Communications between intelligent machines
Spaces of semantic information

Huawei Advanced Wireless Technology Lab. Paris

www.huawei.com

Author: Jean-Claude Belfiore - Joint work with Daniel Bennequin
Place and Date: WWRF46 - Paris/Online - December 1st, 2021

HUAWEI TECHNOLOGIES CO., LTD.

HUAWEI

Part I

Who?

Outline

Nowadays

Historical figures

Collaborator

Daniel Bennequin
completed his secondary education at Lycée Condorcet and then graduated from the École normale supérieure. He received his habilitation (Doctoral d'Etat) in 1982 from the University of Paris VII under Alain Chenciner with thesis Entrelacements et équations de Pfaff. He was a professor at the University of Strasbourg before becoming a professor at the University of Paris VII (Institut Mathématique de Jussieu). Bennequin's dissertation was a major contribution to contact geometry, in which he gave the first example of an exotic contact structure embedded in Euclidean 3-space. On the basis of their work in the 1980s, Bennequin and Yakov Eliashberg might be considered the founders of contact topology. Bennequin has also worked on motion planning. He was a member of Bourbaki. *(Source Wikipedia)*

Collaborator

entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

The Homological Nature of Entropy †

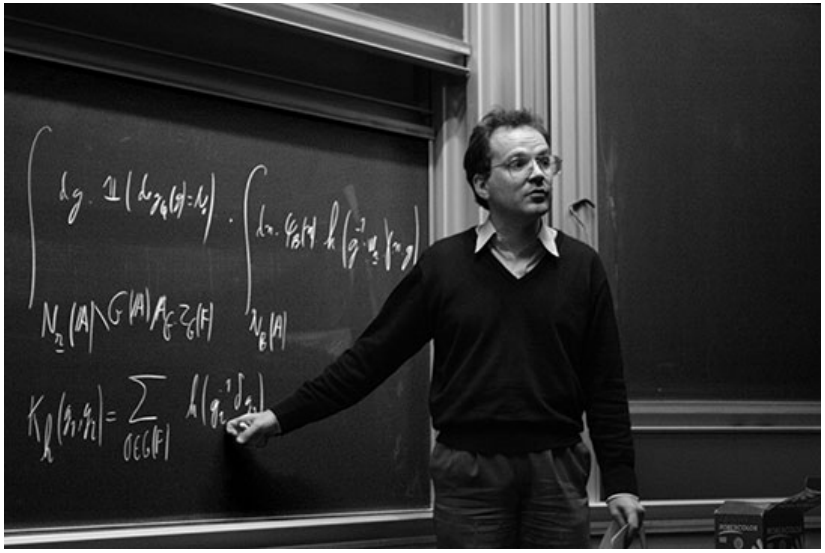
Pierre Baudot [1,*] and Daniel Bennequin [2]

[1] Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany
[2] Universite Paris Diderot-Paris 7, UFR de Mathematiques, Equipe Geometrie et Dynamique, Batiment Sophie Germain, 5 rue Thomas Mann, 75205 Paris Cedex 13, France;
E-Mail: bennequin@math.univ-paris-diderot.fr

† This paper is an extended version of our paper published in Proceedings of the MaxEnt 2014 Conference on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Amboise, France, 21–26 September 2014.

* Author to whom correspondence should be addressed; E-Mail: baudot@mis.mpg.de;
Tel.: +49-341-9959-601.

Inspirator 1: Laurent Lafforgue



Laurent Lafforgue is a French mathematician. He has made outstanding contributions to Langlands' program in the fields of number theory and analysis, and in particular proved the Langlands conjectures for the automorphism group of a function field. The crucial contribution by Lafforgue to solve this question is the construction of compactifications of certain moduli stacks of shtukas. The proof was the result of more than six years of concentrated efforts. In 2002 at the 24th International Congress of Mathematicians in Beijing, China, he received the **Fields Medal** together with Vladimir Voevodsky.

