

Artificial Intelligence: Representation of Knowledge & Beyond Part - 1



Niladri Chatterjee, Ph.D.
Chair Professor of Artificial Intelligence
Indian Institute of Technology Delhi
Email: niladri@maths.iitd.ac.in

Preamble...

In the last less than one decade two terms have gained huge attention in the computer science community:

- AI and Machine Learning
- Data Analytics

Preamble...

In the last less than one decade two terms have gained huge attention in the computer science community:

- AI and Machine Learning
- Data Analytics

This attention is visible across different domains :
Finance, Medical, Legal, Administration

Preamble...

In the last less than one decade two terms have gained huge attention in the computer science community:

- AI and Machine Learning
- Data Analytics

This attention is visible across different domains :

Finance, Medical, Legal, Administration

In short, in modern times any domain dealing with huge database and aim at drawing inference from there relies heavily on AI based techniques

Preamble...

In the last less than one decade two terms have gained huge attention in the computer science community:

- AI and Machine Learning
- Data Analytics

This attention is visible across different domains :

Finance, Medical, Legal, Administration

In short, in modern times any domain dealing with huge database and aim at drawing inference from there relies heavily on AI based techniques

Library and Information Sciences is no exception.

Preamble...

Natural question would be: How & Why

Preamble

Natural question would be: How & Why

To understand this we have go back to the history of Artificial Intelligence

Artificial intelligence (AI) is an area of computer science that aims at reducing the gap between man and machine. Or in other words, it aims at creating machines that behaves or reacts like humans.

Artificial intelligence (AI) is an area of computer science that aims at reducing the gap between man and machine. Or in other words, it aims at creating machines that behaves or reacts like humans.

Hence machines are empowered with different human abilities both mental and physical - which lead to various directions of research:

Artificial intelligence (AI) is an area of computer science that aims at reducing the gap between man and machine. Or in other words, it aims at creating machines that behaves or reacts like humans.

Hence machines are empowered with different human abilities both mental and physical - which lead to various directions of research:

How to make a computer understand language: **Natural Language Processing**

Artificial intelligence (AI) is an area of computer science that aims at reducing the gap between man and machine. Or in other words, it aims at creating machines that behaves or reacts like humans.

Hence machines are empowered with different human abilities both mental and physical - which lead to various directions of research:

How to make a computer understand language: **Natural Language Processing**

How to make a computer see & react: **Computer Vision**

Artificial intelligence (AI) is an area of computer science that aims at reducing the gap between man and machine. Or in other words, it aims at creating machines that behaves or reacts like humans.

Hence machines are empowered with different human abilities both mental and physical - which lead to various directions of research:

How to make a computer understand language: **Natural Language Processing**

How to make a computer see & react: **Computer Vision**

Note that

All these AI based systems require certain amount of expert knowledge to be coded into the system for ready access.

For illustration:

Medical Systems :

Disease, Symptom, Cause, Medicine, Virus

For illustration:

Medical Systems :

Disease, Symptom, Cause, Medicine, Virus

Legal System:

Laws, by-laws, Clauses, Articles, Past judgments

For illustration:

Medical Systems :

Disease, Symptom, Cause, Medicine, Virus

Legal System:

Laws, by-laws, Clauses, Articles, Past judgments

Library System:

Books, Authors, Journals, Publishers, Subjects

For illustration:

Medical Systems :

Disease, Symptom, Cause, Medicine, Virus

Legal System:

Laws, by-laws, Clauses, Articles, Past judgments

Library System:

Books, Authors, Journals, Publishers, Subjects

Hence to make a computer act like human being it has to be imparted with knowledge.

And this thinking started many years back:

For example:

1956: the term Artificial intelligence - John McCarthy

1959 (~) : General Problem Solver - Simon Shaw & Newel

And this thinking started many years back:

For example:

1956: the term Artificial intelligence - John McCarthy

1959 (~) : General Problem Solver - Simon Shaw & Newel

Hence research was on collecting and storing all human knowledge
and store in a machine.

Thus two important aspects of AI or AI systems are:

Thus two important aspects of AI or AI systems are:

- Knowledge Acquisition
- Knowledge Representation

Thus two important aspects of AI or AI systems are:

- Knowledge Acquisition
- Knowledge Representation

In early AI systems this knowledge was acquired from domain experts.

Thus two important aspects of AI or AI systems are:

- Knowledge Acquisition
- Knowledge Representation

In early AI systems this knowledge was acquired from domain experts.

