Программа учебной дисциплины «Неопределённость и нечёткость при анализе данных и принятии решений»

Направление подготовки 01.04.02. Прикладная математика и информатика

Магистерская программа «Науки о данных»

Форма обучения: очная

Степень: магистр

Утверждена
Академическим советом ОП
Протокол № от __.____.2022

Разработчик	А.Е. Лепский, проф. деп. математики ФЭН
Число кредитов	4
Контактная работа (час.)	56
Самостоятельная работа (час.)	96
Курс, образовательная программа	1
Формат изучения дисциплины	без использования онлайн курса

1. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Неопределенность и нечеткость при анализе данных и принятии решений» являются ознакомление студентов с основными положениями и прикладными возможностями применительно к анализу данных и принятию решений теории нечетких множеств, теории свидетельств и др. теорий моделирования неопределенности.

В результате освоения дисциплины студент должен:

- знать основные подходы описания неопределенности в рамках различных теорий неточных вероятностей и теории нечетких множеств;
- иметь представление об оценивании риска и принятии решений в условиях нечеткой, неточной и неполной информации;
- иметь навыки (приобрести опыт) решения некоторых задач описания неопределенности в рамках теории нечетких множеств и теорий неточных вероятностей и их применения при анализе данных и принятии решений.

Выпускник по направлению подготовки 01.04.02 «Прикладная математика и информатика» с квалификацией (степенью) магистр, должен обладать следующими компетенциями:

Общее число зачетных единиц (ЗЕ)	Профессиональные компетенции (ПК)								
	Инструментальные компетенции (ИК)								Профессиональные задачи
	ПК-9	ЗЕ	ПК-14	ЗЕ	ПК-16	ЗЕ	ПК-18	ЗЕ	
4	РБ; СД	1	РБ; СД	1	РБ; СД	1	РБ; СД	1	НИД-1, НИД-2, НИД-3, НИД-4, НИД-5, НИД-6, НИД-7, НИД-8; ПТД-1, ПТД-2, ПТД-3, ПТД-4, ПТД-5, ПТД-6, ПТД-7, ПТД-8, ОУД-1

Данная учебная дисциплина относится к вариативной части цикла дисциплин специализаций «Интеллектуальные системы и структурный анализ», «Технологии моделирования сложных систем».

Изучение данной дисциплины базируется на следующих дисциплинах:

- дискретная математика;
- геометрия и алгебра;
- математический анализ;
- теория вероятностей и математическая статистика;
- выпуклый анализ;
- методы оптимальных решений;
- анализ данных;
- принятие решений.

Для освоения учебной дисциплины, студенты должны владеть знаниями и компетенциями (в рамках учебных программ НИУ ВШЭ) по линейной алгебре, математическому анализу, теории вероятностей, математической статистике, дискретной математике, выпуклому анализу.

Основные положения дисциплины должны быть использованы в следующих дисциплинах:

- современные методы принятия решений;
- методы машинного обучения и майнинга данных;
- при подготовке и защите выпускной квалификационной работы (ВКР).

2. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Разработчик ПУД может совместить детализацию тематического содержания в таблице, или разместить темы (разделы дисциплины) отдельно, может избежать табличного отображения и привести разбиение на контактные часы и самостоятельную работу в виде описания.

Тема (раздел дисциплины)	**Объем в часах**[1]	**Планируемые результаты обучения (ПРО), подлежащие контролю**	**Формы контроля**
	лк		
	см		
	onl/ср		
Тема 1. Элементы теории нечетких множеств и их применение в анализе данных и принятии решений.	16	решает типовые задания обработки и анализа нечетких данных	домашнее задание на 4-6 недель с защитой в виде презентации
	22		
	50		
Тема 2. Элементы теорий неточ-	8	Решает типовые задачи	

[1] Не заполняется для ПУД, которые не вошли в УП ОП и не запланированы в расписании учебных занятий

ных вероятностей (теории свидетельств, теории монотонных мер и др.) и их применение в анализе данных и принятии решений.	10 36	анализа информации в рамках рассмотренных моделей описания неопределенности	экзамен – письменная работа на 60-100 минут
Часов по видам учебных занятий:	**24** **32** **96**		
Итого часов:	**152**		

Формы учебных занятий:

лк – лекции в аудитории;
см – семинары/ практические занятия/ лабораторные работы в аудитории;
onl – лекции или иные виды работы студента с помощью онлайн-курса;
ср – самостоятельная работа студента.

2.1 Содержание разделов дисциплины

Тема 1. Элементы теории нечетких множеств и их применение в анализе данных и принятии решений.

Нечеткие множества и операции над ними

1. Определение и способы задания нечетких множеств (НМ). Основные понятия теории НМ.
2. Алгебраические операции над НМ и их свойства.
3. Способы построения функций принадлежности НМ.
4. Определение расстояния между НМ. Основные способы определения степени размытия НМ. Применения степени размытия в задачах обработки изображений и анализа согласованности позиций экспертов.

Обобщение операций над нечеткими множествами

5. Понятие t-нормы и t-конормы. Генерация t-норм и t-конорм. Примеры применения.[2]
6. Понятие инвертора (нечеткого отрицания). Описание инверторов в терминах генерирующих функций.

