

범용 인공지능의 현황과 과제

2018. 4. 24.

추형석

기술·공학연구실

소프트웨어정책연구소

목 차

1. 개 요

2. 범용 인공지능의 분류

3. 범용 인공지능의 현황

4. 결 론

목 차

1. 개 요

2. 범용 인공지능의 분류

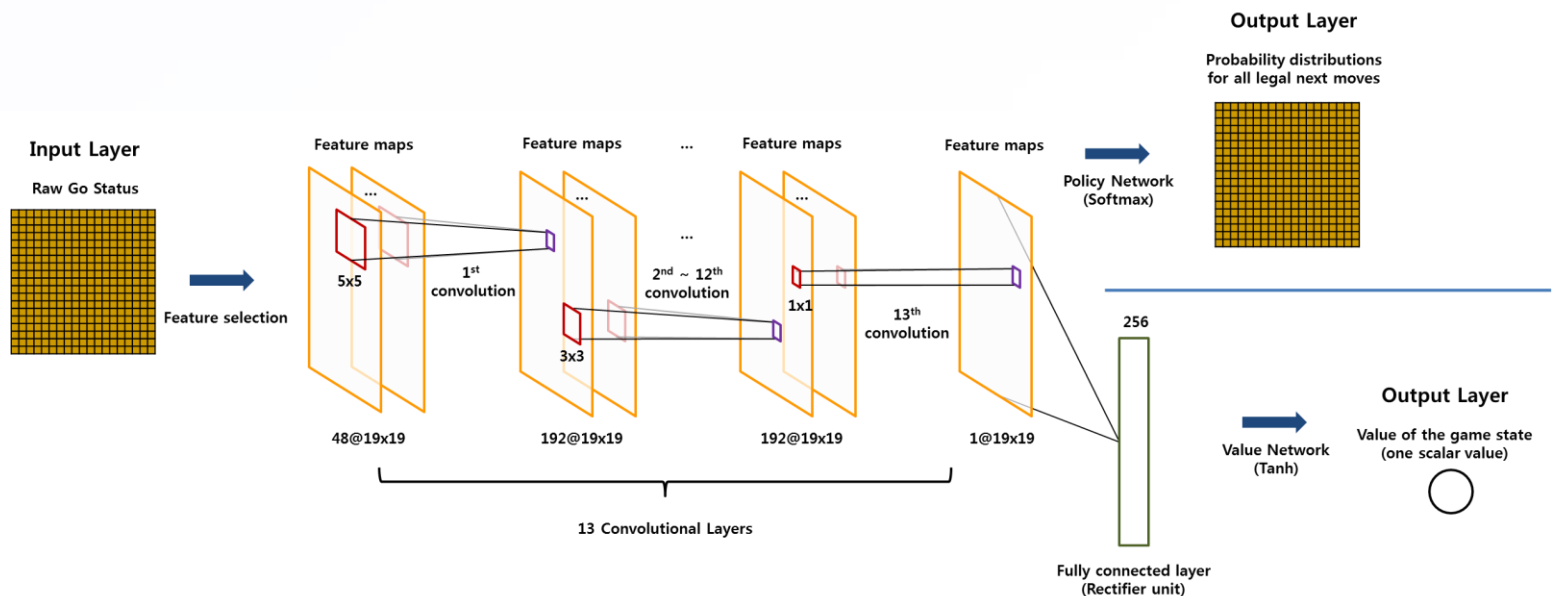
3. 범용 인공지능의 현황

4. 결 론

개 요

● 바둑 인공지능 프로그램 AlphaGo의 진화

- AlphaGo Fan : 바둑 유럽 챔피언 판 후이와 대결하여 5:0 승리
- AlphaGo Lee : 이세돌 9단과 대결하여 4:1 승리 (TPU 활용)
- AlphaGo Master : 온라인 바둑 Tygem에서 60전 전승, 커제 9단과 대결에서 3:0 승리
- AlphaGo Zero : 인간의 기보 없이 스스로 학습하여 최정상 바둑실력 달성
- AlphaZero : AlphaGo Zero의 접근을 일반화하여 체스, 장기 등에서 최정상 달성



범용 인공지능의 부상

● 범용 인공지능(Artificial General Intelligence) 이란?