But elicitation of knowledge from experts have many difficulties:

- Expert may not be available

Eg. Accident Prediction, Dengue spread prediction, Share Price Prediction

- Expert may not be available

 - Eg. Accident Prediction, Dengue spread prediction, Share Price Prediction

- Experts may differ in opinion

 - Eg. Legal systems, Medical systems

- Expert may not be available

 - Eg. Accident Prediction, Dengue spread prediction, Share Price Prediction

- Experts may differ in opinion

 - Eg. Legal systems, Medical systems

- Experts may not be able to articulate knowledge

 - Eg. From Intuition, Experience

- Expert may not be available

 - Eg. Accident Prediction, Dengue spread prediction, Share Price Prediction

- Experts may differ in opinion

 - Eg. Legal systems, Medical systems

- Experts may not be able to articulate knowledge

 - Eg. From Intuition, Experience

And more interestingly

Expert may not even exist!!

Hence question is :

“Where from the knowledge required for developing Modern AI systems may be acquired”

Hence question is :

“Where from the knowledge required for developing Modern AI systems may be acquired”

The solution comes from a novel perspective – viz. data

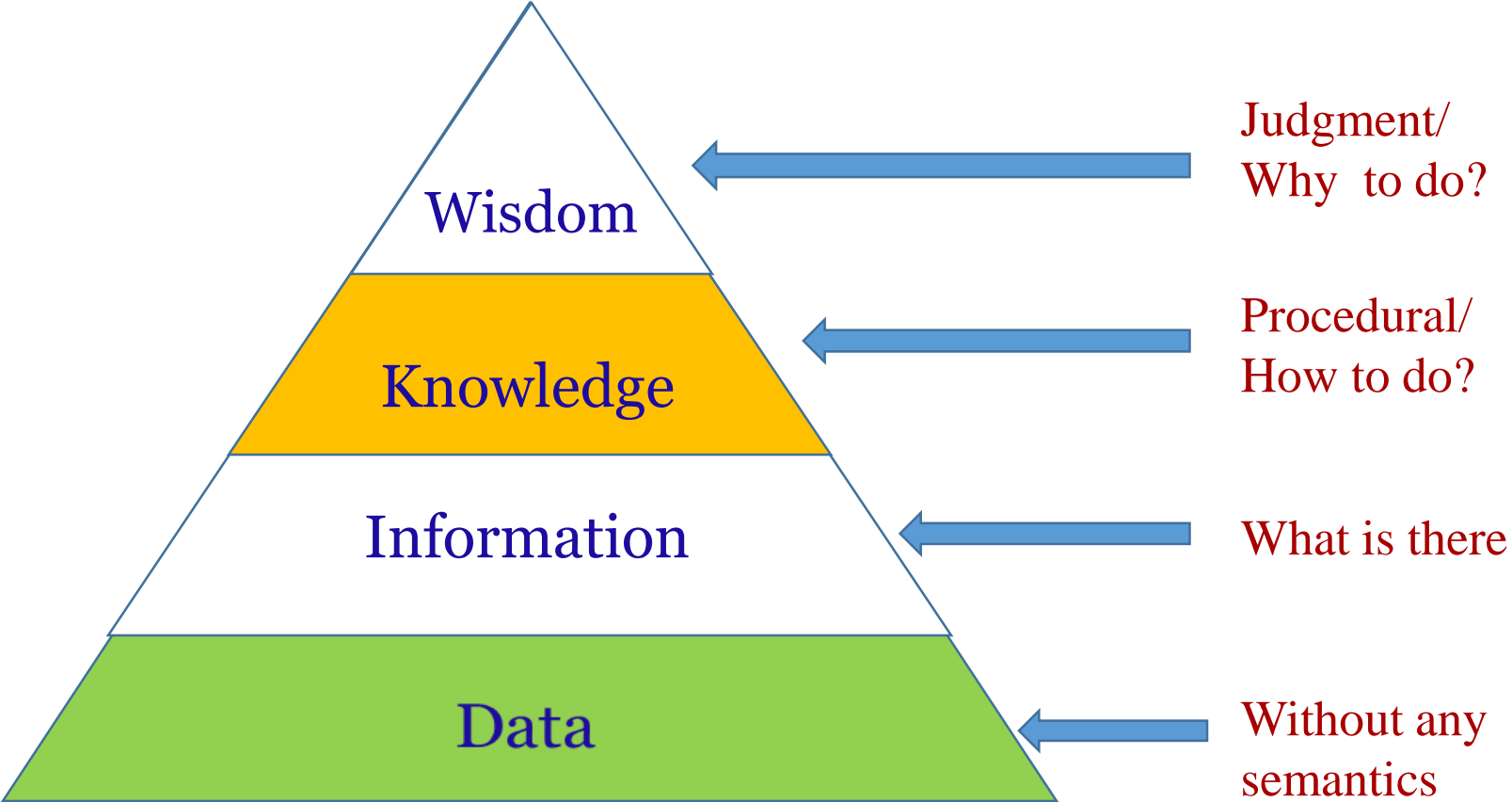


Illustration...

