

**Программа утверждена на заседании кафедры вычислительной математики
Протокол № 4 от 27 ноября 2014 г.**

Рабочая программа дисциплины (модуля)

1. Код и наименование дисциплины (модуля): Методы анализа молекулярных графов в задаче «структура-свойство».
2. Уровень высшего образования – подготовка научно-педагогических кадров в аспирантуре.
3. Направление подготовки «09.06.01 Информатика и вычислительная техника». Направленность программы: «Теоретические основы информатики» (научная специальность 05.13.17).
4. Место дисциплины (модуля) в структуре ООП: вариативная часть ООП, элективный курс по выбору кафедры, обязателен для освоения не позднее второго года обучения.
5. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции (код компетенции)	Планируемые результаты обучения по дисциплине (модулю)
УК-1	У1 (УК-1) УМЕТЬ: анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов
УК-4	З1 (УК-4) ЗНАТЬ: методы и технологии научной коммуникации на государственном и иностранном языках
ОПК-1	В2 (ОПК-1) ВЛАДЕТЬ: навыками решения теоретических и практических задач при помощи современных программных средств
ПК-2	З1 (ПК-2) ЗНАТЬ: современные методы машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска

	<p>У1 (ПК-2) УМЕТЬ: применять современные методы машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска в научно-исследовательской работе.</p> <p>В1 (ПК-2) ВЛАДЕТЬ: современными навыками решения теоретических и практических задач в области теоретической информатики.</p>
--	---

6. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических или астрономических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся:

Объем дисциплины (модуля) составляет 2 зачетные единицы, всего 72 часа, из которых 44 часа составляет контактная работа аспиранта с преподавателем (34 часа занятия лекционного типа, 2 часа занятия семинарского типа (семинары, научно-практические занятия, лабораторные работы и т.п. 8 часов мероприятия промежуточной аттестации), 28 часов составляет самостоятельная работа аспиранта.

7. Входные требования для освоения дисциплины (модуля), предварительные условия.

Для того чтобы формирование данной компетенции было возможно, обучающийся, приступивший к освоению программы аспирантуры, должен

Владеть: навыками решения задач из следующих разделов современной математики: математический анализ, линейная алгебра и аналитическая геометрия, математическая логика, дискретная математика

Знать: основные направления, проблемы, теории и методы современной математики, в том числе основы математического анализа, линейной алгебры, дискретной математики, математической логики, теории вероятностей и других смежных областей математики.

Уметь: решать стандартные задачи математического анализа, линейной алгебры, дискретной математики и математической логики и применять идеи, использованные в их решениях, для решения аналогичных задач.

Иметь навык: программирования на языке высокого уровня типа Си или C++.

.8. Формат обучения -спецкурс по выбору кафедры

9. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических или астрономических часов и виды учебных занятий

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы из них						Самостоятельная работа обучающегося, часы из них		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости коллоквиумы, практические занятия и др)*	Всего	Выполнение домашних заданий	Подготовка рефератов и т.п..	Всего
Тема 1: Представление молекулярных графов в компьютере, графические редакторы	10	6					6	4		4
Тема 2: Формирование МД-матриц на основе молекулярных дескрипторов, зависимость и отсеивание признаков. Структурные символьные спектры М-графов	20	10					10	8	2	10
Тема 3: Линейные нотации представления молекулярных графов в ЭВМ: WLN нотация	14	6	2				8	6		6

Виссвесера, SMILE-код М-графа, волновой алгоритм формирования матрицы топологических расстояний										
Тема 4: Геометрический подход к анализу данных. Близость точек в евклидовом. Метрики М-графов. Алгоритмы кластерного анализа	10	6					6	4		4
Тема 5: Эволюционные алгоритмы- метод группового учета аргумен-тов, генетические алгоритмы.	10	6					6	4		4
Промежуточная аттестация: экзамен	8						8			
Итого	72	34	2	0	0	0	44	26	2	28

10. Перечень учебно-методического обеспечения для самостоятельной работы аспирантов по дисциплине (модулю).
Список литературы, см. 12.

11. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю).

- У1 (УК-1), 31(УК-4), В2 (ОПК-1), 31 (ПК-2), У1 (ПК-2), В1 (ПК-2).
- Описание шкал оценивания: *экзамен с оценкой по пятибалльной шкале.*
- Критерии и процедуры оценивания результатов обучения по дисциплине (модулю), характеризующих этапы формирования компетенций.

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модулю)	ШКАЛА оценивания					ПРОЦЕДУРЫ ОЦЕНИВАНИЯ*
	1	2	3	4	5	
У1 (УК-1) УМЕТЬ: анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов	Отсутствие умений	Частично освоенное умение анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов	В целом успешно, но не систематически осуществляемые анализ альтернативных вариантов решения исследовательских и практических задач и оценка потенциальных выигрышей/проигрышей реализации этих вариантов	В целом успешно, но содержащие отдельные пробелы анализ альтернативных вариантов решения исследовательских задач и оценка потенциальных выигрышей/проигрышей реализации этих вариантов	Сформированное умение анализировать альтернативные варианты решения исследовательских и практических задач и оценивать потенциальные выигрыши/проигрыши реализации этих вариантов	оценка реферативного отчета на экзамене в форме индивидуального собеседования
З1 (УК-4) ЗНАТЬ: методы и технологии научной коммуникации на государственном и иностранном языках	Отсутствие знаний	Фрагментарные знания методов и технологий научной коммуникации на государственном и иностранном языках	Неполные знания методов и технологий научной коммуникации на государственном и иностранном языках	Сформированные, но содержащие отдельные пробелы знания методов и технологий научной коммуникации на государственном и иностранном языках	Сформированные и систематические знания методов и технологий научной коммуникации на государственном и иностранном языках	оценка реферативного отчета на экзамене в форме индивидуального собеседования
В2 (ОПК-1) ВЛАДЕТЬ: навыками решения теоретических и практических задач при помощи современных	Отсутствие навыков	Фрагментарное применение навыков построения и анализа математических моделей, решения	В целом успешное, но не систематическое применение навыков построения и анализа математических	В целом успешное, но содержащее отдельные пробелы применение навыков построения и анализа математических моделей, решения	Успешное и систематическое применение навыков построения и анализа математических моделей, решения	оценка реферативного отчета на экзамене в форме индивидуального собеседования

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модулю)	ШКАЛА оценивания					ПРОЦЕДУРЫ ОЦЕНИВАНИЯ*
	1	2	3	4	5	
программных средств		задач при помощи современных программных средств	моделей, решения задач при помощи современных программных средств	задач при помощи современных программных средств	задач при помощи современных программных средств	
31 (ПК-2) ЗНАТЬ: современные методы машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска	Не имеет базовых знаний	Допускает существенные ошибки	Демонстрирует частичные знания	Демонстрирует знания с небольшими пробелами	Раскрывает полное содержание математических основ и концептуальных моделей теоретической информатики, включая современные методы машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска	экзамен в форме индивидуального собеседования
У1 (ПК-2) УМЕТЬ: применять современные методы машинного обучения, анализа данных, управления данными, представления знаний и	Не умеет и не готов формулировать	Имея базовые представления о предмете, не готов формулировать задачи и выбирать методы их решения.	Не учитывает специфики современного состояния предмета	Не вполне готов выбирать методы анализа и интерпретировать	Готов и умеет корректно ставить задачи в терминах машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска, выбирать методы их анализа и решения, представлять и интерпретировать	письменное решение задач