- 일반적인 개념 : 사람 수준(Human-level) 혹은 사람 수준을 뛰어 넘는 기계
 - 감각적인 정보를 취합하여 문제를 인식하거나 발견하고, 나아가 해결책을 다양하게 모색
- 범용 지능(General Intelligence) : 사람의 일반적인 지능적 행동을 지칭
 - 범용 지능의 정의가 모호하기 때문에, 범용 인공지능을 정의하기는 어려움
- The Core AGI Hypothesis (Ben Goertzel, 2017)
 - 충분히 넓고(인간 수준) 강한 범용성을 갖는 인공지능의 창조나 연구. 질적인 측면에서 상당히 좁고 약한 범용성과 반대되는 개념
 - 이 가설이 맞다면, 범용 인공지능은 좁은 인공지능(narrow AI)의 반대개념으로 정의됨

● 현재 범용 인공지능은 어디까지 왔는가?

- 현재 활발하게 진행 중인 범용 인공지능 연구과제의 소개
- 윤리, 위험, 정책 관점의 범용 인공지능 R&D 분석 소개

참고 문헌

Artificial General Intelligence : Concept, State of the Art, and Future Prospects, Ben Goertzel (2017)



ARTIFICIAL GENERAL INTELLIGENCE

Artificial General Intelligence: Concept, State of the Art, and Future Prospects

early preliminary draft, for comment-solicitation only

Ben Goertzel

BEN@GOERTZEL.ORG *Novaneme LLC*

Editor:

Abstract

Since the AI discipline's founding 6 decades ago, it has yielded many interesting technologies and theoretical results. However it has proved relatively unsuccessful in achieving some of the original core goals of the field, such as the creation of systems with general intelligence as distinct from specialized capabilities and performance on narrowly defined tasks. Subsequently a broad community of researchers has emerged, focusing on the creation and study of synthetic intelligence with broader (e.g. human-level) scope and generalization capability. This paper surveys this diverse community and its progress. Approaches to defining the concept of Artificial General Intelligence (AGI) are reviewed, including mathematical formalisms and more engineering or biology inspired perspectives. The spectrum of designs for AGI systems is reviewed, including systems with symbolic, emergentist, hybrid and universalist characteristics. Prospects for a general theory of general intelligence are considered. Metrics for general intelligence are evaluated, with a conclusion that, although metrics for assessing the achievement of human-level AGI may be relatively straightforward (e.g. the Turing Test, or a robot that can graduate from elementary school or university), metrics for assessing partial progress remain more controversial and problematic.

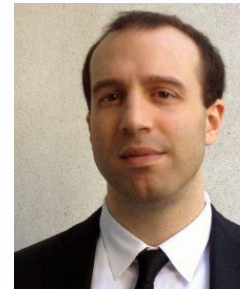
1. Introduction

How can we best conceptualize and approach the original problem regarding which the AI field was founded: the creation of thinking machines with general intelligence comparable to, or greater than, that of human beings? The standard approach of the AI discipline, as it has evolved in the 6 decades since the field's founding, views artificial intelligence largely in terms of the pursuit of discrete capabilities or specific practical tasks. But while this approach has yielded many interesting technologies and theoretical results, it has proved relatively unsuccessful in terms of the original central goals of the field.

Ray Kurzweil (Kurzweil, 2006) has used the term "narrow AI" to refer to the creation of systems that carry out specific "intelligent" behaviors in specific contexts. For a narrow AI system, if one changes the context or the behavior specification even a little bit, some level of human reprogramming or reconfiguration is generally necessary to enable the system to retain its level of intelligence. This is quite different from natural generally intelligent systems like humans, which have a broad capability to self-adapt to changes in their goals or circumstances, performing "transfer learning" (Taylor, Kuhlmann, and Stone, 2008) to generalize knowledge from one goal or context to others. The concept of "Artificial General Intelligence" has emerged as an antonym to "narrow

1

A Survey of Artificial General Intelligence Projects For Ethics, Risk, and Policy, Seth Baum (2017)



Global Catastrophic Risk
INSTITUTE

<http://gcrinstitute.org>

A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy

Seth D. Baum
Global Catastrophic Risk Institute

Global Catastrophic Risk Institute Working Paper 17-1

Cite as: Seth D. Baum, 2017. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. Global Catastrophic Risk Institute Working Paper 17-1.