He won 2 silver medals at International Mathematical Olympiad (IMO) in 1984 and 1985. He entered the École Normale Supérieure in 1986. In 1994 he received his Ph.D. under the direction of Gérard Laumon in the Arithmetic and Algebraic Geometry team at the Université de Paris-Sud. Currently he is a research director of CNRS, detached as permanent professor of mathematics at the Institut des Hautes Études Scientifiques (IHÉS) in Bures-sur-Yvette, France. He received the Clay Research Award in 2000, and the Grand Prix Jacques Herbrand [fr] of the French Academy of Sciences in 2001. His younger brother Vincent Lafforgue is also a notable mathematician. On 22 May 2011 Lafforgue was awarded an honorary Doctor of Science from the University of Notre Dame *(Source Wikipedia)*. On September 2021, he joined **Huawei Paris Research Center**.

Inspirator 2: Olivia Caramello

Olivia Caramello is an Italian mathematician. She holds a national Rita Levi-Montalcini associate professorship at the University of Insubria in Como, Italy, and Gelfand Chair at the Institut des Hautes Études Scientifiques, France. She is known for her work in topos theory and for pioneering the technique of **toposes as bridges**. She authored the 2017 book Theories, Sites, Toposes: Relating and studying mathematical theories through topos-theoretic bridges.

Caramello earned her bachelor's degree in mathematics at the University of Turin and her Diploma in Piano at the Conservatorio di Cuneo at the age of 19. In 2009, she obtained her Ph.D. in Mathematics at the University of Cambridge (UK), as a Prince of Wales Student of Trinity College, with a thesis entitled "The duality between Grothendieck toposes and geometric theories" under the supervision of Peter Johnstone. In 2016, she obtained her Habilitation at Paris Diderot University with a habilitation thesis entitled "Grothendieck toposes as unifying bridges in Mathematics". Caramello has held a research fellowship at Jesus College, Cambridge and post-doctoral appointments at the De Giorgi Center of the Scuola Normale Superiore di Pisa, Paris Diderot University and the University of Milan (as holder of a Marie Curie Fellowship of the Istituto Nazionale di Alta Matematica) and the Institut des Hautes Etudes Scientifiques *(Source Wikipedia)*.

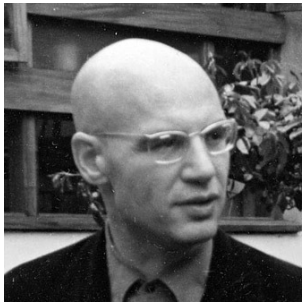
Outline

Nowadays

Historical figures

Historical figures

Claude E. Shannon Information Theory

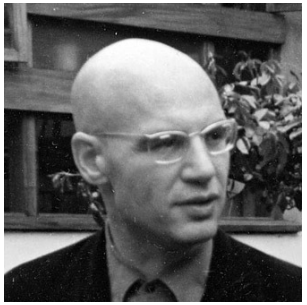


Alexander Grothendieck
Topos Theory Stacks, derivators
Homotopy hypothesis

Figure: Great figures

Historical figures

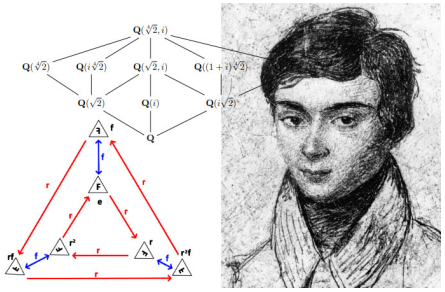
Claude E. Shannon Information Theory



Alexander Grothendieck

Topos Theory Stacks, derivators

Generalization of homotopy



Evariste Galois

A theory of ambiguity

Figure: Great figures

Part II

Why and what is a “Semantic Communication”?

Outline

6 G : building a network of AIs

What do we need to communicate?

Where does AI go ?

- **Energy consumption**

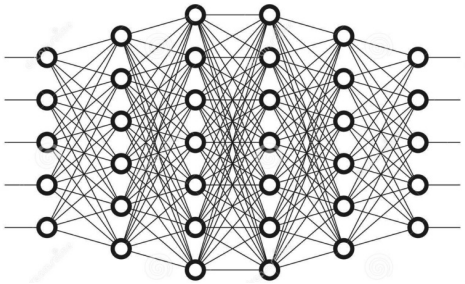


Figure: Huge Neural Networks

Where does AI go ?

