Московский государственный технический университет имени Н. Э. Баумана
(национальный исследовательский университет)

Выпускная квалификационная работа бакалавра

«Метод распознавания эмоций по звучащей речи на основе скрытой марковской модели»

Студент: Казаева Татьяна Алексеевна ИУ7-86Б

Научный руководитель: Строганов Юрий Владимирович

Москва, 2022 г.

Цель и задачи работы

Цель - разработать метод определения эмоций по звучащей речи на основе скрытой марковской модели

Задачи:

- проанализировать существующие эмоциональные корпусов и выбрать наиболее подходящий для обучения классификатора
- проанализировать информативные признаки, характеризующих речь и способы их выделения
- проанализировать классификатороы, чаще всего используемые в анализе речевых эмоций
- спроектировать и реализовать метод детектирования эмоций
- рассчитать качественные характеристики классификатора

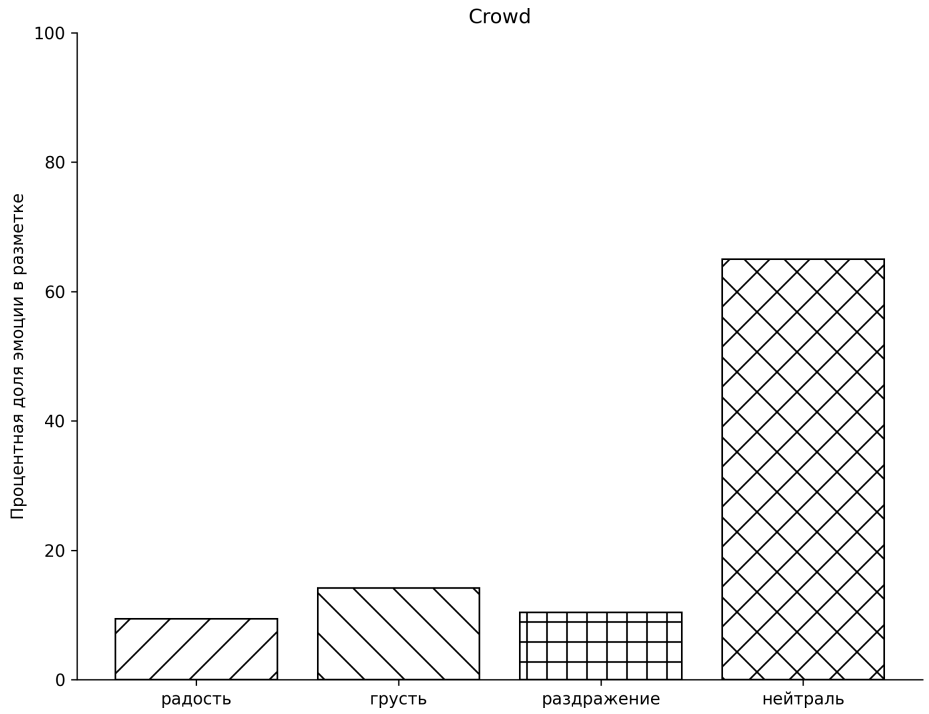
Определение эмоций

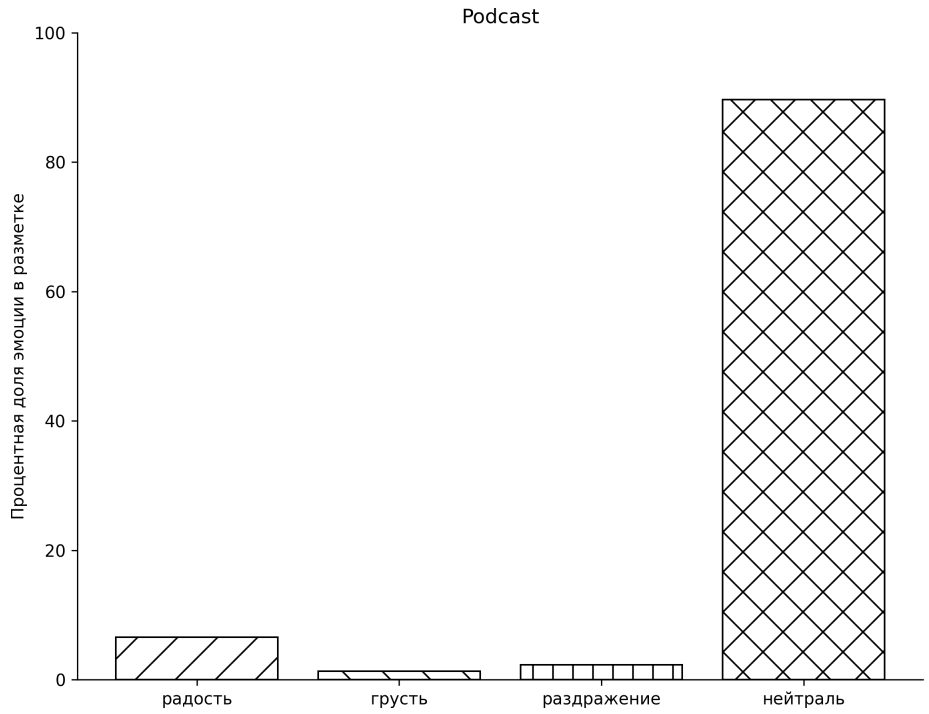
Подход	*Основа подхода*	*Категроии*	*Примеры решений*
Дискретный	Выделение базовых эмоций	Некоторый набор дискретных эмоций	Affectiva, RAVDESS, SAVEE, EmoDB...
Многомерный	Координатное многомерное пространство	Валентность, активация, интенсивность	RECOLA, колесо эмоций Плутчика
Гибридный	Комбинация дискретного и многомерного подходов	Базовые эмоции и уровни силы эмоций	"Песочные часы эмоций"