Note from the author: Comments welcome. Revisions expected. This version 12 November 2017.

Working papers are published to share ideas and promote discussion. They have not necessarily gone through peer review. The views therein are the authors' and are not necessarily the views of the Global Catastrophic Risk Institute.

범용 인공지능의 중요성

- 현대 인공지능 기술은 유례없이 빠르게 발전하는 속도로 인해 미래 인공지능에 대한 기대와 우려가 공존
 - 최초의 범용 인공지능(AGI First) 개발을 위한 경쟁 치열
 - 범용 인공지능 시대를 준비하기 위한 인문·사회적 연구 활발
 - 악의적인 인공지능 활용에 대한 대비는?
 - 범용 인공지능을 공개SW화 하는 것이 합당한 논리인가?
 - 군사적 수요에 의해 추진되는 범용 인공지능을 제어할 방법은?

“카이스트 보이콧” 선언한 세계 학자들, AI무기 개발 비판

등록 :2018-04-05 16:34 수정 :2018-04-05 20:46



인공지능·로봇 분야 50여명 “관계 끊겠다” 공개편지
‘킬러로봇’ 우려에 카이스트 “자율무기 연구 아니다”

범용 인공지능의 측정 방법 (Metric)

- **Turing Test (Alan Turing)**

- 기계가 인간과 얼마나 비슷하게 대화할 수 있는지 판별하는 테스트

- **Coffee Test (Wonzniak)**

- 커피를 만드는 과정을 테스트 (커피 머신 인지, 커피 탐색 등)

- **Robot University Student Test (Goertzel)**

- 인공지능이 사람과 비슷하게 대학 수업을 수강하고 졸업하는 테스트

- **Employment Test (Nilson)**

- 기계가 경제적으로 중요한 일자리에서 사람만큼 업무를 수행하는지 테스트

- **Artificial Scientist Test (Goertzel)**

- 과학적인 연구 수행 가능성을 테스트 (연구 주제 탐색, 관련 문헌 습득, 논문 작성)

목 차

1. 개 요

2. 범용 인공지능의 분류

3. 범용 인공지능의 현황

4. 결 론

범용 지능의 다양한 해석

접근 방법	내 용
실용적 관점	<ul style="list-style-type: none"> - 사람의 지능과 인공지능의 기능적 측면을 비교하는 접근 - 대표적인 예 : 튜링 테스트
심리학적 관점	<ul style="list-style-type: none"> - 지능의 해석 : 개인 내(intra-individual) 다양성, 개인 간(inter-individual) 다양성 - 인간의 지능 분류 : ① 언어적, ② 논리적·수학적, ③ 음악적, ④ 운동감각적, ⑤ 공간적, ⑥ 대인관계, ⑦ 자기 내부적, ⑧ 자연주의적 지능 (Gardner, 1999)
인지 아키텍처 관점	<ul style="list-style-type: none"> - SW의 관점에서 범용 지능의 해석 - 다양한 작업을 수행함에 있어 SW구조는 고정돼야 함
수학적 관점	<ul style="list-style-type: none"> - 범용 지능을 구현하기 위해서는 무한대의 계산 자원이 필요 - 제한된 계산 자원으로는 상대적으로 우수한 범용 지능을 구현
적응주의적 관점	<ul style="list-style-type: none"> - 제한된 자원을 활용해 환경에 적응한다는 접근

범용 인공지능의 분류 (Seth Baum)

지성주의 (Intellectualist)

- 고도화된 지식을 추구
- 안전에 대한 고려가 낮음
- 인공지능 학문의 본질적인 연구주제
- 범용 인공지능 구현 자체에 가치를 둠

인도주의 (Humanitarian)

- 인류를 이롭게 하는 인공지능 개발 목표
- 안전에 대한 고려가 높음
- 인류가 필요로 하는 인공지능 개발에 중점
- 인간적인 윤리를 인공지능에 구현하기 어려움

지성주의 + 인도주의 (Hybrid)

범용 인공지능의 분류 (Ben Goertzel)

기호적 범용 인공지능 (Symbolic AGI)

