Федеральное государственное автономное образовательное учреждение высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"

Факультет компьютерных наук
Департамент анализа данных и искусственного интеллекта

Рабочая программа дисциплины Алгоритмы и структуры данных – 2

для образовательной программы «Прикладная математика и информатика»
направления подготовки 01.03.02 «Прикладная математика и информатика»
бакалавра

Разработчик(и) программы
Объедков С.А., к.т.н., sobiedkov@hse.ru

Одобрена на заседании департамента анализа данных и искусственного интеллекта
«___»_____ 2016 г.
Руководитель департамента
С.О. Кузнецов_____

Утверждена Академическим советом образовательной программы
«___»_____ 2016 г., № протокола_____

Академический руководитель образовательной программы
А.С. Конушин _____

Москва, 2016

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.

1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Алгоритмы и структуры данных – 2», учебных ассистентов и студентов направления подготовки 01.03.02 «Прикладная математика и информатика», обучающихся по образовательной программе «Прикладная математика и информатика».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 01.03.02 «Прикладная математика и информатика»;
- Образовательной программой «Прикладная математика и информатика» направления подготовки 01.03.02 «Прикладная математика и информатика» бакалавра.
- Объединенным учебным планом университета по образовательной программе «Прикладная математика и информатика», утвержденным в 2016 г.

2 Цели освоения дисциплины

Целями освоения дисциплины «Алгоритмы и структуры данных» являются ознакомление студентов с основами алгоритмической теории сложности, приближенными и вероятностными методами решения труднорешаемых задач, в том числе задач, возникающих в анализе данных.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Уровни формирования компетенций:

РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);

СД – способы деятельности, составляющие практическое ядро данной компетенции;

МЦ – мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен вести исследовательскую деятельность, включая анализ проблем, постановку целей и задач, выделение объекта и предмета исследования, выбор способа и методов исследования, а также оценку его качества	УК-6	РБ	Умеет проводить анализ корректности и временной сложности алгоритмов; распознает класс сложности задач.	Лекции, практические занятия, домашние задания	Домашние задания, контрольные работы, экзамен
Способен выполнить программную	ПК-2	РБ, СД	Способен программировать, отлаживать и тестировать	Лекции, практические	Домашние

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
реализацию математического метода, алгоритма, модели данных, описанных в научно-технической публикации или техническом задании			алгоритмы и структуры данных.	занятия, домашние задания	задания, контрольные работы
Способен анализировать тексты и документы по математике и компьютерным наукам на русском (государственном) и английском языках	ПК-11	РБ, МЦ	Способность находить и самостоятельно осваивать нужную информацию из общедоступных источников	Домашние задания	Домашние задания, контрольные работы, экзамен
Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и аппарат математических и компьютерных наук	ПК-15	РБ, СД	Умеет формализовать условие задачи, требующей алгоритмического решения, сформулировать и объяснить доказательство теоремы в устной и письменной форме	Лекции, практические занятия, домашние задания	Домашние задания, контрольные работы, экзамен
Способен реализовать законченную программную систему с использованием технологий баз данных, параллельных вычислений, пользовательских и сетевых интерфейсов	ПК-17	РБ, СД	Умеет разбить задачу на подзадачи, эффективно реализовать программные компоненты для отдельных подзадач и связать их воедино.	Практические занятия, домашние задания	Домашние задания

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к профессиональному циклу дисциплин, является базовой для подготовки бакалавра по направлению 01.03.02 «Прикладная математика и информатика».

Изучение данной дисциплины базируется на следующих дисциплинах:

- Дискретная математика
- Основы и методология программирования
- Алгоритмы и структуры данных – 1

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Способен учиться, приобретать новые знания, умения, в том числе в области, отличной от профессиональной;
- Способен критически оценивать и переосмыслять накопленный опыт (собственный и чужой), рефлексировать профессиональную и социальную деятельность;
- Способен провести теоретическую и экспериментальную оценку математического метода, алгоритма, модели данных;

- Способен использовать объектно-ориентированный подход для разработки архитектуры программного модуля или системы;
- Знает о наиболее важных алгоритмах и структурах данных и основных принципах их проектирования и анализа;
- Умеет обосновывать корректность алгоритмов, проводить теоретическую и экспериментальную оценки их временной сложности;
- Имеет навыки реализации алгоритмов на языках Python и C++.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