Нечеткие отношения

7. Понятие нечетких отношений (НО) и способы их задания. Операции над НО. Понятие композиции бинарных НО и ее свойств. Примеры.
8. Определения рефлексивных, симметричных и транзитивных НО на декартовом квадрате. Понятие транзитивного замыкания НО. Примеры.
9. Различные виды НО порядка. Критерий предпорядка, связь с отношением неразмытого порядка.
10. Понятие НО подобия, различия и сходства. Связь НО различия с ультраметрикой. Приложение: анализ нечеткого отношения согласованности рекомендаций финансовых аналитиков.

[2] Подразделы, выделенные мелким шрифтом, изучаются при наличии резерва времени.

Принцип обобщения Заде и нечеткие числа

11. Принцип обобщения Заде и его применения.
12. Понятие нечетких чисел (НЧ) и операций над ними. Теорема о связи выпуклости НЧ с выпуклостью срезов. Общий вид и примеры НЧ.
13. Основное аналитическое свойство НЧ. Правила интервальной арифметики. Арифметические операции над НЧ и их свойства.
14. Нечеткие функции, уравнения, системы. Примеры.

Числовые характеристики и расстояния между нечеткими числами

15. Числовые характеристики нечетких чисел.
16. Метрики на множестве нечетких чисел.

Сравнение нечетких чисел

17. Сравнение случайных величин. Основные способы сравнения (ранжирования) НЧ.
18. Применение НЧ в анализе нечетких социальных сетей.

Нечеткая регрессия

19. Классическая задача регрессии и линейная регрессия с нечеткими параметрами.
20. Линейная регрессия с нечеткими данными.

Принятие решений при нечетких данных

17. Общая постановка многокритериальной задачи принятия решений при нечетких данных и методы решений.
18. Модель взвешенной суммы и ее нечеткий аналог. Метод TOPSIS и его нечеткий аналог.

Нечеткая классификация и кластеризация

19. Нечеткая классификация: нечеткие классы – четкие объекты, нечеткие классы – нечеткие объекты. Аксиоматика мер близости нечетких множеств.
20. Нечеткая кластеризация: нечеткий алгоритм с-средних, нечеткий алгоритм Густафсона-Кесселя, возможностный нечеткий алгоритм с-средних.

Элементы нечеткого логического вывода

21. Логические нечеткие высказывания. Понятие нечеткой импликации.
22. Нечеткие и лингвистические переменные.
23. Нечеткие правила дедуктивного вывода. Нечеткое моделирования.

Тема 2. Элементы теорий неточных вероятностей и их применение в анализе данных и принятии решений.

Задачи принятия решений, приводящие к моделям неточных вероятностей

24. Неопределенность при принятии решений.
25. Задачи, приводящие к моделям неточных вероятностей. Кредальные множества.

26. Модели принятия решений на основе теории неточных вероятностей. Парадокс Эллсберга.

Теория свидетельств Демпстера-Шейфера

27. Меры доверия и правдоподобия.
28. Правила комбинирования свидетельств.
29. Метод дисконтирования.
30. Применение теории свидетельств к анализу экспертной информации, агрегированию классификаторов и т. д.

Измерение неопределенности в теории неточных вероятностей.

31. Классические меры неопределенности: энтропия Шеннона и мера Хартли.
32. Меры неопределенности в теории функций доверия. Меры противоречивости (конфликта) в теории функций доверия.

3. ОЦЕНИВАНИЕ

Элементами контроля являются домашнее задание и экзамен.

В домашнем задании предлагается разобрать один из методов работы с нечеткими данными и/или в условиях неопределенности по статейному материалу, применить этот метод к реальным и/или тестовым данным, презентовать результаты исследования.

На экзамене студенту предлагается письменно ответить на ряд вопросов теоретически-практического содержания, в ходе выполнения которых необходимо продемонстрировать как владение теоретическим материалом по курсу, так и навыки решения практических задач.

Оценки всех элементов контроля выставляются по 10-ти балльной шкале с точностью до 0.1 и не округляются.

Предварительная итоговая оценка $О_{Итог}$ вычисляется с учетом неокругленных оценок, полученных при защите домашнего задания $О_{Дз}$ и на экзамене $О_{Экз}$, по формуле

$$О_{Итог} = 0{,}4 \cdot О_{Дз} + 0{,}5 \cdot О_{Экз}.$$

Предварительная оценка округляется до итоговой по арифметическому правилу: если дробная часть числа <0.5, то – в меньшую стороны; в противном случае – в большую сторону.

Студенты, набравшие не менее 7 итоговых баллов (после округления), имеют право на получение 1–2 бонусных баллов (до 10-ти итоговых баллов) по результатам дополнительного собеседования (решение дополнительных задач повышенной сложности, ответы на вопросы повышенной сложности по курсу и пр.).

4. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Пример экзаменационного задания.