- 물리적인 기호 시스템 가설에 의해 지능은 기호를 다루는 행위로 정의하는 분야
- 장점 : 기호 시스템은 인간과 동물이 구분되는 가장 큰 특징으로 기호적 관점이 일반화에 가장 적합한 도구
- 단점 : 기호적 관점은 간단한 지각이나 동기로 인한 기호의 진화를 고려해야 함
- 예시 : 기호적 인지 아키텍처

창발적 범용 인공지능 (Emergentist AGI)

- 인간의 뇌의 작동 방식을 모사하여 구현하는 인공지능
- 장점 : 인간의 뇌는 고차원적인 감각 정보를 통해 복잡한 행동을 수행하기 때문에, 뇌를 모사한 접근방법이 우수할 가능성이 높음
- 단점 : 인간의 뇌는 신경망의 자가 구성으로 범용 지능을 달성하므로 이를 구현해야 함
- 예시 : 계산 뇌과학, 인공 생명, 로봇틱스

혼합 범용 인공지능 (Hybrid AGI)

- 기호적·창발적 접근을 혼합한 형태
- 인간 수준의 인공지능을 달성하기 위해서는 기호적·비기호적 접근의 융합이 필요(Nilson)

목 차

1. 개 요

2. 범용 인공지능의 분류

3. 범용 인공지능의 현황

4. 결 론

범용 인공지능 R&D 현황

		Humanitarian	Both	Intellectualist	Profit	Other	Unspecified
Engaged On Safety	Academic		AERA LIDA MIL NARS	FLOWERS			
	Corporate	CommAI Vicarious DeepMind Gusano Real AI	GoodAI	Maluuba	Maluuba	Gusano	
	Other	OpenAI Whole Brain Architecture Initiative	AIDEUS				CogPrime
Not Engaged On Safety	Academic		Blue Brain Human Brain Project	ACT-R CLARION Icarus Leabra Sigma SNePS SOAR MicroPsi SiMA		Human Brain Project	AIXI DeSTIN MLECOG Alice in Wonderland Animats
	Corporate	Nigel Uber AI Labs	HTM MSR AI	NNAISENSE	NNAISENSE	Victor	Cyc Baidu Research Tencent AI Lab
	Other	SingularityNet	China Brain Project	Research Center for Brain-Inspired Intelligence		SingularityNet	Becca DSO-CA MIL

Corporate-Humanitarian-Engaged On Safety Cluster

Academic-Intellectualist-Not Engaged On Safety Cluster

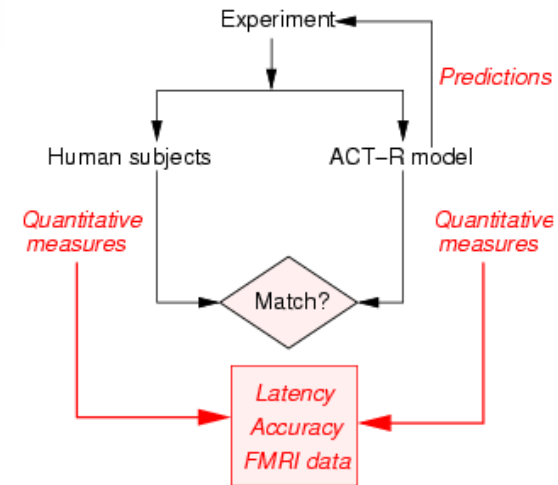
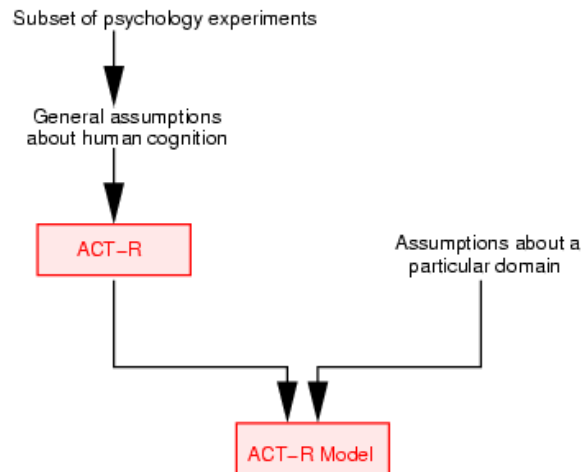
US academic military sub-cluster

