TDT 4137 COGNITIVE ARCHITECTURES

Rudolf Mester
Department of Computer Science (IDI), Data and ARTificial Intelligence group (DART)

Course Curriculum (Fall term 2020)

Lecture 1: **Introduction**.
History and motivation. Course structure and organisation. The Nature of Cognition. Some elements of Cognitive Science. Paradigms of Cognitive Systems

Lecture 2: **The Paradigms of Cognitive Architecture**s.
Background, overview. Cybernetics. The cognitivist systems paradigm. Physical symbol system hypothesis, Heuristic search hypothesis (Newell & Simon). Symbol system models. The knowledge level.
The emergent systems paradigm: What are "emergent approaches"? Competence vs. comprehension. Connectionist models, connectionist systems

Lecture 3: **Building blocks of Cognitive Architecture**s
Symbolic vs. sub-symbolic representations. Perception. Attention. Memory. Symbolic vs. sub-symbolic: What is "symbolic"; what are "symbols"? Chinese Room thought experiment (Searle). What is sub-symbolic? Hybrid representations, combine symbolic and sub-symbolic processing More on the building blocks: Perception. Vision. David Marr's model. Image Understanding. The role of deep learning. Audition. Multi-sensory input to Cognitive Systems. Perceptual attention. Action selection, Planning vs. dynamic action selection. Memory: Multi-store memory concept, Long-term memory vs. short-term memory, Sensory memory, Working memory. Knowledge representation.

Lecture 4: **The Soar cognitive architecture**
Soar as an exemplary symbolic cognitive achitecture. Behaviour = Architecture + Content. Core notion in Soar: Problem spaces. "Impasses" and how to deal with them. Memory structure in Soar. Production systems. Recognise-act cycle. Conflict resolution. Working Memory. Decision procedure. Preferences.
Soar example: "Baseball reasoning". Semantic and episodic memories. Preference operators. Chunking. Example: Soar Tool. New elements to Soar: Reinforcement learning in Soar, Visual memory, Perception-Motor interface.

Lecture 5: **ACT-R and ICARUS**
ACT-R : founded in psychological research, links mental functions to brain structures
ACT-R as a theory and a framework (What is a framework?). Relation to physiological measurements (e.g. fMRI). ACT-R cortical models. Declarative memory, Procedural memory. The role of buffers in ACT-R. Buffers and operators. Subsymbolic and symbolic level in ACT-R. Production utility. Chosing productions. Chunk activation. Relevance of context. Retrieval process.
ICARUS: focus on physical agents in their environment. Cognition link to perception and action. Common features with Soar and ACT-R. Distinctive features of ICARUS: focus on embodiment. Categories and skills as separate cognitive entities. Short-term elements, long-term structures. Organization of long-term knowledge. Inference and execution vs. problem solving. ICARUS elements: physical symbol system, list structures, symbolic patterns and pattern matching, relational rules and dynamic composition. Concepts, rules/clauses, beliefs. Cascaded integration in ICARUS. Conceptual inference. Skills and skill execution. intentions. Theory of skill acquisition. Learning in ICARUS.

Lecture 6: **Connectionism and Neural networks**
Reprise: what are emergent approaches and emergent architectures?
Characteristics of connectionist systems: parallel processing. non-symbolic distributed activation patterns in networks. Neural networks as the prominent instantiation of connectism. Biological

inspiration. Another look into the brain. Biological and artificial neurons. History of artificial neural networks (ANNs). Examples: ANN-based and "engineered" systems for autonomous driving in the late 1980ies. Activation functions. Image features recognized by ANNs. Linear classifiers. Geometry of pattern classification. Probabilistic background, Likelihood ratio. The correlation type of linear units. Activation function as "decision maker", Perceptron rule. Limits of single layer networks. Training for Multilayer networks. Towards deep learning. Learned feature hierarchies. Convolutional neural networks. Examples from autonomous driving.

Lecture 7: **Belief-Desire-Intention Agents (BDI Structure). Subsumption Architecture.**
Agent architectures: historical view. A critical look on symbolic AI. Reactive vs. deliberative agents. Hybrid approaches. Belief-Desire-Intention. Autonomy. Proactiveness. Reactivity. Social Ability. Delibeartion. Means-end reasoning. Detailed view on Intentions. Use of intentions in the Observe-think-act loop.
Reactive architectures: R.Brooks: Intelligence is an emergent property of complex behavioural systems. Subsumption Architecture. The role of levels in subsumption architectures. Criticisms of subsumption architectures. Example: Luc Steels' simulated Mars explorer.
Hybrid architectures: horizontal layering, vertical layering. A quick look at the Touring Machine and InteRRaP.

Lecture 8: **Emergence** (guest lecture K. Downing).
Complex Systems, Swarm Intelligence, Complexity. Logistic equation. Chaos and Complexity. Attractors. Dependence on initial conditions. Cellular Automata. Emergence. emergent pattern formation. Synchronic emergence, diachronic emergence. 1st order, 2nd order, 3rd order emergence. Emergent Intelligence., Hypothesis of spatiotemporal levels of emergent intelligence.

Lecture 9: **Reasoning with Uncertainty I**.
Basic inference methods: Deduction, Abduction, Induction. Uncertainty : different forms of uncertainty. Reasoning methods Example: Mycin. Probability Basics, Bayes Theorem. Fuzzy sets. Fuzzy logic reasoning. Fundamentals, Hedges. Mamdani method, Sugeno method. Examples: Restaurant: how should I tip? Mobile robot intersection decision making.

Lecture 10:**Reasoning with Uncertainty II: Bayesian Reasoning**.
Sources of uncertainly. Decision making with uncertainly. Rational behaviour. Maximum expected utility principle. Probability basics (reprise): statistical independence. Conditional independence. Bayes rule. Bayes classifier. Naive Bayesian reasoning. Examples: Text topic classification, Bag of Words approach. Bayesian Networks. Examples: most probable explanation reasoning. Belief revision. Example: student performance modelled by a Bayesian network.

Lecture 11: **Case-based reasoning. Analogy-based reasoning**.
What is a "concept": classical view / probabilitistic view / exemplar view. Knowledge based systems reconsidered. The Case-based-reasoning (CBR) cycle. History of CBR: Roger Schank' and Schema Theory. Scripts. Dynamic memory. From schemata / scripts to cases. MOPs. Example: CYRUS system (Kolodner). Indexing and choosing the right abstraction level. Case retrieval. Example: CREEK. Other CBR approaches.
Analogy-based reasoning. A first definition of analogy. Analogical reasoning. Example: Solar System vs. Atom model. Kinds of similarity. Relations vs. attributes. Example: Dunker's radiation therapy problem. Gentner's structural mapping theory. Systematicity principle. Orders of expressions. An architecture for analogical reasoning. Example: "Heat is like water" analogy. Mapping base and target. Application of the systematicity principle. Candidate inference. MAC/FAC model of similarity-based retrieval. Four phases of analogy reasonin retrieval, elaboration, mapping, justification.

Lecture 12: **Visual Attention in Humans and Machine**s (guest lecture John Tsotsos).

Lecture 13: **Summary: Linking this all together**: Comparisons, Critical evaluation. New Trends.