1. Понятие t-нормы и t-конормы. Архимедовы и строгие t-нормы. Описание t-норм в терминах генерирующих функций. С помощью степени относительно t-нормы $T(x,y)=\dfrac{2xy}{3-x-y+xy}$ найдите функцию принадлежности для нечеткого множества чисел «очень-очень близких к 2», если $A=(0,2,4)$ – треугольное нечеткое число «близких к 2» чисел.
2. Понятие нечетких чисел (НЧ) и операций над ними. Выпуклость НЧ. Положительные и отрицательные НЧ. Теорема о связи выпуклости НЧ с выпуклостью срезов. Общий вид НЧ. Примеры наиболее популярных НЧ. Найдите в общем виде произведение двух треугольных чисел $A=(a-\alpha,a,a+\alpha)$ и $B=(b-\beta,b,b+\beta)$, если $0<\alpha<a$, $0<\beta<b$.
3. Нечеткая кластеризация: алгоритм c-means. Сделайте один шаг методом c-means при кластеризации на два кластера данных $X=\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3\}$, $\mathbf{x}_1=(0,0)$, $\mathbf{x}_2=(1,2)$, $\mathbf{x}_3=(2,1)$ с $m=2$ и начальными кластерами $A_1=(0.4,0.6,0.8)$, $A_2=(0.6,0.4,0.2)$.
4. Меры доверия и правдоподобия и их свойства. Пусть μ – симметричная мера на алгебре 2^X, где $X=\{x_1,x_2,x_3,x_4\}$. Опишите условия, при которых μ является мерой доверия.
5. Модели принятия решений на основе теории неточных вероятностей. Пусть $X=\{x_1,x_2,x_3\}$, $\mu=0.5\eta_{\langle X\rangle}+0.5P$, где $P\in M_{pr}$ и $P(\{x_i\})=1/3$, $i=1,2,3$, $\xi: X\to\mathbb{R}$, где $\xi(x_i)=i$. Найдите $\underline{E}_{\mathbf{P}}\left[\xi\right]=\inf\limits_{P\in\mathbf{P}(\mu)}\sum\limits_{i=1}^{n}\xi(x_i)P(\{x_i\})$.

5. РЕСУРСЫ

5.1 Рекомендуемая основная литература

№п/п	Наименование
1	Броневич А.Г., Лепский А.Е. Нечеткие модели анализа данных и принятия решений: учебное пособие. — М.: Изд. дом Высшей школы экономики, 2022. — 264 с. (библиотека НИУ ВШЭ)
2	Classic Works of the Dempster-Shafer Theory of Belief Functions. – Editors R.R. Yager, L. Liu, Springer-Verlag Berlin Heidelberg 2008, (база данных электронных книг Springer Book электронной библиотеки НИУ ВШЭ)

5.2 Рекомендуемая дополнительная литература

№п/п	Наименование
3	Wang X., Ruan D., Kerre E.E. Mathematics of Fuzziness – Basic Issues. – Berlin Heidelberg: Springer-Verlag, 2009. (база данных электронных книг Springer Book электронной библиотеки НИУ ВШЭ)
4	Уткин Л. В. Анализ риска и принятие решений при неполной информации. – Санкт-Петербург: Наука, 2007.
5	Grabisch M. Fuzzy integral in multicriteria decision making// Fuzzy Sets and Systems, 69(3), 1995, pp.279-298. (база данных электронных ресурсов Science

	Direct электронной библиотеки НИУ ВШЭ)
6	Grabisch M., Nicolas J.-M. Classification by fuzzy integral: Performance and tests// Fuzzy Sets and Systems, 65(2–3), 1994, pp.255-271. (база данных электронных ресурсов Science Direct электронной библиотеки НИУ ВШЭ)
7	Zadeh L.A. Fuzzy sets// Information and Control, 1965, 8, pp.338-353. (база данных электронных ресурсов Science Direct электронной библиотеки НИУ ВШЭ)
8	Baas S.M., Kwakernaak H. Rating and ranking of multiple-aspect alternatives using fuzzy sets// Automatic 13, 1977, pp.47-58. (база данных электронных ресурсов Science Direct электронной библиотеки НИУ ВШЭ)
9	Lee K.H. First Course on Fuzzy Theory and Applications – Berlin Heidelberg: Springer-Verlag, 2005. (база данных электронных книг Springer Book электронной библиотеки НИУ ВШЭ)
10	The Society for Imprecise Probability: Theories and Applications/ электронный ресурс http://www.sipta.org/index.php

5.3 Программное обеспечение

№п/п	Наименование	Условия доступа/скачивания
1	Python, R	свободное лицензионное соглашение

5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№п/п	Наименование	Условия доступа/скачивания
1	браузер	из внутренний сети университета

5.5 Материально-техническое обеспечение дисциплины

Проектор для лекций и семинаров.

6. ОСОБЕННОСТИ ОРГАНИЗАЦИИ ОБУЧЕНИЯ ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ И ИНВАЛИДОВ

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося и предоставления подтверждающих медицинских документов), а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

- для лиц с нарушениями зрения: индивидуальные задания и консультации;
- для лиц с нарушениями слуха: индивидуальные задания и консультации;
- для лиц с нарушениями опорно-двигательного аппарата: индивидуальные задания и консультации.