공개된 정보를 활용해 수집한 45개의 주요 범용 인공지능 R&D 현황
 기준 : 기관 형태, 공개SW 여부, 군사적 연결성, 국가, 목표, 안전에 대한 고려, 규모

기호적(Symbolic) 범용 인공지능 (1/2)

● ACT-R (Adaptive Control of Thought – Rational)

- 인지 아키텍처(Cognitive Architecture)의 한 종류로 사람의 인지 과정을 표현한 이론 (Theory)
- 일종의 SW 프레임워크로 하노이 탑 문제, 문서나 단어의 기억, 언어 이해, 의사 소통, 비행기 조종 등 다양한 분야에 적용
- 책임자 / 기관 : John Robert Anderson / CMU
- 최신 버전 : 7.5 (2017.07.11.)



기호적(Symbolic) 범용 인공지능 (2/2)

● Cyc

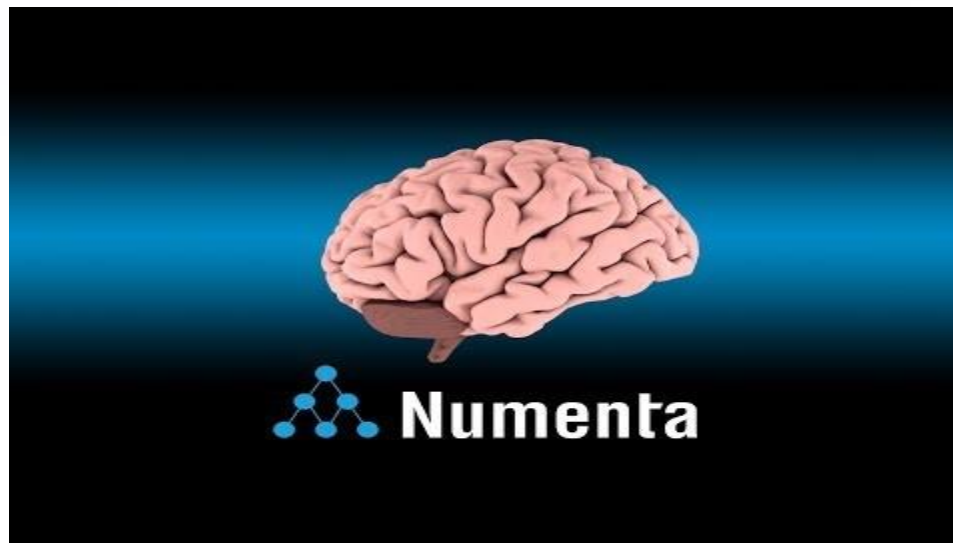
- 1984년 부터 시작된 범용 인공지능 프로젝트로, 수백만 가지의 사람의 일반 상식을 지식 베이스(Knowledge Base)로 구축
 - 지식 베이스 구축에 전문 인력이 직접 참여
- 책임자 / 기관 : Douglas Lenat / Cycorp
- 최신버전 : ResearchCyc 6.1 (2017.11.27.)



창발적(Emergentist) 범용 인공지능 (1/3)

- **HTM (Hierarchical Temporal Memory)**

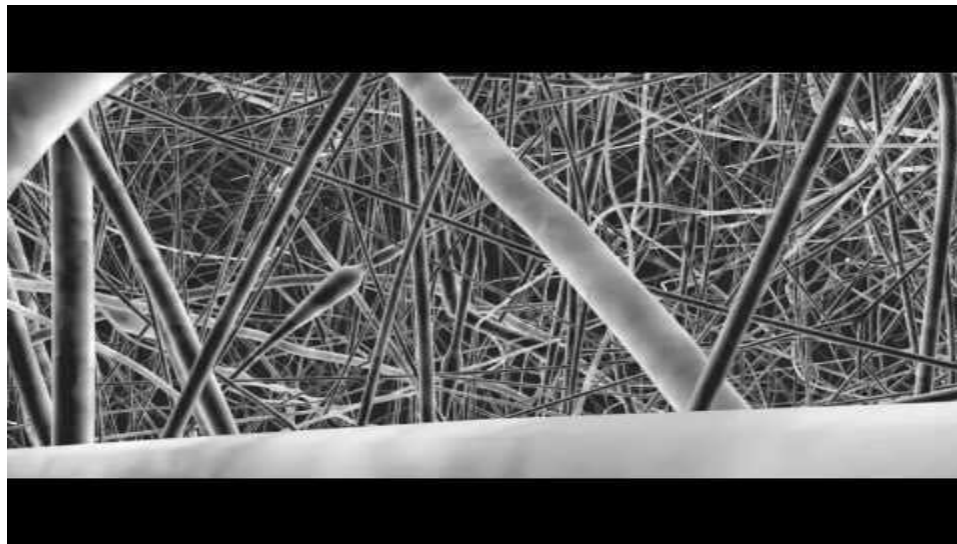
- 인간의 뇌(신피질 부위)의 동작 방식을 이해하고 모델링 함으로써 범용 인공지능을 구현하는 연구 수행 (2002년 최초 설립)
- 주요 적용 분야 : 개인별 에너지 소비 예측, 데이터를 통한 복잡한 기계의 오작동 예측
- 기관 : Numenta
- 공개SW 커뮤니티 적극 운영



