

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт «Умные материалы и технологии»

УТВЕРЖДЕНО:
Директор Института «Умные
материалы и технологии»
И.А. Курзина

Оценочные материалы по дисциплине

Хемоинформатика

по направлению подготовки

27.03.05 Инноватика

Направленность (профиль) подготовки:
**Tomsk International Science Program, с профессиональным модулем Молекулярная
инженерия / Molecular Engineering**

Форма обучения
Очная

Квалификация
Инженер

Год приема
2024

СОГЛАСОВАНО:
Руководитель ОП
И.А. Курзина

Председатель УМК
Г.А. Воронова

1 Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-3. Способен применять современные информационные компьютерные технологии, обрабатывать и использовать новую информацию в предметной области.

ПК-2. Способен решать профессиональные задачи на основе знаний в сфере биотехнологии и молекулярной инженерии на основе знаний естественных, математических и технических наук, а также математических методов и моделей.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

РООПК-3.1. Знает принципы работы современных информационных компьютерных технологий, программ и сред.

РООПК-3.2. Умеет осваивать и применять современные информационные компьютерные технологии, программы и среды для обработки и получения информации в профессиональной сфере.

РОПК-2.1. Знает существующие подходы к решению профессиональных задач, в том числе на основе математических методов и моделей.

2 Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- практическое задание;
- тест.

2.1 Практическое задание (РООПК-3.1, РООПК-3.2, РОПК-2.1):

С использованием базы ChEMBL отберите данные по активности соединений против циклооксигеназы-2 (COX-2). Отберите наиболее надежные данные. Проведите химическую и математическую чистку данных. Отберите 10% набора в качестве внешнего валидирующего набора случайным образом. С помощью программы ISIDA-MLR и различных типов фрагментных дескрипторов постройте несколько QSAR моделей (не менее 10). Для валидации модели используйте 5-кратную кросс-валидацию и внешний валидирующий набор. Отберите лучшую модель. Используйте эту модель для скрининга данных из базы ZINC. В качестве набора возьмите только «лекарствоподобные соединения» базы. Из числа соединений, показавших наибольшую активность ($IC_{50} < 10^{-7}$ моль/л), отберите 5%, имеющих наименьшую сумму расстояний от соединений, использованных для тренировки модели, в пространстве фрагментных дескрипторов.

Метод рекомендации по выполнению:

1. Начать подготовку отчета по практической работе заранее, не менее чем за 3 дня до занятия
2. Подобрать по теме литературу и воспользоваться материалами, предоставленными на занятии преподавателем
3. Изучить вопросы практической работы в разных источниках
4. По необходимости проконсультироваться у преподавателя
5. Подготовиться к сдаче отчета по практической работе

Критерии оценивания:

Оценка «зачтено» выставляется в трех случаях:

5 баллов: если студент активно работает в течение всего практического занятия, дает полные ответы на вопросы преподавателя в соответствии с планом практического занятия и показывает при этом глубокое овладение лекционным материалом.

4 балла: если студент активно работает в течение практического занятия, вопросы освещены полно, изложения материала логическое, обоснованное фактами, со ссылками на соответствующие нормативные документы и литературные источники, освещение вопросов завершено выводами, студент обнаружил умение анализировать факты и события, а также выполнять учебные задания, но в ответах допущены неточности, некоторые незначительные ошибки.

3 балла: если студент в целом овладел сути вопросов по данной теме, обнаруживает знание лекционного материала и учебной литературы, пытается анализировать факты и события, делать выводы и решать задачи, но на занятии ведет себя пассивно, отвечает только по вызову преподавателя, дает неполные ответы на вопросы, допускает грубые ошибки при освещении теоретического материала.

Оценка «не зачтено» (2 и менее баллов) выставляется в случае, когда студент обнаружил несостоятельность осветить вопрос, вопросы освещены неправильно, бессистемно, с грубыми ошибками, отсутствуют понимания основной сути вопросов, выводы, обобщения, обнаружено неумение решать учебные задачи.