- **Energy consumption**

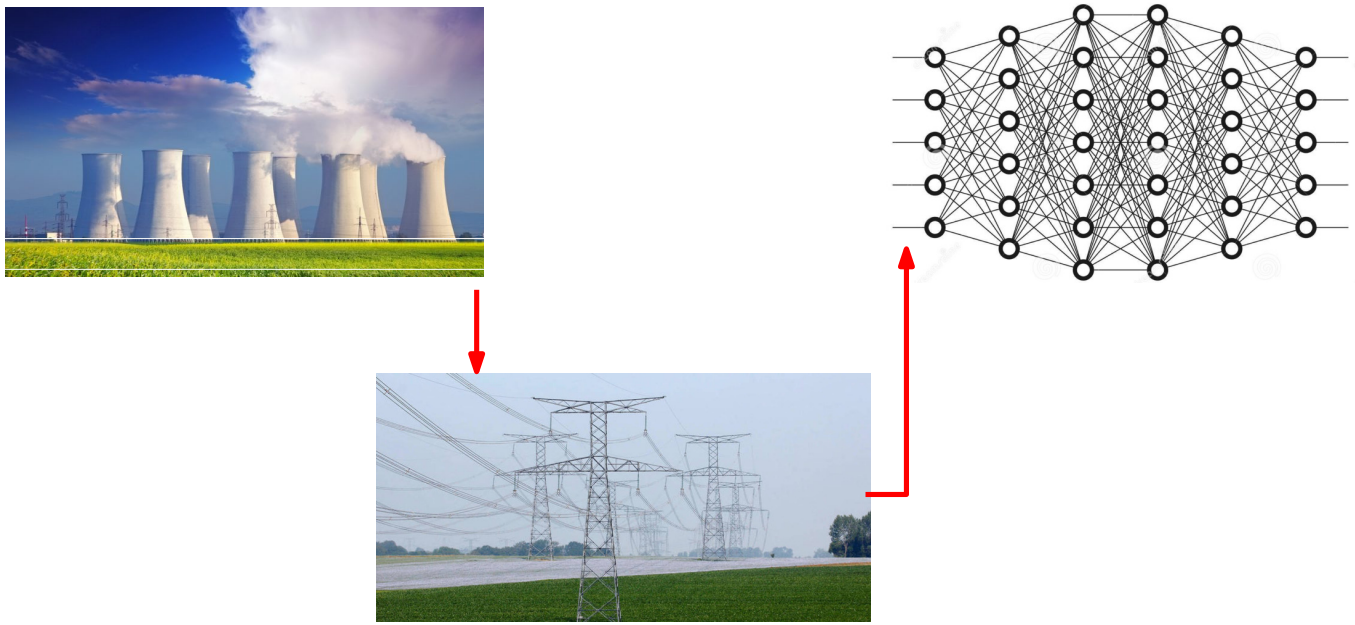


Figure: Huge Neural Networks are Energy consuming

Where does AI go ?

- **Training time of huge networks**

Figure: Training is tooooooo long

6G can be an enabler

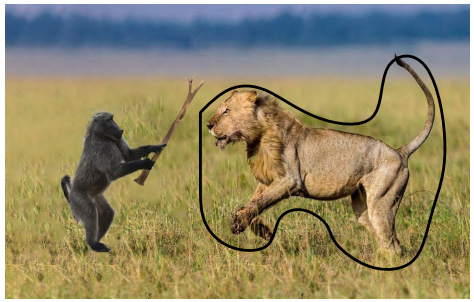
- Semantic communication between agents
  - Instead of having a huge centralized network, choose decentralization
  - Many modules specialized in small tasks (Modular Neural Networks)
  - Distributed learning
  - 6G should be the wide scale enabler for it.
- Notion of **Semantic Information**semantic information : efficient encoding of semantic content.
- **Language** interpretation is of paramount importance.

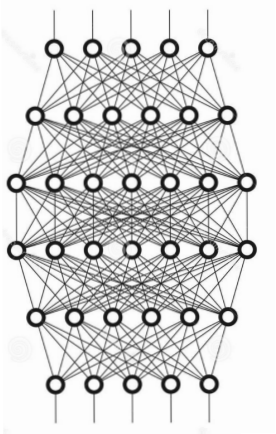
Outline

6 G : building a network of AIs

What do we need to communicate?

