

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики и
информационных технологий

**Методы распознавания образов в разработке железорудных
месторождений**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 421 группы
направления 09.03.01 «Информатика и вычислительная техника»
факультета компьютерных наук и информационных технологий
Черенкова Дмитрия Игоревича

Научный руководитель
к. ф.-м.н., доцент

Ю.А. Бродская

Заведующий кафедрой
к. ф.-м.н., доцент

Л.Б. Тяпаев

Саратов 2018 год

Введение. С каждым годом все более необходимо минеральное сырье в потребностях нашей промышленности. Это приводит к возникновению различных наиболее сложных задач, одна из которых обнаружение скрытых месторождений полезных ископаемых по косвенным признакам. Такой метод куда дешевле трудозатрат на поиск и разведку прогнозируемых рудных месторождений. Не во всех случаях можно дать точную оценку, выбранному участку, так как существует огромное число различных вариаций образования полезных ископаемых. Порой, даже именитые ученые не в силах ответить на вопрос об образовании породы. Однако, некоторые способы очень часто встречаются и имеют схожий набор признаков, отталкиваясь от этих данных можно строить наши прогнозы. В практику геологоразведческих работ в настоящее время фундаментально вошли математические методы и электронные вычислительные устройства. Таким образом с помощью вычислительной техники, можно автоматизировать процессы по распознаванию чего-либо.

Цель данной дипломной работы в том, чтобы показать, как можно использовать идею распознавания образов в промышленных целях, на примере распознавания образов полезных ископаемых. В данной отрасли, наиболее важно возобновление природных ресурсов, иначе мы можем истощить все ныне действующие карьеры.

Задачи заключаются в изучении теории распознавания образов, анализе различных мест применений в промышленности, а так же в создании и написании программы распознавания образов железорудных ископаемых. Приложение должно удовлетворять требованиям:

- быстрая работа программы;
- удобный пользовательский интерфейс;
- возможность выбора алгоритма распознавания;
- поддержка кроссплатформенности (работы из разных операционных систем);

Работа состоит из введения, трех частей, заключения, списка используемой литературы и двух приложений.

В первой части рассмотрены основные понятия, и простейшую реализацию задач распознавания образов.

Во второй части рассмотрены такие понятия как признаки, класс, классификация, прогнозирование, разобрана обработка исходных данных и виды их представлений.

Третья часть является практической частью, в которой описана разработка приложения распознавания образов.

1 Основные понятия и определения. В современном мире существует множество задач, требующих принятия решений в зависимости фотографии объекта или отталкиваясь от его признаков.

Рассмотрим основные элементы модели классификации.

Класс — множество объектов, имеющих общие свойства. Объекты внутри класса должны обладать "схожестью". Количество классов в задаче распознавания не ограничено и зависит от нужных объектов имеющие одинаковые наборы свойств.

Классификация — процесс назначения объекту некоторого класса. Классификатор — устройство, которое по набору признаков, определяет принадлежность классу.

Верификация — сопоставление экземпляра объекта с эталонной моделью для выбранного класса.

Образ — наименование области в пространстве признаков, содержащая множество объектов. Признак — количественное описание свойства исследуемого объекта.

Пространство признаков — N-мерное пространство определяемое для заданной задачи распознавания, где N фиксированное число измеряемых признаков для любых объектов.

Распознавание образов можно определить, как выбор класса для объекта с помощью анализа параметров это объекта. Под анализом будем понимать выделение только тех признаков, что находятся в эталонной модели, а так же их оценка.[20]

2 Классификация методов распознавания образов. Более детально остоновимся на существующих трех группах методов распознавания образов:

- Сравнение с образцом. В эту группу входит классификация по ближайшему среднему, классификация по расстоянию до

ближайшего соседа. Также различные структурные методы распознавания.[21]

- Статистические методы. Статистические методы используют некоторую статистическую информацию при решении задач распознавания. Метод определяет принадлежность объекта к конкретному классу на основе вероятности. В ряде случаев это сводится к определению апостериорной вероятности принадлежности объекта к определенному классу, при условии, что признаки этого объекта приняли соответствующие значения.[8,14]
- Нейронные сети. Обособленный класс методов распознавания. Отличается особенностью самообучения. Но и является самым сложным из перечисленных.[5]

3 Алгоритм распознавания. Любые эвристические методы распознавания построены на основе подражания действиям специалиста в этой области, можно выделить схему действий специалиста, который изучает данные об объекте исследования и систематизирует их. Основные этапы распознавания: построение классификаций и классифицирование объектов.

