

**Программа утверждена на заседании кафедры вычислительной математики
Протокол № 1 от 7 сентября 2015 г.**

Рабочая программа дисциплины (модуля)

Код и наименование дисциплины (модуля): СПЕЦКУРС (полугодовой). Методы анализа молекулярных графов в задаче «структурасвойство».

2. Уровень высшего образования – специалитет.

3. Направление подготовки 01.05.01 Фундаментальная математика и механика. Специализация: Фундаментальная математика.

4. Место дисциплины (модуля) в структуре ООП: вариативная часть ООП. Является специальной дисциплиной (спецкурсом) для студентов 3-6 годов обучения, специализирующихся в данной научной области или смежной научной области, спецкурсом по выбору студента.

Освоение дисциплины необходимо для последующего изучения дисциплин образовательной программы: курсовая работа, научно-исследовательская практика, преддипломная практика, выпускная квалификационная работа.

5. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

Формируемые компетенции	Планируемые результаты обучения по дисциплине (модулю)
СПК-1 современными навыками решения теоретических и практических задач в области химической информатики	Знать современные методы машинного обучения, методы и алгоритмы формирования дескрипторов молекулярных графов; уметь применять современные методы машинного обучения в научно-исследовательской работе, формировать описания молекулярных графов по его разным представлениям

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 3 зачетных единицы, всего 108 часа, из которых 44 (46*) часа составляет контактная работа студента с преподавателем (34 (36*) часа занятия лекционного типа, 12 часов мероприятия текущего контроля успеваемости и промежуточной аттестации), 64 (62*) часа составляет самостоятельная работа студента.

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 математики

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

Очная форма обучения, лекционные занятия.

9. Содержание дисциплины (модуля), структурированное по темам (Перечень тем см. Приложения).

Наименование и краткое содержание разделов и тем дисциплины (модуля), форма промежуточной	Всего (часы)	В том числе
---	--------------	-------------

аттестации		Контактная работа (работа во взаимодействии с преподавателем), часы из них					Самостоятельная работа обучающегося, часы из них			
		Занятия лекцио нного типа	Занят ия семинарског о типа	Групп овые консу льтац ии	Индиви дуальн ые консуль тации	Учебные занятия, направленные на проведение текущего контроля успеваемости коллоквиумы, практические контрольные занятия и др)*	Всего	Выполне ние домашних заданий	Подгот ов-ка рефера тов и т.п..	Всего
Тема 1	6	2					2	4		4
Тема 2	6	2					2	4		4
Тема 3	6	2					2	4		4
Тема 4	6	2					2	4		4
Тема 5	6	2					2	4		4
Тема 6	6	2					2	4		4
Тема 7	6	2					2	4		4
Тема 8	6	2					2	4		4
Текущий контроль успеваемости	6					2	2	4		4
Тема 9	6	2					2	4		4

Тема 10	6	2					2	4		4
Тема 11	6	2					2	4		4
Тема 12	6	2					2	4		4
Тема 13	6	2					2	4		4
Тема 14	6	2					2	4		4
Тема 15	6	2					2	4		4
Тема 16	4						0	4		4
Тема 17*	2*							2*		2*
Промежуточная аттестация <u>экзамен</u>	8 (6*)					2	2	6(4*)		6 (4*)
Итого	108	30				4	34	74		74

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).

- Перечень компетенций: ПСК-1.
- 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)	КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ по дисциплине (модулю) и ШКАЛА оценивания					ПРОЦЕДУРЫ ОЦЕНИВАНИЯ
	1	2	3	4	5	
Знание методов машинного обучения, алгоритмов формирования дескрипторов молекулярных графов	Отсутствие знаний	Фрагментарные знания методов машинного обучения, алгоритмов формирования дескрипторов молекулярных графов	Общие, но не структурированные знания методов машинного обучения алгоритмов формирования дескрипторов молекулярных графов	Сформированные, но содержащие отдельные пробелы знания методов машинного обучения, алгоритмов формирования дескрипторов молекулярных графов	Сформированные систематические знания методов машинного обучения, алгоритмов формирования дескрипторов молекулярных графов	Экзамен (зачет) в форме индивидуального собеседования
уметь применять методы машинного обучения в научно-исследовательской работе, формировать описания молекулярных графов	Отсутствие умений	Частично освоенное умение применять методы машинного обучения в научно-исследовательской работе, формировать описания М- графов	В целом успешное, но не систематически осуществляемое умение применять методы машинного обучения в научно-исследовательской работе, формировать описания М- графов	В целом успешное, но содержащие отдельные пробелы умение применять методы машинного обучения в научно-исследовательской работе, формировать описания М- графов	Сформированное умение применять методы машинного обучения в научно-исследовательской работе, формировать описания М- графов	Письменное решение задач

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

12. Ресурсное обеспечение:

Перечень основной учебной литературы: см. Приложение

Перечень дополнительной учебной литературы: см. Приложения

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»: см. Приложения.