The famous "Goal-oriented" communication paradigm





$x \in \{\text{Lions}\}$? $x \notin \{\text{Lions}\}$

- **Traditional communication :** transmit the whole image (the receiver reconstructs it according to some distortion criterion).
- **Naive approach :** here, just transmit 1 bit which encodes the presence or absence of a lion in the image (the receiver just wants to know it; it is not interested in the baboon).
- *NOT ENOUGH* at all !!!!
- the receiver receives one bit related to the type Lion. But it does not know (a priori) its meaning, its *semantics*.
- This semantics as well, needs to be communicated $\Longrightarrow$ *Semantic Communication*.

Part III

Categories

Outline

Categories

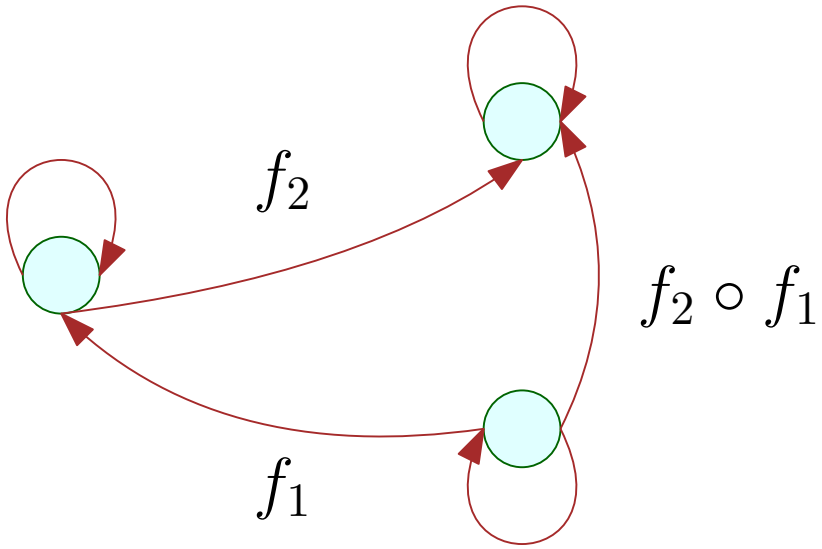
Functors

Sheaves and topos

Group equivariance

A (very) short primer in category theory

- A category is a collection of “objects” and “arrows” satisfying some rules.



- The composition of morphisms (or arrows) is associative.
- There is an identity morphism.

- **Example 1:** The category of Sets
  - Objects are sets
  - Arrows are functions
- **Example 2:** The category Top of topological spaces
  - Objects are topological spaces
  - Arrows are continuous functions
- **Example 3:** Any oriented graph is a small category
  - Objects are vertices
  - Arrows are edges

Outline

Categories

Functors

Sheaves and topos

Group equivariance

Functors

- "Every sufficiently good analogy is yearning to become a functor." -- **John Baez**

Functors

- "Every sufficiently good analogy is yearning to become a functor." -- **John Baez**

- **Functor:** A functor F is a map between two categories C and D satisfying the constraints,
  - $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ for any object $X \in C$.
  - $F(g \circ f) = F(g) \circ F(f)$

- A way to "approximate" objects in D relative to the image by F of objects in C.
- Structure is transported.