Построение классификации отражает уровень наших знаний об объекте и по этим знаниям можно группировать объекты по принципу сходства и различия. В этом и есть вся суть классификации — сравнение объектов между собой, между эталонными объектами. Тут и происходит первое отличие вычислительной машины от человека. Машина всегда определяет вероятность принадлежности классу, когда человек пользуется интуитивными соображениями на основе своих суждений. Очень просто определить класс, когда у нас пара признаков, и очень сложно, когда их сотни. Данный процесс можно описать в качестве черного ящика, у которого есть известные входные параметры объектов, выходом будет является

принадлежность объектов к какому-либо классу, и каким-то набором не известных нам методов сравнения этих объектов. [1]

Построение классификации исследуемых объектов, а также классификация новых, составляет единый процесс распознавания. При этом когда новые объекты пополняют классификацию, они так же влияют на содержимое образа или абстрактной модели. Поэтому прежде, чем распознавать новые объекты необходимо грамотно построить классификацию эталонных объектов, чтобы последующие новые не сильно влияли на построенную модель. если же влияние будет значительное, то есть вероятность ошибки при классифицировании нового набора значений. Проблема распознавания образов очень глобальна, включает в себя такие элементы как классифицирование, построение эталонной модели и обучение. Обучение необходимо для плавной адаптации смены эталонной модели, на основе распознанных объектов. [2]

4 Организация исходных данных. В первую очередь при постановке задачи стоит определиться с объектами исследования. Так как данная работа затрагивает обширную область геологии, то объекты исследования могут быть от химических элементов до земной коры в целом. В данной работе под объектом будем понимать участки земной коры разных размеров, вычисляемые с целью получения прогноза полезных ископаемых. В использовании геологических карт под объектами обычно понимают квадратные площади. Но не стоит забывать о том, что наши элементы поиска руды, должны представлять собой элементы исследуемых объектов. Так определяется основное множество объектов и направление сбора необходимой информации.[3,7]

На саму структуру влияет целевое задание. Первое ограничение которое может там накладываться это указание цели : медь, железо, золото и т.д. Это влечет за собой ряд признаков характерных только для конкретной

цели, и характером проявления данных месторождений. Именно тут максимально используется накопленные знания геологов в результате исследований выбранных целевых объектов. Возможно, определить качественный и количественный состав месторождения. Качественное прогнозирование — это фактически ответ на вопрос, если ли в данной области целевые месторождения, если ответ положительный, то к данной области можно попробовать применить количественный прогноз, для определения примерной массы этих месторождений. Структура правильно организованного признакового пространства в общем виде должна быть иерархической, а для решения когерентных вопросов необходимо использовать те или иные ее части. Например, региональные признаки необходимы для оценки обнаружение там месторождений, а локальные для оценки промышленной значимости объектов.[19]

5 Формальные типы признаков. Признаки описывающие геологические группы объектов разделяют на две группы.

- Метрические. Образуют свойства, выражение которым можно осуществить на качественном, полуколичественном и количественном уровнях.[2]
- Не метрические. Образует свойство измеримое по шкале наименований. Обычно это совокупность метрических свойств, которые можно объединить. Например текстура, включает в себя: размер зерен, состав, их взаимодействие и т. д.[9]

За единицу характеристики метрических свойств объекта принято само свойство в целом, то качественные признаки описывают способность обладать этим свойством в целом, например магнитный - не магнитный. Полуколичественные признаки получаются делением свойств на крупные части, которые отображают степень или интенсивность проявления свойства. Данные величины проще описать количественными значениями, поэтому они

и получили название полуколичественные. Количественные признаки объекта, это что можно точно описать значением, например плотность породы, такие признаки помогают более детально описать объекты. На практике же, часто встречается не точные измерения объектов что приводит к количественные признаки выражается в полуколичественные или качественные. Но тут все зависит от важности данного признака, в некоторых случаях, это даже полезно, сильно упрощает работу по классификации.[4,6,10-14]

6 Практические способы реализации задачи распознавания железорудных ископаемых. В практике будем отталкиваться от ранее написанной теории. Разберем несколько алгоритмов распознавания образов в данной сфере. Для дальнейшей классификации определим сами классы. Возьмем несколько эталонных образцов и выделим самые необходимые признаки.

Программа реализована на высокоуровневом языке программирования Java. Будем использовать эталонные объекты как сущности.

Основываясь на таблице из приложения, определим класс сущности минерал, который содержит в себе, все поля таблицы. В данной таблице представлены лишь часть значимых признаков для распознавания железорудных месторождений. Остальные не так значительны при задаче распознавания. Для любой другой предметной области можно так же выделить признаки явно определяющие классификацию объектов, и конечно же эталонные объекты для распознавания образов.

Дополним структуру еще одним параметром - точность классификации выставив который можно повышать точность классифицируемых объектов, но есть и свои минусы, при привышении данной точности, мы получаем объект который нельзя классифицировать, то есть как был не опознанным так им и останется.

7 Архитектура Web-приложения и стек использованных технологий. Основой серверной части в данном приложении выступает Spring Boot. Это фреймворк для работы с клиент-серверными приложениями включающий в себя такие технологии как: Spring MVC, Apache Tomcat, JDBC API и другие. Система сборки проекта — Apache Maven. [16]

В качестве пользовательского интерфейса возьмем Vaadin. Данный стек технологий позволяет построить полноценное Web-приложение.

