.
  - Flat на основе KNN.
  - Flat на основе NB.
- Разработана система тестирования иерархических классификаторов.

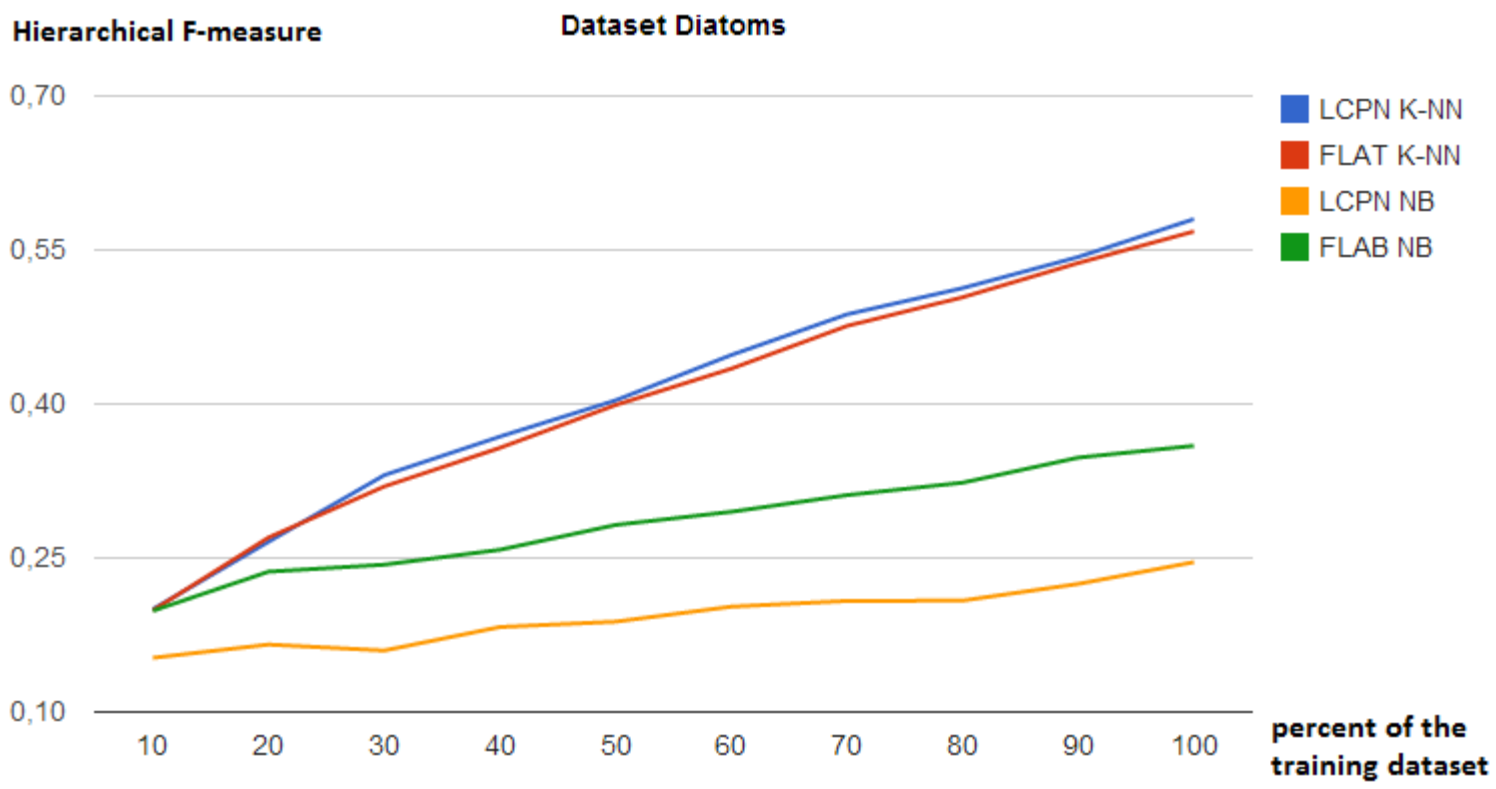
Результаты эксперимента 1



Результаты эксперимента 2



Результаты эксперимента 3



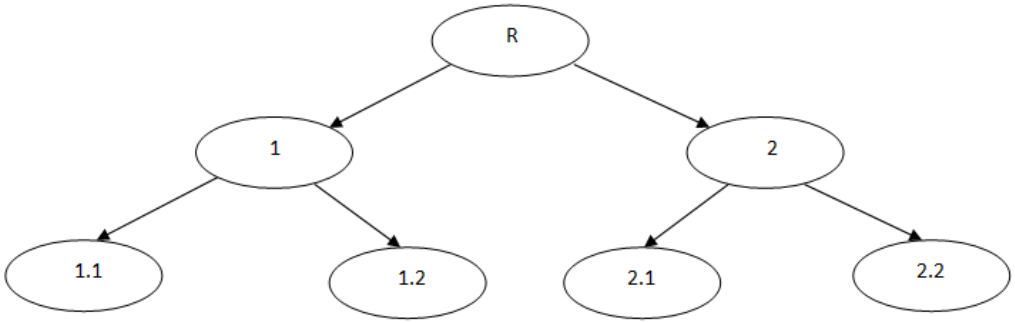
Результаты

- Исследованы существующие методы иерархической классификации и дообучения классификаторов.
- Разработаны методы иерархической классификации с дообучением на основе различных локальных классификаторов.
- Проведено тестирование .

Спасибо за внимание!

- Вопросы?

Почему FLAT NB лучше LCPN NB в экспериментах



Объект\Признак	1	2	3	4	5	6	Класс объекта
1	0	1	1	1	1	1	1.1
2	0	1	0	1	1	1	1.2
3	1	1	1	1	1	1	2.1
4	1	0	0	0	0	0	2.2
5	1	1	0	1	1	1	?

LCPN NB классифицирует объект 5 как принадлежащий к классу 1.2, в то время как FLAT NB классифицирует его в класс 2.2.