РЕЗУЛЬТАТ ОБУЧЕНИЯ по дисциплине (модулю)	ШКАЛА оценивания					ПРОЦЕДУРЫ ОЦЕНИВАНИЯ*
	1	2	3	4	5	
информационного поиска в научно- исследовательск ой работе.					полученные результаты	
В1 (ПК-2) ВЛАДЕТЬ: современными навыками решения теоретических и практических задач в области теоретической информатики.	Не владеет методами и навыками.	Владеет отдельными приемами	Владеет приемами и навыками решения основных стандартных задач	Владеет методами и навыками, но не готов оценивать востребованность конкретных задач в современной науке.	Полностью владеет методами машинного обучения, анализа данных, управления данными, представления знаний и информационного поиска, алгоритмами для решения задач; навыками создания и исследования новых актуальных теорий и направлений, востребованных в современной науке	экзамен в форме индивидуальног о собеседования

- Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих этапы формирования компетенций.

Контрольные вопросы.

- Информатика как наука, изучающая информацию и ее свойства в естественных, искусственных и гибридных системах.
- Классификация информационных продуктов и услуг. Жизненный цикл информационного продукта.
- Методы управления производством и распределением информационных продуктов.
- Задача машинного обучения. Объекты и признаки. Основные понятия: метод обучения, функционал качества, обобщающая способность, скользящий контроль.
- Представление молекул. Молекулярные графы – топологическое и геометрическое представления. Методы расчета.
- Химическое пространство в задаче «структура-свойство» (СС-задача)
- Дескрипторы М-графов. Топологические индексы и их построение.
- Топологических индекс Винера и его использование. Пространственная версия индекса.
- Топологических индекс Рандича. Индексы Кира-Холла и их использование.
- Поиск в химическом пространстве. Гипотеза компактности. Кластерный анализ.
- Задача структура свойство. Общая постановка. Задача построения описания. Задача поиска корреляций.
- Задачи поиска корреляций «структура-свойство» - SAR/QSAR/QSPR
- Химические базы данных. Состав и принципы проведение поиска. Использование для формирования обучающих выборок.
- Навигация в химическом пространстве. Кластеры и их идентификация
- Генетический алгоритм. Его конструирование для поиска значимых дескрипторов в СС-задаче
- Генетический алгоритм. Его конструирование для поиска оптимальной метрики - выбор дескрипторов в метрику
- Двухфазная схема построения классификаторов в СС-задаче.
- Метод группового учета аргументов для поиска значимых дескрипторов в СС-задаче
- Метод группового учета аргументов для поиска оптимальной метрики - выбор дескрипторов в метрику
- Маркеры атомов. Структурные символьные спектры.
- Пространственные дескрипторы М-графов. Нечеткое определение интервалов расстояний и конформации молекул.
- Задача машинного обучения. Объекты и признаки. Основные понятия: метод обучения, функционал качества, обобщающая способность, скользящий контроль.
- Задача классификации. Анализ формальных понятий.
- Задача классификации. Метод опорных векторов.
- Задача классификации. Алгоритм к ближайших соседей.

- Задача классификации. Байесовские классификаторы.
 - Задача классификации. Алгоритм AdaBoost.
 - Задача классификации. Факторный анализ.
 - Задача кластеризации. Алгоритм k-средних.
 - Задача кластеризации. Графовые алгоритмы.
 - Задача классификации. Алгоритм C4.5.
 - Задача кластеризации. Иерархическая кластеризация.
 - Процесс дизайна молекул с заданными свойствами. Основные этапы - сценарии использования и задачи, решаемые наЭВМ.
- Методические материалы, определяющие процедуры оценивания.
Итоговая экзаменационная оценка формируется как минимум оценки реферата и оценки ответа по экзаменационному билету.
Экзаменационный билет состоит из двух контрольных вопросов и практического задания.

Образцы билетов.

