

Руководство по изучению курса «Интеллектуальный анализ данных»

Общие рекомендации

Для изучения каждой темы курса предлагаются различные информационные ресурсы: базовый учебник, дополнительная литература, фрагменты электронного пособия, Интернет ресурсы, видео лекции.

Студентам при работе над курсом *рекомендуется* придерживаться предложенной последовательности его освоения:

1. При изучении *каждой темы*:

- ознакомиться со слайдами лекции (их можно рассматривать как расширенный план темы)
- прочитать соответствующий теоретический материал в базовом учебнике или другую предложенную литературу по теме (дополнительный материал)
- выполнить (если есть) практические работы или пройти тест.

2. Выполненные *практические задания* прикрепляются по соответствующей ссылке, например,

Ответ на задание N1

Задание считается выполненным в срок, если выполняется до указанной после задания даты (и в нижеприведенной таблице). В каждом задании есть задачи и вопросы, которые оцениваются указанным количеством баллов. Если задание не выполнено в указанный срок – оценка снижается на 2 балла.

3. Тест можно проходить до 3 раз (включительно) с интервалом между попытками – не менее двух дней. Итоговая оценка

- если тест пройден один раз – совпадает с оценкой за эту попытку;
- если попыток было две – формируется по формуле:
$$(0.6 * \text{Оценка1} + 0.4 * \text{Оценка2});$$
- если попыток было три – формируется по формуле:
$$(0.5 * \text{Оценка1} + 0.3 * \text{Оценка2} + 0.2 * \text{Оценка3}).$$

4. В курсе предусмотрены *лабораторные работы*, которые выполняются с использованием пакета Statistica/Win 7.0.

5. При прохождении материала последнего раздела студенту необходимо подготовить и прислать на проверку *реферат* по одной из предложенных тем и выступить на семинаре. Если реферат не выполнен в указанный срок – оценка снижается на 2 балла.

8. Курс завершается *устным экзаменом*. Для допуска к нему студент должен выполнить практические задания, лабораторные работы, пройти тестирование по всем тестам с результатом не менее, чем на 50% и подготовить реферат по одной из предложенных тем на положительную оценку (от 4 и выше).

В следующей таблице приведены темы курса и соответствующие задания.

	Тема				
		Лекция (часы)	Практ. задание	ЛР	Тест
Раздел 1. История различных этапов анализа данных. Процесс интеллектуального анализа данных. Понимание данных и их подготовка					<u>01.10.2011</u>
	Вводная лекция	2			
	Типы данных	1			
	Качество данных	2	<u>01.10.2011</u>		
	Способы визуального представления данных	1		2	
Раздел 2. Алгоритмы классификации данных с учителем и без					<u>29.10.2011</u>
	Методы кластер-анализа	6	<u>15.10.2011</u>	2	
	Валидация результатов кластер-анализа	2			
	Дискриминантный анализ.	4		2	
	Компьютерные интенсивные методы и их применение к процедуре проверки качества классификации	2			
Раздел 3. Статистические методы решения задачи прогнозирования					<u>12.11.2011</u>
	Регрессионный анализ	2			
	Темпоральные данные. Анализ временных рядов	2	<u>12.11.2011</u>		
Раздел 4. Проблема снижения размерности и методы ее решения					<u>26.11.2011</u>
	Метод главных компонент	2	<u>26.11.2011</u>	2	
	Факторный анализ	4			
Раздел 5. Дополнительные темы курса					
	Web-Mining	2			
	TEXT-Mining	2			
	Семинары	6	<u>До 24.12.2011</u>		
	ИТОГО	40		8	

Информационное обеспечение курса

1. Базовый учебник

И.Яцкив. Многомерный статистический анализ: классификация и снижения размерности. Учебное пособие. Rīga, TSI, 2005. 207р.

содержит материал, необходимый для:

- освоения тем 2 и 4 раздела,
- выполнения практических заданий 2, 3 и 4.
- выполнения лабораторных работ 2, 3 и 4.

Для остальных тем курса приводятся фрагменты будущего электронного пособия автора курса или материалы магистров, писавших выпускные работы под руководством автора курса.

2. Рекомендуемая литература

1. M.Berthold, D.J.Hand. *Intellegent Data Analysis*. N.Y.Springer-Verlag. 2003
2. *Data mining: New Generation Challenges and Future Directions*. Edited H.Kurgupta, A.Joshi, K.Sivakumar, Y.Yesha. MITPress, 2004.
3. А.А.Барсегян, М.С.Куприянов, В.В.Степаненко. *Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP*. СПб: БХВ-Петербург, 2007, 384р.
4. Carlo Vercellis. *Business Intelligence: Data Mining and Optimization for Decision Making*. UK. Wiley, 2009, 410p
5. M.Bramer. *Principles of Data Mining*. Springer, 2007.

3. Дополнительная литература

1. Margaret H.Dunham. *Data Mining: Introductory and Advanced Topics*, Pearson Education, 003, 315p.
2. B.Mirkin. *Clustering for Data Mining: A Data Recovery Approach*. NewYork: Chapman & Hall, 2005.
3. S.Srivastava and ot. *Credit card fraud detection using Hidden Markov Model.*, India, 2008
4. Ronen Feldman, James Sanger. *The Text Mining Handbook. Advanced Approaches in Analyzing Unstructed data*. Cambridge University Press, 2007, 410p. ISBN 0-521-83657-3
5. Daniel T.Larose. *Discovering Knowledge in Data: n introduction in data mining*. Wiley, 2005.
6. *Data Analysis*. Edited Gerard Govaert. ISTE, 2009.
7. R. Nisber, J.Elder, G.Miner. *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier, 2009. ISBN: 978-0-12-374765-5

Полезные интернет ресурсы

- ✓ <http://www.kdnuggets.com/news/index.html>
- ✓ <http://www.thearling.com/index.htm#wps>
- ✓ <http://www.intuit.ru/department/database/datamining/>
- ✓ *Fraud The Facts*. Cardwatch, UK, 2008 citeseerx.ist.psu.edu
- ✓ Rob Walker. *The Evolution and Future of Business Intelligence*. http://www.information-management.com/infodirect/2009_140/business_intelligence_bi-10016145-1.html