Outline

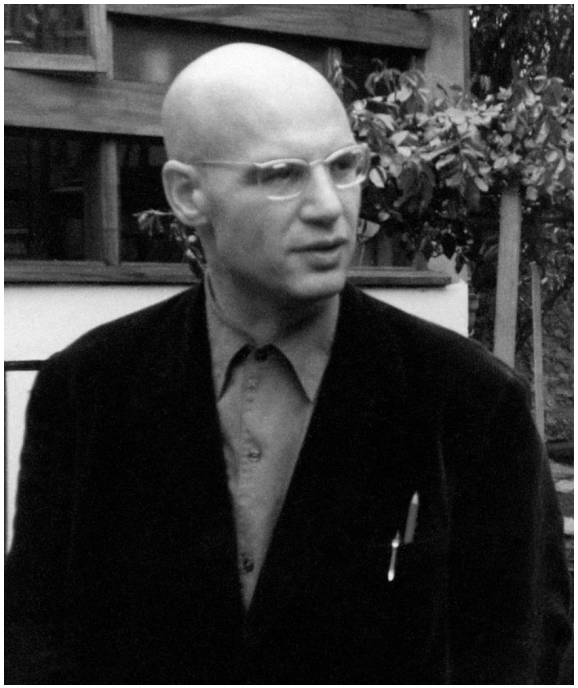
Categories

Functors

Sheaves and topos

Group equivariance

Then came Grothendieck

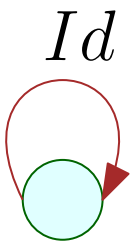


*““It is the **topos** theme which is this “bed” or “deep river” where come to be married geometry and algebra, topology and arithmetic, mathematical logic and category theory, the world of the “continuous” and that of “discontinuous” or discrete structures. It is what I have conceived of most broad to perceive with finesse, by the same language rich of geometric resonances, an “essence” which is common to situations most distant from each other coming from one region or another of the vast universe of mathematical things”.*

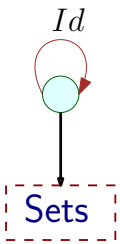
Alexander Grothendieck

The topos of sets

- The small category $\mathscr{C}$ is the simplest one



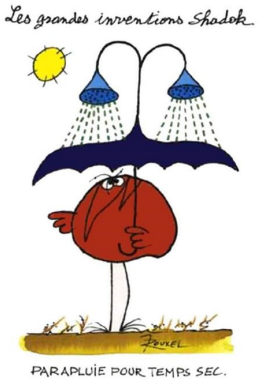
- Sheaves of sets on $\mathscr{C}$



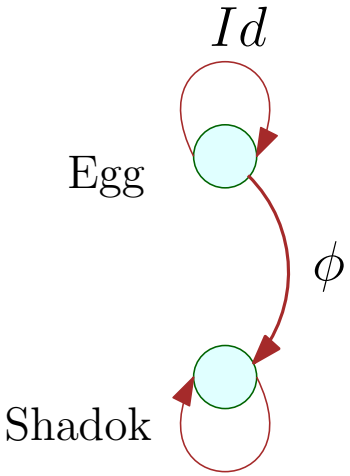
$\sim$ Category of Sets

- Other toposes have the same fundamental properties as the category of sets.
- Computation is possible.

The topos of shadoks

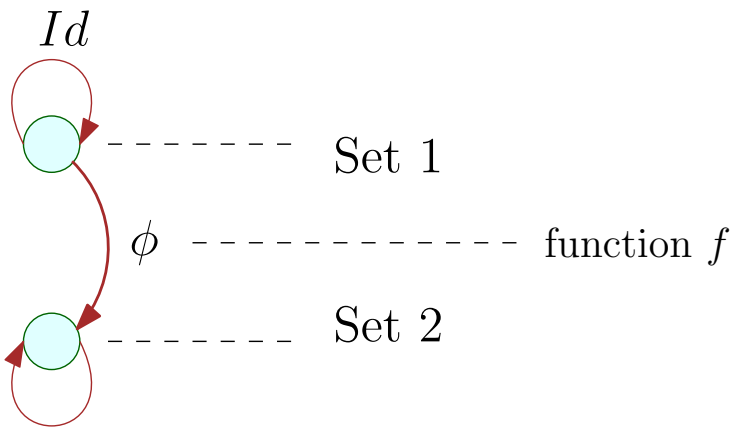


- Introduced by Alain Prouté to explain topos
- The small category is



- Arrow means: belongs to.