Корпуса звучащей речи

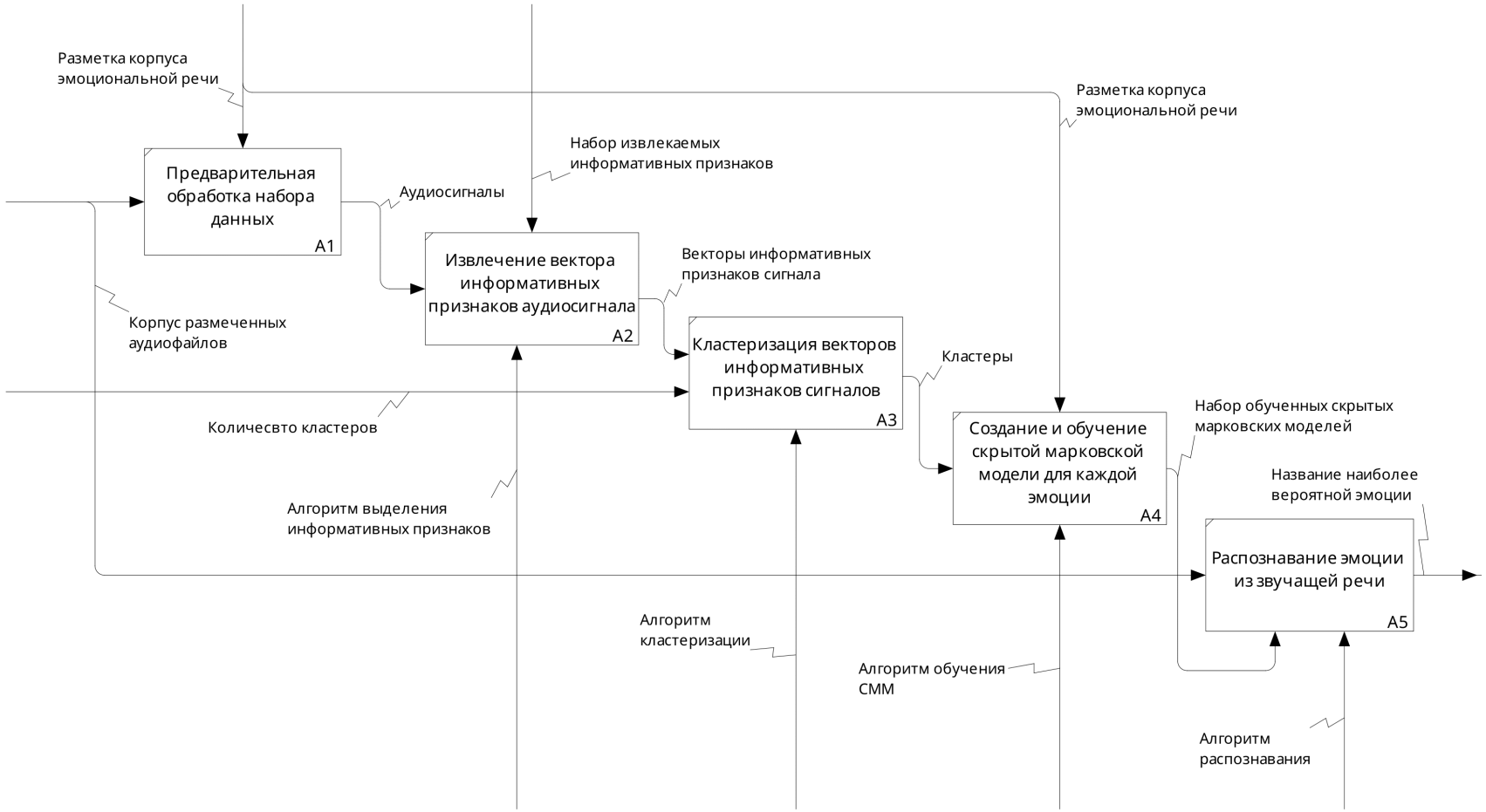
Название	*Количество эмоций*	*Количество голосов*		*Лексикон*	*Публичный*	*Поддержка русского языка*
		М	Ж			
RAVDESS	7	12	12	2 предл.	да	нет
SAVEE	6	4	0	15 предл.	да	нет
Emo-DB	6	5	5	10 предл.	да	нет
TESS	7	0	2	200 слов	да	нет
RUSLANA	4	12	49	10 предл.	нет	да
DUSHA	**4**	**-**	**-**	**обширный**	**да**	**да**
REC	-	-	-	обширный	нет	да

Распределение классов разметки корпуса DUSHA





Предлагаемый метод



Обучающий набор данных, составленный из корпуса DUSHA

В обучающий набор было включено 1500 аудиофайлов каждого класса разметки.

Подгруппа	*Всего*	*Тренировочная выборка*	*Тестовая выборка*
раздражение	1 ч. 02 мин. 47 сек.	50 мин. 05 сек.	12 мин. 41 сек.
нейтраль	1 ч. 02 мин. 23 сек.	50 мин. 02 сек.	12 мин. 20 сек.
радость	1 ч. 02 мин. 01 сек.	50 мин. 32 сек.	12 мин. 29 сек.
грусть	1 ч. 02 мин. 53 сек.	51 мин. 57 сек.	12 мин. 55 сек.

Шумоочистка к аудиофайлам не применялась.

Просодические признаки речи

Признаки оцениваются в баллах: 1 – низший балл, 3 – высший балл.

	Устойчивость к шуму	*Информативность*	*Емкость представления*
Частота основного тона	3	1	1
Интенсивность	3	2	3
Темп речи	3	3	3
Паузация	1	1	3

Спектральные признаки речи

Признаки оцениваются в баллах: 1 - низший балл, 3 - высший балл.