창발적(Emergentist) 범용 인공지능 (2/3)

● Blue Brain Project

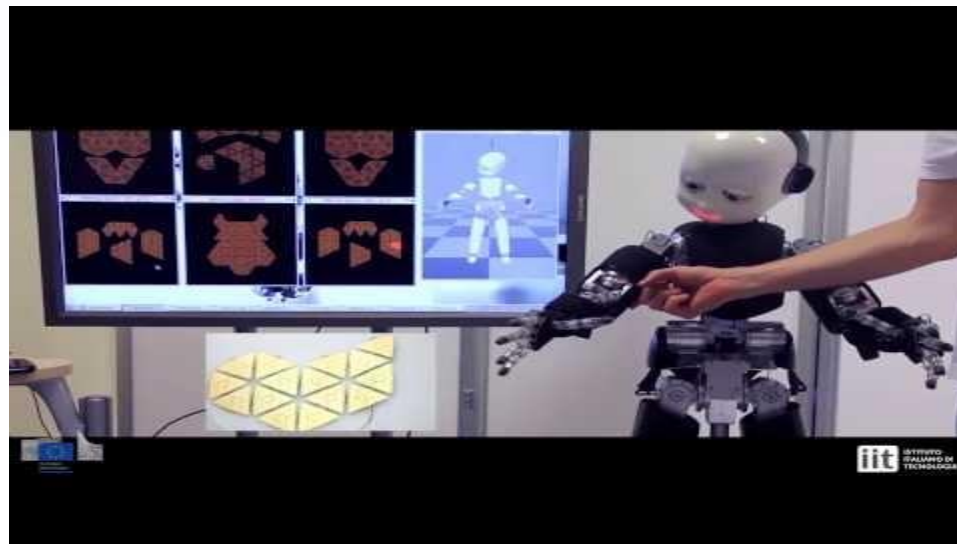
- 역 공학을 통해 포유류의 뇌를 디지털화하여 재구축 하는 것이 목표 (슈퍼컴퓨터 활용)
- 뇌 기능에 대한 실험적인 데이터를 활용, 궁극적으로 사람의 뇌를 모사
- 책임자 / 기관 : Henry Markram / EPFL (École Polytechnique Fédérale de Lausanne)



창발적(Emergentist) 범용 인공지능 (3/3)

- iCub(Cognitive Universal Body) Humanoid Robot

- 어린이 수준의 지능과 행동을 구현하기 위한 로봇으로, 주변 환경을 인식하고 학습을 통해 지능적 행동을 수행
- 가능한 행동 : 기어가기, 3차원 미로 찾기, 화살 쏘기, 안면 감정 표현, 힘 제어 등
- 기관 : Italian Institute of Technology
- 오픈소스 휴머노이드 로봇



목 차

1. 개 요

2. 범용 인공지능의 분류

3. 범용 인공지능의 현황

4. 결 론

결론

- 범용 인공지능 기술은 그 정의(definition)도 구체화되지 않았기 때문에, R&D 통해 정립해나가는 과정 상에 있음
- 범용 인공지능의 본격적인 출현은 아직 시기상조라고 판단되나, 범용 인공지능이 세상을 변화시킬 잠재력이 높기 때문에 이를 대비하기 위한 노력이 필요