



COURSE DESCRIPTION CARD - SYLLABUS

Course name

Data compression methods [S1S11E>MKD]

Course

Field of study

Artificial Intelligence

Year/Semester

2/3

Area of study (specialization)

–

Profile of study

general academic

Level of study

first-cycle

Course offered in

English

Form of study

full-time

Requirements

elective

Number of hours

Lecture

15

Laboratory classes

0

Other

0

Tutorials

15

Projects/seminars

0

Number of credit points

3,00

Coordinators

dr hab. inż. Robert Susmaga

robert.susmaga@put.poznan.pl

Lecturers

Prerequisites

Basic knowledge regarding: a) calculus (logarithmic function, exponential function, function derivative), b) linear algebra (vectors, matrices, vector/matrix operations), c) probability theory and statistics (probability, including conditional probability, random variables, probability distributions and probability distribution functions, mean values, expected values, variance), d) data structures (one- and two-dimensional arrays, lists, trees). Basic skills regarding designing, creating and testing computer programs (in a programming language of one's choice) that implement simple processing of static (vectors and matrices) and dynamic (lists, trees) data structures. (recommended) A fair amount of cognitive curiosity and not less perseverance in pursuing the goals of personal development.

Course objective

The objective of the course is to present aspects of Data Compression in the context of the Information Theory, one of the most fundamental theories underlying theoretical Computer Science of modern-day. The Information Theory deals with representing, storing and communicating information expressed in the form of symbols. Owing to the fact that many important applications of this theory reach far beyond the core of Computing Science, the presented selection of aspects will be confined to the most fundamental ones, mainly those related to such domains of the Computer Science as Data Exploration and Data Compression, in particular: Lossless Data Compression. The fruits of rapid development of notions in the Information Theory, initially disputable and professedly unsolvable, have soon turned out to be incredibly useful and to have a great deal of practical value. In the modern-day these solutions show up in virtually all imaginable computer systems in existence, ones that could hardly survive nowadays without the ubiquitous multimedia content, the popularity and versatility of which has been consistently and unwaveringly influenced by the accessibility of data compression methods. Detailed objectives of the course include sharing skills and knowledge sufficient for: a) understanding the fundamental ideas underlying the Information Theory, especially those pertaining to Lossless Data Compression, b) identifying, formulating and solving basic problems in the Information Theory and Lossless Data Compression, c) designing and creating computer programs that successfully implement the presented methods and algorithms.

Course-related learning outcomes

Knowledge:

The students:

1. have a basic, ordered and well-grounded knowledge essential for important areas of computer science such as algorithmics and programming languages -- [K1st_W2]
2. know and understand the basic techniques, methods, algorithms, and tools used for solving computer problems related to Information Theory and its applications to Data Compression -- [K1st_W4]
3. know and understand the basic techniques, methods, algorithms, and tools used for solving computer problems as well as problems in artificial intelligence, including an automated recognition of patterns in data of different types and their processing -- [K1st_W5]
4. have a basic knowledge of key directions and the most important successes of artificial intelligence understood as an essential sub-domain of computer science, making use of the achievements of other scientific disciplines, like Data Compression to provide solutions with a high practical impact -- [K1st_W5]

Skills:

The students:

1. can collect information from the appropriate sources of different nature, perform its critical analysis, interpretation and synthesis as well as comprehensively draw and justify formulated conclusions regarding the information, especially in the context of Information Theory and its applications to Data Compression -- [K1st_U1]
2. can efficiently plan and carry out experiments, including computer measurements and simulations related to various aspects of Data Compression, interpret the obtained results and draw conclusions based on the experimental outcomes -- [K1st_U4]
3. can retrieve, analyse and transform different types of data and carry out data synthesis to knowledge and conclusions useful for solving a variety of problems that arise in computer science, especially in AI -- [K1st_U10]
4. can -- following a pre-defined specification -- design and create an IT system by first selecting and then using available methods, techniques and computer tools (including programming languages) -- [K1st_U8]
5. can adapt existing algorithms as well as formulate and implement novel ones, including algorithms typical for different streams of AI, using at least one well-known tool -- [K1st_U9]

Social competences:

The students:

1. understand that knowledge and skills quickly become outdated and perceives the need for constant additional training and raising one's qualifications -- [K1st_K1]
2. are aware of the importance of scientific knowledge and research related to computer science and AI in solving problems that are essential for the functioning of individuals, firms, organizations as well as the entire society -- [K1st_K2]

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Learning outcomes presented above are verified as follows:

Formative assessment (laboratory classes): evaluation of the solutions to the assigned programming problems (as they arise).

Final assessment:

- (laboratory classes): evaluation of the solutions to the assigned programming problems (final),
- (lectures): evaluation of the results of a written test (45--60 min) with both multiple choice, short answer and (small) computational questions (mostly: micro-problems to be solved in writing).

Programme content

The course includes, but is not limited to, the following.

Fundamentals of:

- calculus (fundamental properties and graphs of $\exp(x)$, $\log(x)$ and $x \cdot \log(x)$);
- linear algebra (linear / convex combinations of vectors, convex hulls);
- probability theory (discrete random variables): probability, probability distribution functions.