	Устойчивость к шуму	*Информативность*	*Емкость представления*
Мел-кепстральные коэффициенты	**2**	**3**	**3**
Частоты первых четырех формант	2	3	2
Джиттер, шиммер	1	1	1

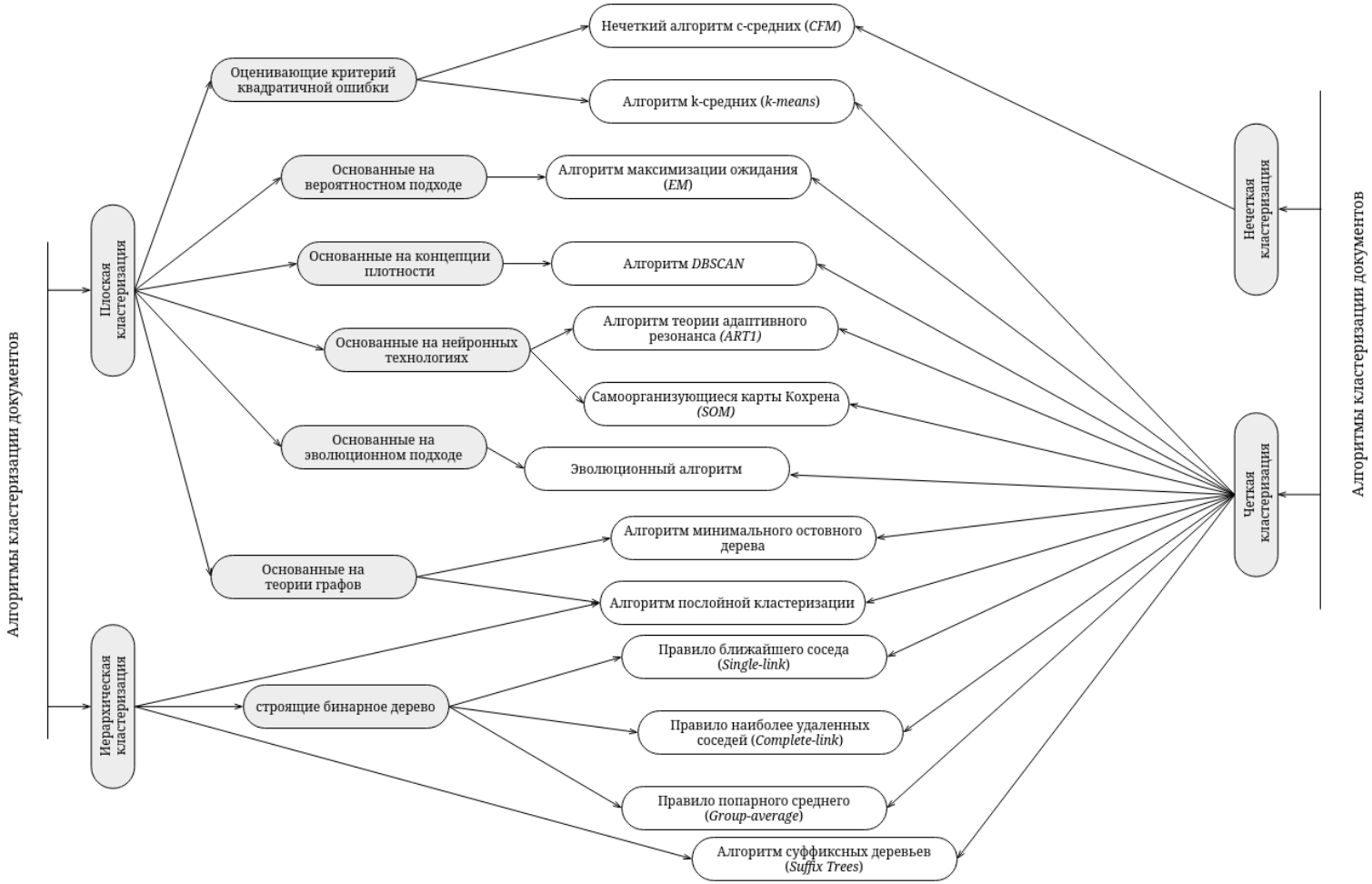
Мел-кепстральные коэффициенты

Мел-кепстральный коэффициент под номером m
вычисляется согласно:

$$c_j(m) = \sum_{m=0}^{M-1} T_j(m) \cos\left(\frac{\pi n \left(m + \frac{1}{2}\right)}{M}\right), \; 0 \leq n < M,$$

где $T_j(m)$ - логарифмическое значение энергии компонент спектра на выходе каждого треугольного мел-фильтра фильтра.