- Машинное обучение – 1
- Методы оптимизации
- Информационный поиск
- Машинное обучение на больших данных
- Дополнительные главы теории алгоритмов
- Сложность вычислений и логика в теоретической информатике – 1
- Основные методы анализа данных
- Машинное обучение и майнинг данных

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы				Самостоя-тельная работа
			Лекции	Семинары	Практические занятия	Другие виды работы[1]	
1	Основы теории вычислительной сложности	48	10		10		28
2	Методы решения труднорешаемых задач	28	6		6		16
3	Задачи и алгоритмы анализа данных	76	16		16		44

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год	Кафедра/подразделение	Параметры
		1		
Текущий	Домашнее задание	*		Решение задач в компьютерной системе и письменное решение задач с последующим устным объяснением. Задания разбиты на части, для каждой из которых указывается срок выполнения.
Итоговый	Экзамен	*		Устный экзамен, 4 часа.

[1] Указать другие виды аудиторной работы студентов, если они применяются при изучении данной дисциплины.

7 Критерии оценки знаний, навыков

При прохождении контроля студент должен продемонстрировать владение различными техниками проектирования и анализа алгоритмов, навыки программирования и тестирования алгоритмов и структур данных, а также умение формулировать и доказывать математические утверждения.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

8 Содержание дисциплины

Раздел 1 Основы теории вычислительной сложности

Машины Тьюринга. Классы P, NP, coNP, примеры задач из этих классов. NP-полнота. Полиномиальные сведения. Теорема Кука – Левина. NP-полные задачи: 3-SAT, Not-All-Equal-3-SAT, задачи о клике, вершинном покрытии, независимом множестве, гамильтоновом пути, максимальном разрезе, сумме подмножества, рюкзаке и др. Теорема об иерархии классов сложности по времени.

Литература по разделу:

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982 (Главы 1 – 3).
2. Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ, 2-е издание. М.: Вильямс, 2007 (С. 88 – 97, 1085 – 1150).
3. Arora S. and Barak B. Computational Complexity: A Modern Approach. Cambridge University Press, 2009 (Chapters 1–3).
4. Moore C. and Mertens S. The Nature of Computation. Oxford University Press, 2011 (Chapters 1–6).
5. Sipser M. Introduction to the Theory of Computation. Boston, Mass.: Thomson Course Technology, 2006 (pp. 245–301, 340–343, 348–351).

Раздел 2. Методы решения труднорешаемых задач

Локальный поиск, градиентный спуск, алгоритм Метрополиса, имитация отжига. Рандомизированные и приближенные алгоритмы. Приближенное решение задачи о максимальном разрезе. Алгоритмы типа Монте-Карло и Лас-Вегас. Простой рандомизированный алгоритм для вычисления означивания переменных, максимизирующего число истинных дизъюнктов в 3-КНФ.

Литература по разделу:

1. Клейнберг Дж., Тардос Е. Алгоритмы: разработка и применение. СПб.: Питер, 2016 (Разделы 12.1, 12.2, 12.4, 13.4).
2. Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ, 2-е издание. М.: Вильямс, 2007 (Глава 35).
3. Moore C. and Mertens S. The Nature of Computation. Oxford University Press, 2011 (Section 10.9).
4. Sipser M. Introduction to the Theory of Computation. Boston, Mass.: Thomson Course Technology, 2006 (Section 10.1).

Раздел 3. Задачи и алгоритмы анализа данных

Сегментация изображений (передний/задний план) с помощью минимального разреза. NP-трудность задачи сегментации изображений на три класса. Приближенный алгоритм сегментации изображений на несколько классов.

Кластеризация на основе минимального остовного дерева, максимизирующая минимальное межкластерное расстояние. NP-трудность кластеризации, минимизирующей максимальное внутрикластерное расстояние. Приближенный алгоритм кластеризации, минимизирующий максимальное расстояние до центра кластера.

Locality-sensitive hashing.

Потоковые алгоритмы.

Эффективное перечисление последовательностей. Полиномиальная задержка, кумулятивная задержка, сложность алгоритма относительно размера входа и выхода. Алгоритмы перечисления частых множеств.