The idea of Shannon information (shortly: information). The measure of information content:

construction and properties. The Hartley measure of information content. The idea of the Shannon entropy (shortly: entropy). Entropy as the measure of information content: construction and properties, in particular: the subdivision property.

Data transfer and the need for encoding/decoding procedures.

Three types of data processing: encoding/decoding, encrypting/decrypting and hashing (with focus on the first).

Encoding/decoding (discrete data): the idea and objectives (including: data transfer reliability). Codes: definitions and examples, tree representations of codes, prefix codes. Kraft's inequality. Data encoding aimed at compressing.

Compression (discrete and continuous data, with focus on the former): idea and objectives (including: data transfer rate). Two basic types of data compression (lossless and lossy, with focus on the former).

Data compression measures (discrete data). Entropy as the compression limit. The universal compressing method.

Lossless compression (discrete data): coding-based compression. Shannon, Shannon-Fano and Huffman encoding: the idea, code trees, optimal codes, algorithms, examples, properties, adaptive approaches.

Lossless compression (discrete data): dictionary-based compression. Lempel-Ziv-Welch method: the idea, representations, dictionaries, algorithms, examples, properties, adaptive approaches.

Miscellaneous compression methods and compression-supporting techniques: Run-Length-Encoding (the idea, algorithms); Move-to-Front encoding (the idea, algorithms), Burrows-Wheeler transform (the idea, algorithms).

Compression systems and platforms.

Optional material (time and progress allowing):

- (encoding-based) adaptive Huffman coding (the idea, algorithms),
- (dictionary-based) Lempel-Ziv-77 and Lempel-Ziv-78 methods (the idea, algorithms),
- (other) arithmetic coding (the idea, coding symbol strings, probability intervals, algorithms).

Course topics

The course includes, but is not limited to, the following.

Fundamentals of:

- calculus (fundamental properties and graphs of $\exp(x)$, $\log(x)$ and $x \cdot \log(x)$);
- linear algebra (linear / convex combinations of vectors, convex hulls);
- probability theory (discrete random variables): probability, probability distribution functions.

The idea of Shannon information (shortly: information). The measure of information content:

construction and properties. The Hartley measure of information content. The idea of the Shannon entropy (shortly: entropy). Entropy as the measure of information content: construction and properties, in particular: the subdivision property.

Data transfer and the need for encoding/decoding procedures.

Three types of data processing: encoding/decoding, encrypting/decrypting and hashing (with focus on the first).

Encoding/decoding (discrete data): the idea and objectives (including: data transfer reliability). Codes: definitions and examples, tree representations of codes, prefix codes. Kraft's inequality. Data encoding

aimed at compressing.

Compression (discrete and continuous data, with focus on the former): idea and objectives (including: data transfer rate). Two basic types of data compression (lossless and lossy, with focus on the former). Data compression measures (discrete data). Entropy as the compression limit. The universal compressing method.

Lossless compression (discrete data): coding-based compression. Shannon, Shannon-Fano and Huffman encoding: the idea, code trees, optimal codes, algorithms, examples, properties, adaptive approaches.

Lossless compression (discrete data): dictionary-based compression. Lempel-Ziv-Welch method: the idea, representations, dictionaries, algorithms, examples, properties, adaptive approaches.

Miscellaneous compression methods and compression-supporting techniques: Run-Length-Encoding (the idea, algorithms); Move-to-Front encoding (the idea, algorithms), Burrows-Wheeler transform (the idea, algorithms).

Compression systems and platforms.

Optional material (time and progress allowing):

- (encoding-based) adaptive Huffman coding (the idea, algorithms),
- (dictionary-based) Lempel-Ziv-77 and Lempel-Ziv-78 methods (the idea, algorithms),
- (other) arithmetic coding (the idea, coding symbol strings, probability intervals, algorithms).

Teaching methods

Lectures: slide show presentations (theoretical elements, explanations, examples, exercises).

Laboratory classes: designing and creating (in a programming language of one's choice) programs that solve the assigned problems (which illustrate the ideas and notions presented during the lectures).

Bibliography

Basic

1. D.J.C. MacKay: Information Theory, Inference, and Learning Algorithms, Cambridge University Press, Cambridge, UK, 2003.
2. T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd Edition, Wiley and Sons, Hoboken, New Jersey, 1991.
3. K. Sayood (red.): Lossless Compression Handbook, Academic Press, Elsevier Science, San Diego, California, 2003.

Additional

1. Lecture notes (slide show presentations)
2. K. Sayood: Introduction to Data Compression, 3rd Ed., Morgan Kaufmann Publishers, San Francisco, California, 2006.
3. A. Drozdek: Wprowadzenie do kompresji danych WNT, Warszawa, 1999 (in Polish).
4. A. Przelaskowski: Kompresja danych. Podstawy, metody bezstratne, kodery obrazów, BTC, Legionowo, 2005 (in Polish).

Breakdown of average student's workload

	Hours	ECTS
Total workload	75	3,00
Classes requiring direct contact with the teacher	30	1,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	45	1,50