2.2 Тестовые задания для проверки текущей успеваемости (РООПК-3.1, РООПК-3.2):

- Какие из указанных SMILES соответствует молекуле аспирина?
 - CC(=O)Oc1ccccc1C(O)=O
 - c1c(C(O)=O)cc(OC(O)C)ccc1
 - OC(=O)c(cccc1)c1OC(=O)C
 - c1(C(=O)O)ccccc1OC(=O)C
- Какие InChI для приведенной молекулы гуанина соответствуют молекуле и являются стандартными?
 - InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)/f/h8,10H,6H2
 - InChI=1S/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)
 - InChI=1S/C6H6N5O2/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)
 - InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)/f/h7,9H,6H2
- Какой SMARTS запроса будет определять выделенную подструктуру в приведённой молекуле? Атомы водорода не принимать во внимание.
 - N~*~*~N
 - NcccN
 - [#7]ccc[#7]
 - [NH2]aaa[NH2]
- Какая из приведенных SMILES удовлетворяет приведенной структуре Маркуша?
 - OCCc1c(C)cccc1
 - OCCCCCCCCc1cc(C(=O)O)ccc1
 - OCCc1ccc(C(C)=O)cc1
 - OCCc1ccc(COC=O)cc1
- Какое из приведенных отнесений отмеченных фармакофорных центров 1,2 и 3 является наиболее полным и корректным (один ответ)? Обозначения: N - negative charge, P - positive charge, H - hydrophobe, Ar - aromatic ring, A - H-acceptor, D - H-donor.
 - 1: N; 2: D; 3: P;
 - 1: N; 2: A, D; 3: P, D;
 - 1: N, A; 2: A, D; 3: P, D, Ar;

- d. 1: N, A; 2: A, D, N; 3: P, A, D, Ar;
6. Каким из приведенных молекул (1,2,3, 4) удовлетворяет приведенный трехточечный топологический фармакофор P? Обозначения: P - positive charge, A - H-ассептор, D - H-donor. Расстояния являются топологическими.
- 1
 - 2
 - 3
 - 4
7. Каким из приведенных молекул (1,2,3,4) соответствует данная структура Маркуша?
- 1
 - 2
 - 3
 - 4
8. Какая из указанных структур (1,2,3,4), содержащихся в базе, будет выдаваться в результате поиска по субструктуре M? Атомы водорода не принимаются во внимание.
- 1
 - 2
 - 3
 - 4
9. Какая из указанных структур (1,2,3,4), содержащихся в базе, будет выдаваться в результате поиска по суперструктуре M?
- 1
 - 2
 - 3
 - 4
10. Две структуры задаются указанными ниже битовыми строками. Какой будет индекс схожести Танимото между данными структурами?
- Mol 1 1 0 0 0 1 1 1 0 1 1
- Mol 2 0 1 0 0 1 0 1 1 0 1
- 6/11
 - 5/8
 - 3/10
 - 3/8
11. Какое из приведенных выражений содержит формулу для вычисления индекса схожести Тверского? a-число активных бит в одной молекуле, b - число включенных бит в другой молекуле, c - число бит, которые являются активными в обеих молекулах.
- $c/(a+b-c)$
 - $2c/(a+b)$
 - $(a+b-2c)/(a+b-c)$
 - $c/(c+\alpha(a-c)+\beta(b-c))$
12. В какой из приведенных баз данных можно найти информацию, характеризующую прочность связывания данного химического соединения с различными белками?
- CAS
 - PubChem
 - ChEMBL
 - ZINC
13. Какую информацию о соединении можно найти в базе ChemSpider?
- Химическая структура
 - Информация об испытании данного соединения на bioassay
 - Кристаллическая структура молекулы

- d. Индекс LASSO, характеризующий насколько данная молекула подходит для связывания с активными центрами различных ферментов
14. Какие этапы входят в процесс осуществления поиска по структуре?
- Стандартизация соединения
 - Генерация хэш-кода
 - Поиск молекулы с помощью скринов
 - Поиск индекса схожести данного соединения с другими соединениями базы

Критерии оценивания тестовых заданий

Оценка «5» (отлично) – выставляется обучающемуся, допустившему до 10 % ошибок в тестовом задании. Оценка «4» (хорошо) – выставляется обучающемуся, допустившему до 24 % ошибок в тестовом задании. Оценка «3» (удовлетворительно) – выставляется обучающемуся, допустившему до 39 % ошибок в тестовом задании. Оценка «2» (неудовлетворительно) – обучающийся допустил более 40 % ошибок в тестовом задании.