Consider a table of paired numbers $\{(x_i, y_i)\}$, where

x_i is the Accession number of the i^{th} book

y_i is the registration number of the i^{th} user

The raw file is your **data**.

Illustration...

Consider a table of paired numbers $\{(x_i, y_i)\}$, where

x_i is the Accession number of the i^{th} book

y_i is the registration number of the i^{th} user

The raw file is your **data**.

Suppose we take the frequency which book is borrowed how many times.

This gives us a shorter table of the form $\{(x_i, n_i)\}$, where

n_i is the number of times the i^{th} book is borrowed

This table gives you **information** about the popularity of the books

Illustration...

Consider a table of paired numbers $\{(x_i, y_i)\}$, where

x_i is the Accession number of the i^{th} book

y_i is the registration number of the i^{th} user

The raw file is your **data**.

Suppose we take the frequency which book is borrowed how many times.

This gives us a shorter table of the form $\{(x_i, n_i)\}$, where

n_i is the number of times the i^{th} book is borrowed

This table gives you **information** about the popularity of the books.

The above information gives the librarian the **knowledge** about the popular subjects and the demand of the related books among the users. This helps the librarian to decide the books to be purchased in the next lot.

Illustration

Consider a table of paired numbers $\{(x_i, y_i)\}$, where
 x_i is the Accession number of the i^{th} book
 y_i is the registration number of the i^{th} user

The raw file is your **data**.

Suppose we take the frequency which book is borrowed how many times.

This gives us a shorter table of the form $\{(x_i, n_i)\}$, where

n_i is the number of times the i^{th} book is borrowed

This table gives you **information** about the popularity of the books.

The above information gives the librarian the **knowledge** about the popular subjects and the demand of the related books among the users. This helps the librarian to decide the books to be purchased in the next lot.

The above knowledge taken over several years and observing their consequences give the librarian the desired **wisdom** about how to plan regarding purchase and writing off the books.

Thus we have

Data => Information => Knowledge => Wisdom

Thus we have

Data => Information => Knowledge => Wisdom

Lot of data is now available in electronic form:

- Social Media
- Publishers catalog
- Advertisements
- Library reports
- Book reviews
- Pamphlets from institutes

Thus we have

Data => Information => Knowledge => Wisdom

Lot of data is now available in electronic form:

- Social Media
- Publishers catalog
- Advertisements
- Library reports
- Book reviews
- Pamphlets from institutes

We need to exploit the data being generated around us to extract knowledge

Thus we have

Data => Information => Knowledge => Wisdom

Lot of data is now available in electronic form:

- Social Media
- Publishers catalog
- Advertisements
- Library reports
- Book reviews
- Pamphlets from institutes

We need to exploit the data being generated around us to extract knowledge

This leads to Machine Learning

How to Manage the Knowledge?

- * The first step in this direction is searching for desired information in data
- * Typical search is over the web pages scattered all over the internet.
- * In general it is a keyword based search.
- * But that does not give the desired result.



Who teaches AI in IIT Delhi



All

News

Images

Videos

Maps

More

Settings

Tools

About 29,500 results (0.25 seconds)



IITs plan master's makeover to arrest high attrition rate

Economic Times - Oct 8, 2019

At IIT Delhi, where the dropout rate has been much higher at 50%, the institute is ... including the roll-out of master's in **artificial intelligence (AI)**.



IIT Madras students develop AI-based 'Eye in the Sky' for ...

Livemint - 19 hours ago

New Delhi: Indian Institute of Technology (IIT) Madras students have developed an **Artificial Intelligence(AI)**-based solution called 'Eye in the ...

IIT Madras students develop an 'Eye in the Sky' for disaster relief

Local Source - **Times of India** - 15 hours ago



Tracking fires using AI

Livemint - Oct 4, 2019

Gurugram-based collective Blue Sky Analytics has been using **AI** and ... Centre of Excellence for Research on Clean Air (Cerca) at IIT, Delhi in ...



All

Images

News

Maps

Videos

More

Settings

Tools

About 10,500 results (0.27 seconds)

[Extramarks Education deploys AI chatbot to solve students ...](#)[The Hindu BusinessLine](#) - Oct 3, 2019

Delhi-based Extramarks Education has launched an **Artificial** ... segment and test preparation modules for **IIT-JEE** and medical entrance exams. ... The company also has **teachers** available at their centres round the clock to ...

[Surgical Strike Helmets, Defence AI: How IIT Ropar Beat Other ...](#)[News18](#) - Sep 17, 2019

Its score card reads: **Teaching** - 28.1; Research - 12.8; Citations – 100; ... The older IITs like Madras, Bombay and Delhi had taken 50 years to ...

[From research cafes to instant placements: 5 reasons you ...](#)[India Today](#) - Sep 18, 2019

IIT Delhi is set to hold its third Industry Day on September 21 and the ... **AI** based innovation in the area of healthcare, clean energy solutions to ... are two main career paths open for research scholars **teaching** or industry.