Классификация алгоритмов кластеризации



Классификаторы, наиболее часто используемые в аффективных вычислениях

Скрытая марковская модель (СММ)

- используется для моделирования *последовательностей данных*
- данные преобразуются в последовательность наблюдений

Искусственная нейронная сеть (ИНС)

- состоят из соединенных и взаимодействующих искусственных нейронов
- данные передаются через слои искусственных нейронов

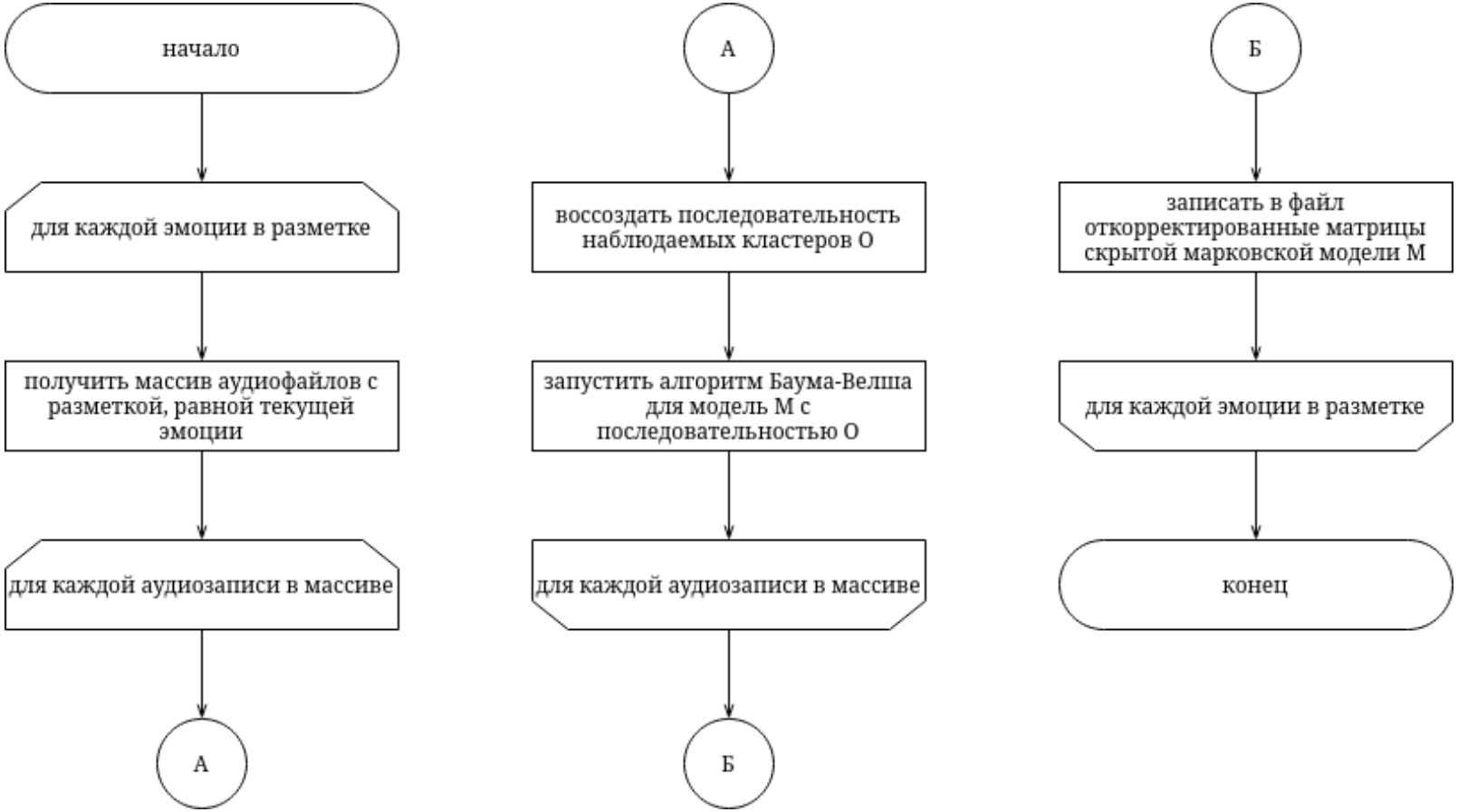
Скрытая марковская модель

Можно описать как *двойной стохастический процесс*:

проявление эмоции -
скрытый стохастический процесс, который невозможно наблюдать напрямую

наблюдаемый набор мел-кепстральных коэффициентов -
процесс, который создает последовательность наблюдений