- Билет 1. (А) Задачи поиска корреляций «структура-свойство» - SAR/QSAR/QSPR. (В) Маркеры атомов. Структурные символьные спектры. (С) Построить матрицу расстояний и вычислить индекс Винера для заданного М-графа.
- Билет 2. (А) Химические базы данных. Состав и принципы проведение поиска. Использование для формирования обучающих выборок. (в) Генетический алгоритм. Его конструирование для поиска значимых дескрипторов в СС-задаче. (с) Вычислить индекс Рандича для заданного М-графа.

Примеры тем рефератов:

Топологические индексы молекулярных графов
Алгоритмы кластерного анализа.
Геометрический подход к классификации данных
Синтаксические методы распознавания образов
Линейная и непараметрическая регрессия

12. Ресурсное обеспечение:

- **Перечень основной и дополнительной учебной литературы:**

1. Прогнозирование свойств химических соединений. Унифицированный Репозиторий моделей «структура-свойство» - Москва, 2012, Изд-во МАКС Пресс, ISSN 978-5-317-04335-3,
2. Система прогнозирования свойств химических соединений. Алгоритмы и модели. –Москва, 2008, Издательство МАКС Пресс, ISBN 978-5-317-02704-9,
3. Химические приложения топологии и теории графов, под ред. Р. Кинга. — М.: Мир, 1987. — 560 с
4. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003 -649 с.
5. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger -Weinheim: Wiley-VCH, 2003 - 1870 с
6. Хельтье Х.-Д. и др. Молекулярное моделирование: теория и практика - пер. с англ. - М: Бином. Лаборатория знаний, 2009.-318 с.
7. В.В. Корнеев, А.Ф. Гареев, С.В. Васютин, В.В. Райх. Базы данных. Интеллектуальная обработка информации. Нолидж, 2000.
8. С. М. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006
9. М. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of Machine Learning, MIT Press, 2012.

Дополнительная литература:

1. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003 - 649 с.
2. Leach A.R., Gillett V.J. An introduction to Chemoinformatics. - Dordrecht: Kluwer Academic Publishers, 2003 - 259 с.
3. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger - Weinheim: Wiley-VCH, 2003 - 1870 с.
4. R. Todeschini, V. Consonni. Molecular Descriptors for chemoinformatics. Т. 1. Изд. 2, доп. - Weinheim: Wiley-VCH, 2009 - 967 с.

- **Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

Virtual Computational Chemistry Laboratory - <http://www.vcclab.org/>

База данных ChEMBL - <https://www.ebi.ac.uk/chembl>

База данных ChemSpider - <http://www.chemspider.com>

База данных PubChem - <http://pubchem.ncbi.nlm.nih.gov/>

База данных Reaxys - <http://www.reaxys.com>

База данных ZINC - <http://zinc.docking.org>

Программа Fragmenter - <http://infochim.u-strasbg.fr/spip.php?rubrique49>

Программа ISIDA-QSPR - <http://www.vpsolovev.ru/programs/>

Программа Weka (Waikato Environment for Knowledge Analysis) - <http://www.cs.waikato.ac.nz/ml/weka/>

- Перечень используемых информационных технологий, используемых при осуществлении образовательного процесса, включая программное обеспечение, информационные справочные системы (при необходимости): по ситуации.

ISIS/Host, ISIS/Base (www.mdli.com)

ChemFinder, ChemOffice (www.cambridgesoft.com)

JChem (www.chemaxon.com)

THOR (www.daylight.com)

MOE (www.chemcomp.com)

ICM Pro (под mySQL) (www.molsoft.com)

CheD (Сергей Трепалин)

UNITY (www.tripos.com)

OrChem (orchem.sourceforge.net)

Bingo (ggasoftware.com/opensource/bingo)

Pgchem::tigris (pgfoundry.org/projects/pgchem)

- Описание материально-технической базы.
аудиторный фонд механико-математического факультета, персональные компьютеры слушателей.

13. Язык преподавания - русский

14. Преподаватель - профессор Михаил Иванович Кумсков

Заведующий кафедрой вычислительной математики
профессор, д.ф.-м.н.

/Кобельков Г.М./