Литература по разделу:

1. Клейнберг Дж., Тардос Е. Алгоритмы: разработка и применение. СПб.: Питер, 2016 (Разделы 4.7, 7.10, 11.2, 12.6).
2. Dahlhaus E., Johnson D.S., Papadimitriou C.H., Seymour P.D., and Yannakakis M. The Complexity of Multiterminal Cuts. *SIAM J. Comput*. 23 (1994), 864–894.
3. Johnson D. S., Yannakakis M., and Papadimitriou C.H. On Generating all Maximal Independent Sets. *Information Processing Letters* 27 (1988), 119–123.
4. Leskovec J., Rajaraman A., and Ullman J.D. Mining of Massive Datasets. 2nd edition. Cambridge University Press, 2014 (Sections 3.1–3.4, Chapters 4 and 6).

9 Образовательные технологии

В преподавании дисциплины сочетаются лекции, практические занятия в компьютерном классе и домашние задания, состоящие из теоретических задач и задач на программирование.

10 Оценочные средства для текущего контроля и аттестации студента

10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

Примерные задания для домашней работы:

1. Покажите, что класс P замкнут относительно пересечения.
2. Дано: формула φ. Спрашивается, существует ли по крайней мере два означивания переменных, при которых формула φ истинна? Докажите, что данная задача является NP-полной.
3. Пусть существует алгоритм $A(G, k)$ с полиномиальным временем работы, позволяющий определить, есть ли в графе G клика размера k. Покажите, что в этом случае существует алгоритм $B(G)$ с полиномиальным временем работы, находящий клику максимального размера в графе G (именно саму клику, а не только ее размер).
4. Рассмотрим модель вычислений, в которой программа представлена в виде ориентированного ациклического графа, все узлы которого кроме двух выделенных помечены переменными. Один выделенный узел помечен нулем, другой — единицей; они соответствуют значениям функции, вычисляемой программой. Из каждого узла, помеченного переменной, исходит два ребра: одно помечено нулем, другое — единицей. У выделенных узлов нет исходящих ребер. Один из узлов программы — начальный.
   Программа получает на вход набор булевых значений для переменных, которыми помечены ее узлы. Вычисление начинается в начальном узле. При достижении узла, помеченного переменной, программа проверяет значение этой переменной: если переменная истинна, программа переходит по ребру, помеченному единицей, иначе — по ребру, помеченному нулем. Вычисление завершается при достижении выделенного узла. Программа возвращает значение, которым помечен этот узел.
   Оцените (асимптотически) количество узлов, необходимое для программы, получающей на вход n булевых переменных и возвращающей 1, если хотя бы половина этих переменных истинна, и 0 — в противном случае. Объясните ответ.
5. Приведите полиномиальный алгоритм сведения задачи о раскраске графа:

   Дано: Неориентированный граф и число k.
   Вопрос: Можно ли раскрасить вершины графа в k цветов так, чтобы цвета всех смежных вершин были различны?

   к следующей задаче о целочисленном решении неравенств:

Дано: Система неравенств. Неравенства могут быть как строгие, так и нестрогие. В неравенствах могут присутствовать переменные и константы из области вещественных чисел. Также в неравенствах допустимы следующие операции: сложение, вычитание, умножение, деление, возведение в степень.
Вопрос: Существует ли целочисленное решение данной системы?

Докажите корректность сведения.

6. Реализуйте приближенный алгоритм сегментации изображений на несколько классов.
7. Реализуйте приближенный алгоритм кластеризации, минимизирующий максимальное расстояние до центра кластера.
8. Реализуйте один из алгоритмов поиска частых множеств.