Therefore knowledge should be organized in a semantic space.

Note: Understanding semantics automatically is very difficult.

Ex 1: The bank is near my house.

Therefore knowledge should be organized in a semantic space.

Note: Understanding semantics automatically is very difficult.

Ex 1: The bank is near my house.

Ex 2: The computer is printing data. It is fast.

The computer is printing data. It is alphanumeric.

Therefore knowledge should be organized in a semantic space.

Note: Understanding semantics automatically is very difficult.

Ex 1: The bank is near my house.

Ex 2: The computer is printing data. It is fast.

The computer is printing data. It is alphanumeric.

Ex 3: The book is interesting.

Do you think the book is interesting?

Therefore knowledge should be organized in a semantic space.

Note: Understanding semantics automatically is very difficult.

Ex 1: The bank is near my house.

Ex 2: The computer is printing data. It is fast.

The computer is printing data. It is alphanumeric.

Ex 3: The book is interesting.

Do you think the book is interesting?

Ex 4: Deepika and Ranveer are married now.

Virat and Ranveer are married now.

- The search should not be *key based*
- Knowledge based search capabilities on *conceptual spaces*.
- Search should spread over several documents
- Query answering capabilities - enabling users to find, share, and combine information more easily
- The information can be readily interpreted by machines, without human intervention

- The search should not be *key based*
- Knowledge based search capabilities on *conceptual spaces*.
- Search should spread over several documents
- Query answering capabilities - enabling users to find, share, and combine information more easily
- The information can be readily interpreted by machines, without human intervention

But how to do it?

Information Semantics...

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.

Information Semantics...

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.
- At word level: **synonyms, homonyms** are often useful.

Information Semantics...

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.
- At word level: **synonyms, homonyms** are often useful.
- For structured information – we look for **Structures!**
And try for mapping between structures.

Information Semantics...

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.
- At word level: **synonyms, homonyms are often useful.**
- For structured information – we look for **Structures!**
And try for mapping between structures.

But most often one-to-one mappings are not applicable.

Information Semantics...

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.
- At word level: **synonyms, homonyms** are often useful.
- For structured information – we look for **Structures!**
And try for mapping between structures.

But most often one-to-one mappings are not applicable.
- Hence discovering semantics is primary.

Information Semantics

- The primary aim is to achieve: **Interoperability**
Each Semantic conflict needs to be resolved for this.
- At word level: **synonyms, homonyms** are often useful.
- For structured information – we look for **Structures!**
And try for mapping between structures.

But most often one-to-one mappings are not applicable.
- Hence **discovering semantics** is primary.
- One simple technique is called: **Semantic Annotation.**

Semantic Annotation

Creating semantic labels/tags within documents to allow automated processing of documents

A Realistic Example

The term "rabi" means "spring" in Arabic, and the rabi crops are grown between the months mid November to April. The water that has percolated in the ground during the rains is main source of water for these crops. Rabi crops require irrigation. So a good or bountiful rain may tend to spoil the Kharif crops but it is good for Rabi crops.

Annotation

The term “<*rabi* : **croptype**>” means “<*spring* : **season**>” in <*Arabic* : **language**>, and the <*rabi* : **croptype** ><*crops*: **agrithing**> are grown between the months mid <*November*: **month**> to <*April*: **month**>. The <*water* : **naturalthing**> that has percolated in the <*ground* : **naturalthing**> during the <*rains* : **naturalthing**> is main source of <*water* : **naturalthing**> for these crops. <*Rabi* : **croptype** > crops require irrigation. So a good or bountiful <*rain* : **naturalthing**> may tend to spoil the <*kharif*: **croptype**> <*crops*: **agrithing**> but it is good for <*rabi* : **croptype**><*crops*: **agrithing**>.

Such representation is much more explicit form knowledge
Sharing and interoperability.

One can extrapolate and discover:

- November and April are months.
- November to April is a sequence of months.
- A sequence of months is a season

A Thing is something that physically exist.

- Things may be of several types: NATURAL, AGRI etc.
- Crop is a thing of Agriculture type.
- Soil, water are Natural things

Such representation is much more explicit form knowledge
Sharing and interoperability.

One can extrapolate and discover:

- November and April are months.
- November to April is a sequence of months.
- A sequence of months is a season

A Thing is something that physically exist.

- Things may be of several types: NATURAL, AGRI etc.
- Crop is a thing of Agriculture type.
- Soil, water are Natural things

Thus some knowledge can be elicited.

But how to store it conceptually for computer to use?

The knowledge extracted from a set of documents can be represented in many different ways.

- Graph Based Representation.
- Markup language based textual representation

Etc.



THANK YOU