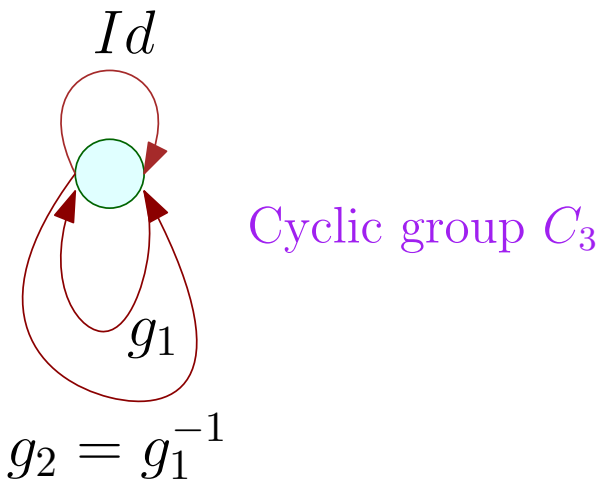
- Sheaves of sets are



The topos on a group: Action of a group on sets

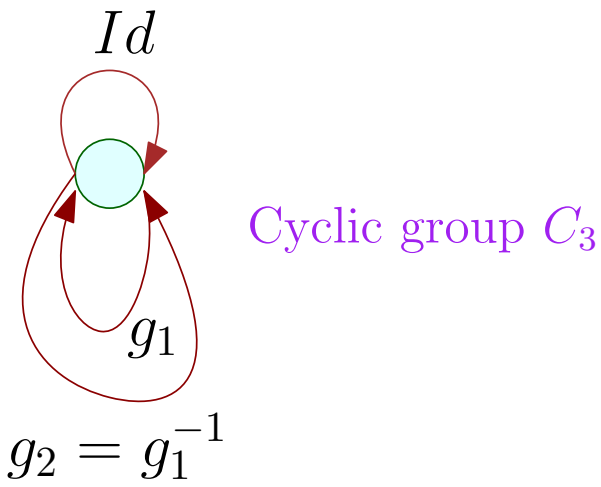
- Consider a given group G. It is a small category (monoid)

- The action of G on a set is a functor $F : G \to \text{Sets}$.



The topos on a group: Action of a group on sets

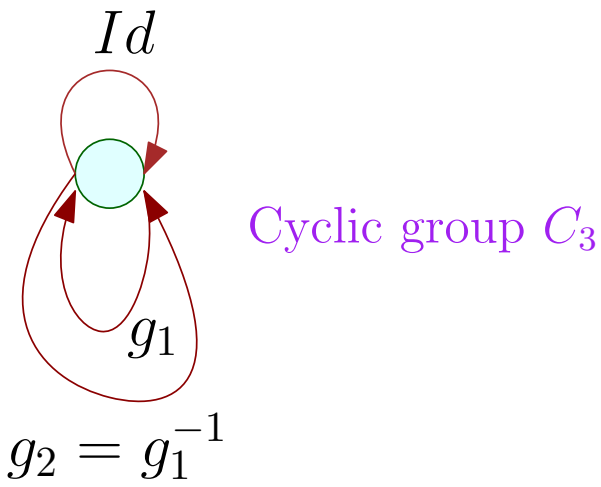
- Consider a given group G. It is a small category (monoid)
- Consider the **Grothendieck pre-sheaves topos** on G, named $\hat{G}$ and fix a context Γ (object of the topos). It is a $G-$Set, a set on which G acts.
- The action of G on a set is a functor $F : G \to$ Sets.



- The object $G-$Set makes concrete the group G.

The topos on a group: Action of a group on sets

- Consider a given group G. It is a small category (monoid)
- Consider the **Grothendieck pre-sheaves topos** on G, named $\hat{G}$ and fix a context Γ (object of the topos). It is a $G-$Set, a set on which G acts.
- The action of G on a set is a functor $F : G \to \text{Sets}$.



- The object $G-$Set makes concrete the group G.
- C_3 is made concrete from its action on sets.