3 Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Зачет по курсу «Хемоинформатика» проводится в форме устного опроса студентов.

3.1 Примеры вопросов для зачета:

- Основные способы представления объектов в хемоинформатике;
- Линейные представления;
- Матричные представления;
- Битовые представления;
- Стандартные файлы в хемоинформатике;
- Хеометрика и хемоинформатика;
- Биоинформатика и хемоинформатика;
- Дизайн библиотек данных;
- Разбросанные и сфокусированные библиотеки;
- Генерация структур;
- RECAP;
- Кластеризация молекул;
- Иерархические и неиерархические подходы;
- Отбор молекул без кластеризации;
- Основные способы валидации моделей;
- Классификационные модели;
- Регрессионные модели;
- Отбор дескрипторов;
- Навигация в химическом пространстве как способ моделирования;
- Методы машинного обучения.

Оценка «зачтено» выставляется в случае, если студент выполнил правильно ответил на два вопроса.

Оценка «незачтено» выставляется в случае, если студент не ответил или ответил на вопросы с существенными ошибками.

4 Оценочные материалы для проверки остаточных знаний (сформированности компетенций) (РОПК-3.1, РОПК-2.1):

- Какова основная цель хемоинформатики?
 - Интеграция и анализ химической информации с помощью вычислительных методов
 - Изучение физических свойств веществ

- В. Создание новых химических соединений
 - Г. Исследование взаимодействий между биомологическими системами
2. Какую роль играет химическая информация в современных научных исследованиях?
- + А. Обеспечивает базу данных для разработки новых веществ
 - Б. Служит исключительно для академических публикаций
 - В. Ограничивает исследования до традиционных методов
 - Г. Не имеет практического применения в промышленности
3. Какой из следующих форматов используется для представления химических структур в компьютерах?
- + А. SMILES
 - Б. IUPAC название
 - В. Молекулярная формула
 - Г. Химическая реакция
4. Какой тип графа обычно используется для представления молекул?
- + А. Связный граф
 - Б. Дерево
 - В. Циклический граф
 - Г. Орфографический граф
5. Какой тип информации хранится в химических базах данных?
- + А. Структурные данные, свойства и сведения о биологической активности соединений
 - Б. Только физико-химические свойства веществ
 - В. Экономическая информация о производстве
 - Г. Исторические данные о химических открытиях
6. Какая из следующих баз данных специализируется на химических соединениях?
- + А. PubChem
 - Б. GenBank
 - В. Protein Data Bank
 - Г. CCDC
7. Какое из следующих понятий описывает разнообразие химических структур и функций?
- + А. Молекулярное разнообразие
 - Б. Молекулярная масса
 - В. Молекулярная комплементарность
 - Г. Химическая однородность
8. Какую информацию можно получить, изучая молекулярное разнообразие?
- + А. Влияние структуры на биологическую активность
 - Б. Почему молекулы одинаковы
 - В. Физические свойства химических элементов
 - Г. Как строить молекулы
9. Что такое молекулярные дескрипторы?
- + А. Количественные характеристики молекул, используемые для прогнозирования их свойств
 - Б. Физические свойства веществ
 - В. Способы представления химических соединений
 - Г. Визуальные модели молекул
10. Какой из следующих дескрипторов может быть использован для оценки липофильности?
- + А. ЛогР
 - Б. Молекулярная масса
 - В. Температура кипения
 - Г. Площадь поверхности
11. В чем заключается основная идея моделирования «структура-свойство»?

- + А. Связь между химической структурой и ее функциональными свойствами
- Б. Изучение кинетики химических реакций
- В. Анализ спектроскопических данных
- Г. Исследование взаимодействий макромолекул

12. Какой метод может быть использован для предсказания свойств соединений на основе их структур?

- + А. QSAR (количественное соотношение структура-свойство)
- Б. Хроматографический анализ
- В. Групповая обобщающая модель
- Г. Метод молекулярной динамики

Информация о разработчиках

Фироз Неда, ассистент, Институт «Умные материалы и технологии» ТГУ.