10.2 Примеры заданий промежуточной аттестации

1. Приведите формальное определение детерминированной и недетерминированной машин Тьюринга.
2. Пусть задано конечное множество S и конечный набор его подмножеств $S_1, S_2, ..., S_k$. Для каждого множества S_i заданы два числа low_i и $high_i$. Требуется выяснить, существует ли подмножество $T \subseteq S$, такое что для каждого i выполняется $low_i \leq |T \cap S_i| \leq high_i$. Докажите, что эта задача является NP-трудной, сведя к ней задачу о выполнимости 3-КНФ (3SAT).
3. Верно ли, что SAT $\in$ coNP?
4. Задача по определению того, содержит ли граф простой (без повторяющихся вершин) путь с числом ребер не менее m является NP-полной. Почему из этого следует, что существование алгоритма, вычисляющего самый длинный простой цикл в графе за полиномиальное время от длины входа, маловероятно?
5. Обозначим за I некоторое множество предметов, имеющих целочисленные «вес» $w(i)$ и «ценность» $v(i)$. Пусть есть оракул, моментально отвечающий на вопрос, можно ли выбрать из множества I предметы, общий вес которых не выше, а общая ценность не ниже заданных. Объясните, как при помощи такого оракула можно за полиномиальное время определить максимальную общую ценность набора предметов из I, общий вес которых не превышает заданный максимальный вес M.
6. Задача о доминирующем множестве:
   DS = $\{\langle G, k\rangle \mid G$ — неориентированный граф (V, E); существует подмножество $C \subseteq V$, такое что $|C| = k$ и каждая вершина в $V \setminus C$ связана ребром из E с некоторой вершиной из $C\}$.
   Докажите, что DS $\in$ NP, описав недетерминированную машину Тьюринга с полиномиальным временем работы, решающую DS. Докажите, что DS — NP-трудная задача, сведя к ней (за полиномиальное время) задачу о вершинном покрытии. (По графу $G = (V, E)$ постройте граф $G' = (V', E')$, в котором $V' \subseteq V \cup E$.)
7. Докажите, что существует язык, разрешимый за время $O(n^3)$ на одноленточной машине Тьюринга, но не разрешимый за время $O(n)$ на многоленточной машине Тьюринга.
8. Рассмотрим задачу о поиске кратчайшего маршрута в неориентированном графе: требуется обойти граф, побывав во всех его вершинах и вернувшись в исходную вершину, так, чтобы сумма весов на ребрах пройденного пути была минимально возможной. Приближенное решение, отличающееся от оптимального не более чем вдвое, можно получить, если построить минимальное остовное дерево графа и обойти его. Такое решение можно найти за время, полиномиальное от размера входных данных. Предложите модификацию описанного алгоритма, работающую за полиномиальное время и вычисляющую решение, которое никогда не хуже решения, полученного исходным алгоритмом, но зачастую лучше.

11 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на семинарских и практических занятиях: правильность решения теоретических задач, корректность и эффективность компьютерных программ, написанных студентом. Оценки за работу на семинарских и практических занятиях преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем - $O_{аудиторная}$.

Накопленная оценка по дисциплине рассчитывается по формуле:

$O_{накопленная} = 0{,}4 \cdot O_{дз1} + 0{,}4 \cdot O_{дз2} + 0{,}2 \cdot O_{аудиторная}$.

В диплом выставляется результирующая оценка по учебной дисциплине.

$O_{результ} = 0{,}6 \cdot O_{накопленная} + 0{,}4 \cdot O_{экз}$

Способ округления результирующей оценки по учебной дисциплине: на усмотрение преподавателя до ближайшего сверху или снизу целочисленного значения.

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Базовый учебник

Кормен Т.Х., Лейзерсон Ч.И., Ривест Р.Л., Штайн К. Алгоритмы: построение и анализ. — 2-е издание: Пер. с англ. — М.: Вильямс, 2007.

12.2 Основная литература

Leskovec J., Rajaraman A., and Ullman J.D. Mining of Massive Datasets. 2nd edition. Cambridge University Press, 2014.

Электронная версия: http://infolab.stanford.edu/~ullman/mmds/bookL.pdf.

12.3 Дополнительная литература

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
2. Клейнберг Дж., Тардос Е. Алгоритмы: разработка и применение. СПб.: Питер, 2016.
3. Arora S. and Barak B. Computational Complexity: A Modern Approach. Cambridge University Press, 2009.
4. Dahlhaus E., Johnson D.S., Papadimitriou C.H., Seymour P.D., and Yannakakis M. The Complexity of Multiterminal Cuts. *SIAM J. Comput*. 23 (1994), 864–894.
5. Johnson D. S., Yannakakis M., and Papadimitriou C.H. On Generating all Maximal Independent Sets. *Information Processing Letters* 27 (1988), 119–123.
6. Moore C. and Mertens S. The Nature of Computation. Oxford University Press, 2011.
7. Sipser M. Introduction to the Theory of Computation. Boston, Mass.: Thomson Course Technology, 2006.

12.4 Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

- интерпретатор языка Python 3.x;
- компилятор языка C++11 (g++ 4.8).