Обучение скрытой марковской модели



Распознавание эмоций в речи

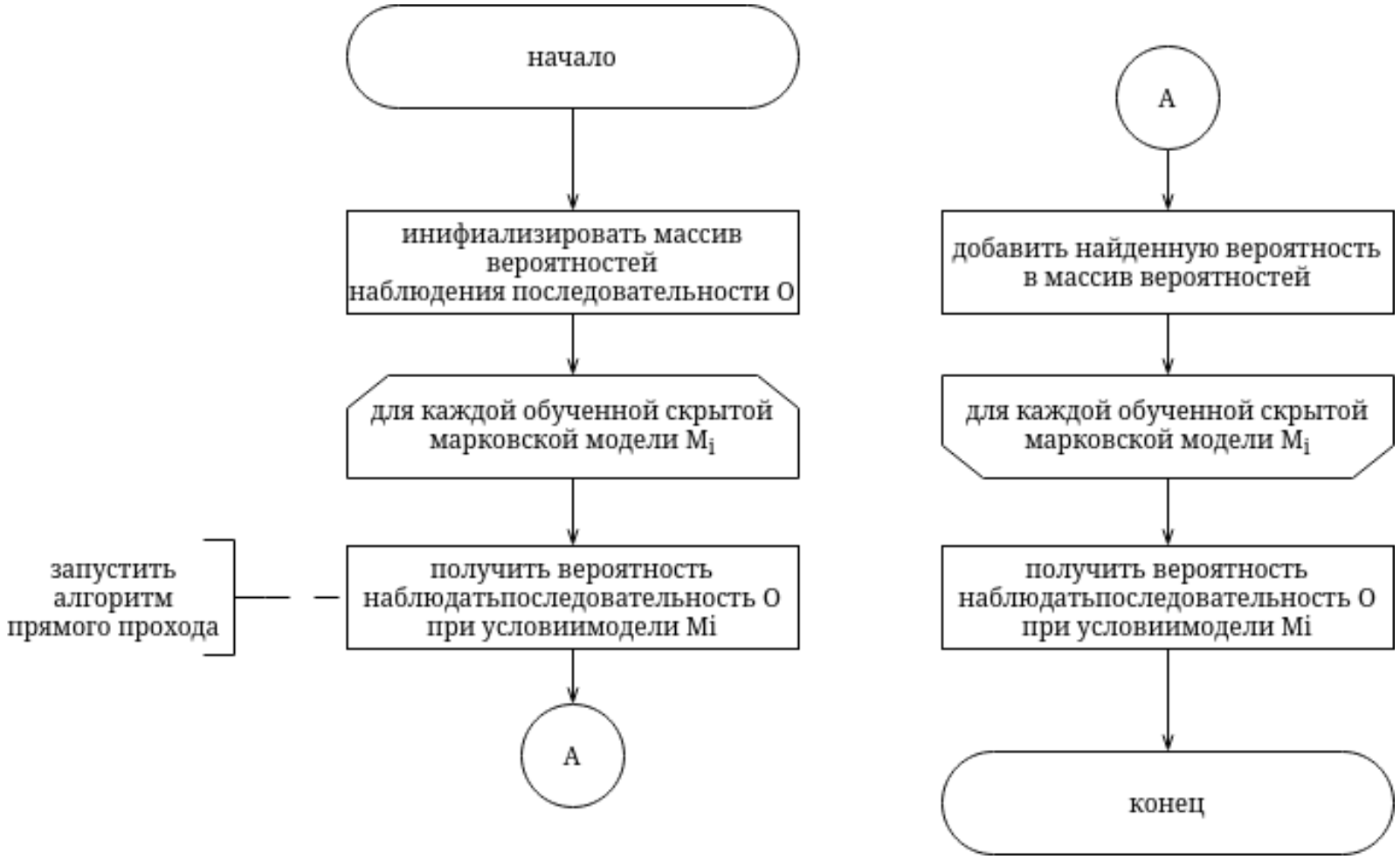
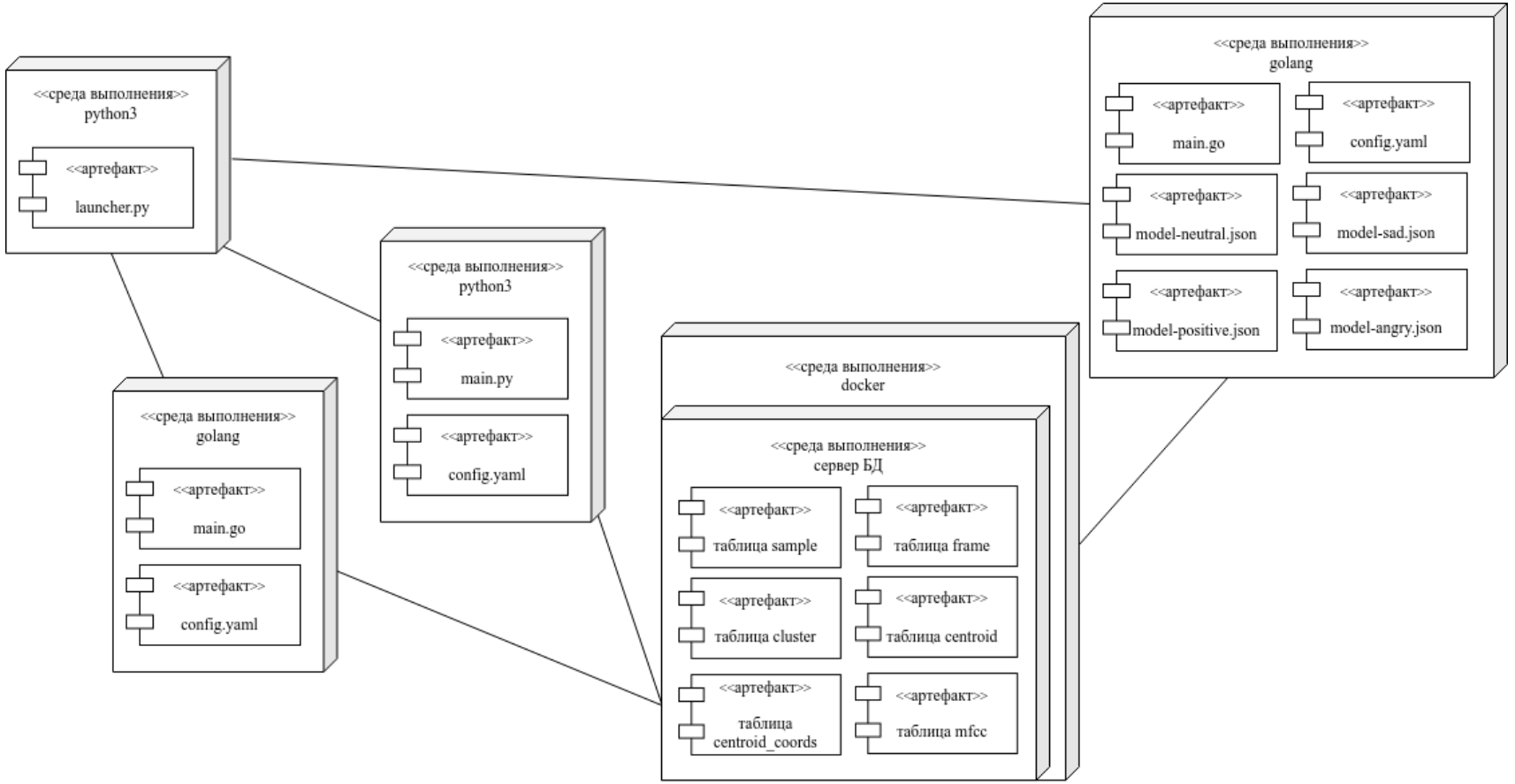