Apache Maven — фреймворк для автоматизации сборки проектов, специфицированных на XML-языке POM(Project Object Model).[17]

Maven — это инструмент для сборки Java проекта: компиляции, создания jar, создания дистрибутива программы, генерации документации

Spring Framework обеспечивает комплексную модель разработки и конфигурации для современных бизнес-приложений на Java — на любых платформах. Spring обеспечивает решения многих задач, с которыми сталкиваются Java-разработчики и организации, которые хотят создать информационную систему, основанную на платформе Java.[18]

Vaadin — свободно распространяемый фреймворк для создания RIA-веб-приложений, разрабатываемый одноимённой финской компанией. В отличие от библиотек на Javascript и специфических плагинов для браузеров, Vaadin предлагает сервер-ориентированную архитектуру, базирующуюся на Java Enterprise Edition.[15]

Заключение. При написании данной работы была изучена соответствующая литература, с помощью которой удалось выполнить поставленную задачу. Было реализовано Web-приложение на высокоуровневом языке программирования Java с помощью, которого основываясь на геологических данных возможно классифицировать полезные ископаемые. Так же изучена основная структура построения Web-приложения и различные фреймворки для работы с серверной и клиентской

частями. Работа актуальна, потому что затрагивает сферу деятельности человека с мало возобновляемыми ресурсами.

Имея вычислительную мощь и знания геологов накопленные годами, возможно обнаруживать скрытые новые месторождения полезных ископаемых.

В каждой части дипломной работы были рассмотрены различные аспекты теории распознавания образов.

В ходе выполнения данной работы были изучены методы распознавания образов и их применение в геологии, так же были решены следующие задачи:

- изучены методы распознавания образов;
- рассмотрены основные существующие алгоритмы теории распознавания образов;
- рассмотрена методы распознавания геологических объектов;
- выбраны методы для практической реализации;
- изучена структура Web-приложения;
- сделана программная реализация выбранных алгоритмов;
- подобраны тестовые примеры.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Г. С. Федосеев, В. В. Бабич, В. В. Зайков, В. И. Лебедев, Н. А. Плохих. Распознавание образов в задачах качественного прогноза рудных месторождений. Наука, Новосибирск. 1980 .
- 2 Г. С. Поротов. Математические методы моделирования в геологии. Санкт-Петербург. 2006 .
- 3 И. И. Абрамов, Л. Н. Дуденко, Ю. И. Михайлов. Математизация в геологии и металлогенический прогноз. Новосибирск. 1972 .
- 4 Ю. А. Воронин, Б. К. Алабин, С. В. Гольдин. Геология и математика. Основные методологические, теоретические и организационные вопросы применения математических методов и ЭВМ в геологии. Наука, Новосибирск. 1967 .
- 5 Уоссермен Ф. Нейрокомпьютерная техника. Мир, Москва. 1992.
- 6 М. Давид. Геостатистические методы при оценке запасов. Недра, Санкт-Петербург. 1980.
- 7 А. Б. Каждан Математические методы в геологии. Недра, Москва. 1990.
- 8 А. Б. Каждан Математическое моделирование в геологии и разведке полезных ископаемых. Недра, Москва. 1979.
- 9 П. Джурье, Т. Айзенауэр Распознавание образов в химии. Мир, Москва. 1977.
- 10 И. К. Волчанская, Ю. М. Фомичев. Математические методы при прогнозе рудоносности. Наука, Москва. 1977.
- 11 Ю. Л. Васильев, А. Н. Дмитриев. Применение математических методов и ЭВМ для решения прогнозных задач нефтяной геологии. Новосибирск. 1973.
- 12 А. Н. Бугаец, А. П. Малак, Ю. А. Садовский. Применение методов дискретного анализа при оценке месторождений полезных ископаемых. Новосибирск. 1970.

- 13 Н. Н. Боровко. Применимость и эффективность различных схем распознавания в типичных геологических задачах. Недра, Санкт-Петербург. 1972.
- 14 П. А. Рыжов. Математическая статистика в гоном деле. Москва, МИГРЭ. 1965.
- 15 Фреймворк для создания пользовательского интерфейса Vaadin. [Электронный ресурс]: (на 10 мая 2018 года) URL: <https://vaadin.com>
- 16 Felipe Gutierrez. Pro Spring Boot. 2016
- 17 Фреймворк для автоматизации сборки проектов [Электронный ресурс]: (на 10 мая 2018 года) URL: <https://maven.apache.org/>
- 18 Фреймворк Spring MVC [Электронный ресурс]: (на 10 мая 2018 года) URL <https://docs.spring.io/spring/docs/current/spring-framework-reference/web.html>
- 19 Ю. И. Журавлев. Распознавание. Классификация. Прогноз. Математические методы и их применение. Наука, Москва. 1989.
- 20 Задача распознавание образов. [Электронный ресурс]: (на 10 мая 2018 года) URL: <https://www.intuit.ru/studies/courses/2265/243/lecture/6241>
- 21 Математические методы распознавания образов. [Электронный ресурс]: (на 10 мая 2018 года) URL: <http://www.ccas.ru/frc/papers/mestetskii04course.pdf>