Описание материально-технической базы: аудитории для проведения лекционных занятий.

13. Язык преподавания: русский (при необходимости – английский).

ПРИЛОЖЕНИЕ

Спецкурс программы специалитета, полугодовой: Методы анализа молекулярных графов в задаче «структура-свойство».

- Преподаватель – профессор М.И. Кумсков
- Аннотация курса: специальный курс для студентов рассматривает основные понятия и методы программной симуляции систем, основное внимание уделяется симуляции микропроцессорных средств и систем
- Тематическое содержание курса:

Тема 1	Ключевые понятия химической информатики.
Тема 2	Задача «структура-свойства», задача QSAR и задача QSPR.
Тема 3	Молекулярный граф и его представления в задаче «структура-свойство»
Тема 4	Топологические индексы молекулярных графов.
Тема 5	Фармакофоры. Фрагментный код суперпозиции подструктур
Тема 6	Элементы представления молекулярного графа. Локальные физ.-хим. свойства молекулы. Символьные маркеры вершин.
Тема 7	Структурный символьный спектр. Матрица «молекула-дескриптор». Теорема о расщеплении столбцов МД-матрицы.
Тема 8	Поиск дескрипторов, адаптированных к заданному свойству – проверка гипотезы химика-эксперта.
Тема 9	Многомерный регрессионный анализ МД-матрицы. Скользящий контроль.
Тема 10	Метод группового учета аргументов (МГУА) на основе линейной регрессии. Семейства моделей.
Тема 11	Генетический алгоритм поиска значимых дескрипторов.
Тема 12	Понятие кластера (таксона). Деревья решений на основе кластеров. Идентификация выбросов.

Тема 13	Алгоритмы кластерного анализа в задаче «структура-свойство».
Тема 14	Двухфазная схема решения задачи «структура-свойство».
Тема 15	Поиск метрики на основе отбора дескрипторов. Использование МГУА.
Тема 16	Представление моделей «структура-свойство» (СС-модели). Унифицированный репозиторий.
Тема 17*	Параллельные вычисления при формировании СС-моделей.

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

Контрольные вопросы.

- Классификация информационных продуктов и услуг химической информатики. Жизненный цикл СС модели.
- 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 расчёта.
- 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. Выбросы.
- 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 k ближайших соседей.
 - Задача классификации. Байесовские классификаторы.
 - Задача классификации. Алгоритм AdaBoost.
 - Задача классификации. Факторный анализ.
 - Задача кластеризации. Алгоритм k-средних.
 - Задача кластеризации. Графовые алгоритмы.
 - Задача кластеризации. Иерархическая кластеризация.
 - Процесс дизайна молекул с заданными свойствами. Основные этапы - сценарии использования и задачи, решаемые на ЭВМ.
- 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 проведения 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.
2. Система прогнозирования свойств химических соединений. Алгоритмы и модели. –Москва, 2008, Издательство МАКС Пресс.
3. Химические приложения топологии и теории графов, под ред. Р. Кинга. — М.: Мир, 1987.
4. Хельтье Х.-Д. и др. Молекулярное моделирование: теория и практика - пер. с англ. - М: Бином. Лаборатория знаний, 2009.
5. В.В. Корнеев, А.Ф. Гареев, С.В. Васютин, В.В. Райх. Базы данных. Интеллектуальная обработка информации. Нолидж, 2000.

Дополнительная литература:

1. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003.
2. Leach A.R., Gillett V.J. An introduction to Chemoinformatics. - Dordrecht: Kluwer Academic Publishers, 2003.
3. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger - Weinheim: Wiley-VCH, 2003.
4. R. Todeschini, V. Consonni. Molecular Descriptors for chemoinformatics. Т. 1. Изд. 2, доп. - Weinheim: Wiley-VCH, 2009.
5. С. М. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006.
6. М. Mohri, А. Rostamizadeh, А. Talwalkar. Foundations of Machine Learning, MIT Press, 2012.
7. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003.
8. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger -Weinheim: Wiley-VCH, 2003.

- 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)

Приложение утверждено на заседании кафедры вычислительной математики

Протокол № 1 от 7 сентября 2015 г.