Диаграмма развертывания



Результат классификации на тренировочной выборке

Матрица неточностей для классификации на тренировочной выборке.

экспертная оценка	*оценка классификатора*			
	angry	*neutral*	*positive*	*sad*
angry	44.3%	6.2%	35.9%	13.6%
neutral	17.7%	11.8%	44.8%	25.8%
positive	30.8%	5.8%	50.3%	13.1%
sad	28.0%	7.3%	34.6%	30.1%

Оценка результата классификации (1/2)

"нейтраль"

	Positive	Negative
Positive	141	780
Negative	1059	4373

$$\text{Precision}_{\text{neutral}} = 0.38\%$$
$$\text{Recall}_{\text{neutral}} = 11\%$$
$$\text{F}_{\text{neutral}} = 28\%$$

"грусть"

	Positive	Negative
Positive	361	4491
Negative	462	799

$$\text{Precision}_{\text{sad}} = 36\%$$
$$\text{Recall}_{\text{sad}} = 30\%$$
$$\text{F}_{\text{sad}} = 32$$

Оценка результата классификации (2/2)

"злость"

	Positive	Negative
Positive	533	920
Negative	467	4095

$$\text{Precision}_{\text{anger}} = 37\%$$
$$\text{Recall}_{\text{anger}} = 44\%$$
$$\text{F}_{\text{anger}} = 36\%$$

"радость"

	Positive	Negative
Positive	603	1206
Negative	1007	4690

$$\text{Precision}_{\text{positive}} = 36\%$$
$$\text{Recall}_{\text{positive}} = 50\%$$
$$\text{F}_{\text{positive}} = 43\%$$

Выводы

На выборке из **6000 элементов** с развномерным распределением классов:

общая F-мера $\approx 35\%$
максимальная F-мера $\approx 43\%$

Класс, распознанный **наиболее качественно** - "радость" (полнота распознавания $\approx 50\%$),
наименее качественно - "нейтраль" (полнота распознавания $\approx 11\%$)

Заключение

Цель работы достигнута: был разработан и реализован метод распознавания эмоций по звучащей речи. Все поставленные задачи были выполнены:

- проанализированы русскоязычные и иностранные корпуса эмоциональной речи, для обучения классификатора был выбран корпус **DUSHA**
- проанализированы признаки, характеризующие эмоцию в речи, для классификации были использованы **мел-кепстральные коэффициенты**
- проведен обзор классификаторов, используемых в анализе речевых эмоций
- спроектирован и реализован метод детектирования эмоций
- с помощью качественных метрик (*F-мера*, *точность*, *полнота*) **оценен результат** классификации.

Дальнейшее развитие

сбор собственного корпуса звучащей речи, содержащего аудиозаписи **студийного качества**, озвученные профессиональными актерами

расширение объема информации в разметке: учет **интонационного контура** (ИК) для каждой аудиозаписи