Outline

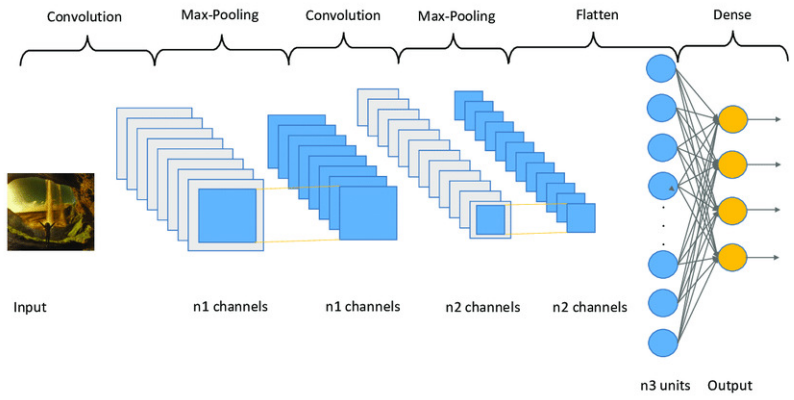
Categories

Functors

Sheaves and topos

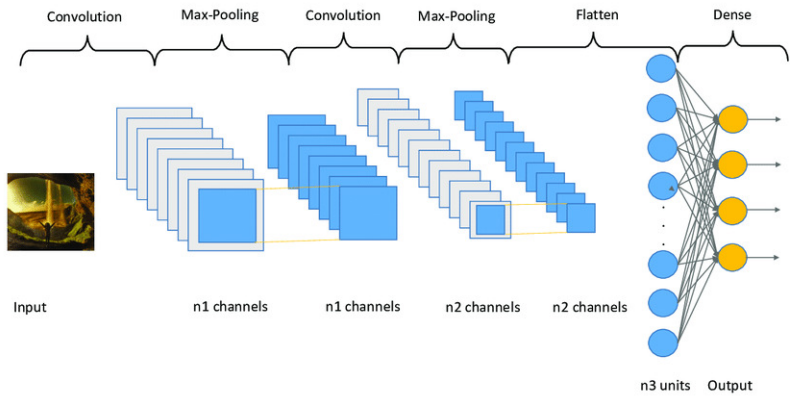
Group equivariance

Convolutional Neural Network



- Images are assumed to be by nature invariant by planar translation.
- It is imposed to a large number of layers to accept a non trivial action of the **group G of $2D-$translations** and to a large numbers of connections between two layers to be compatible with the actions, which implies that the underlying linear part when it exists is made by convolutions with a numerical function on the plane.

Convolutional Neural Network



- Images are assumed to be by nature invariant by planar translation.
- It is imposed to a large number of layers to accept a non trivial action of the **group G of $2D-$translations** and to a large numbers of connections between two layers to be compatible with the actions, which implies that the underlying linear part when it exists is made by convolutions with a numerical function on the plane.

- This doesn't forbid that in several layers, the action of G is trivial, to get invariant characteristics under translations, and here, the layers can be fully connected.
- Other Lie groups and their convolutions have been recently used in DNNs.
- DNNs that analyze images today, for instance in object detection, have several channels of convolutional maps, max pooling and fully connected maps, that are joint together to take a decision.
- It looks as a structure for localizing the translation invariance, as it happens in the successive visual areas in the brains of animals.
- Experiments show that in the first layers, kinds of wavelet kernels are formed spontaneously to translate contrasts, and color opposition kernels are formed to construct color invariance.

Part IV

From Shannon to semantic spaces

Outline

An incorrect formalism: mixing logical information and semantic information

An Outline of a Theory of Semantic Information

Spaces of Semantic Information

- The Carnap - Bar Hillel case
- More general language

What has followed Carnap & Bar-Hillel (1952) since then ?

Towards a Theory of Semantic Communication

Jie Bao
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, NY 03060
baojie@cs.rpi.edu

Prithwish Basu,Mike Dean,Craig Partridge
Raytheon BBN Technologies
Cambridge, MA, 02138
{pbasu,mdean}@bbn.com,craig@aland.bbn.com

Ananthram Swami
US Army Research Lab
Adelphi, MD 20783
ananthram.swami@us.army.mil

Will Leland
Raytheon BBN Technologies
Cambridge, MA, 02138
wel@bbn.com

James A. Hendler
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, NY 03060
hendler@cs.rpi.edu

- In *Bao et al., "Towards a theory of semantic communication,"* the authors use a language and associate to each proposition $\mathscr{P}$ the probability that $\mathscr{P}$ is true.
- Instead of being a starting point for a theory of semantic communication, it is rather a theory of logical communication.
- Here the semantics is a pure probabilistic one which means that it remains Shannonian.

- There is a redundancy given by the language. The propositions can be modelled, using a graph (or better, a simplicial complex).
- Fix some a priori probabilities and run a BP to get the rest.
- The drawback is that the semantics given by artificial neural networks, for instance, is not covered by that approach.
- **Back to Carnap & Bar-Hillel**

Outline

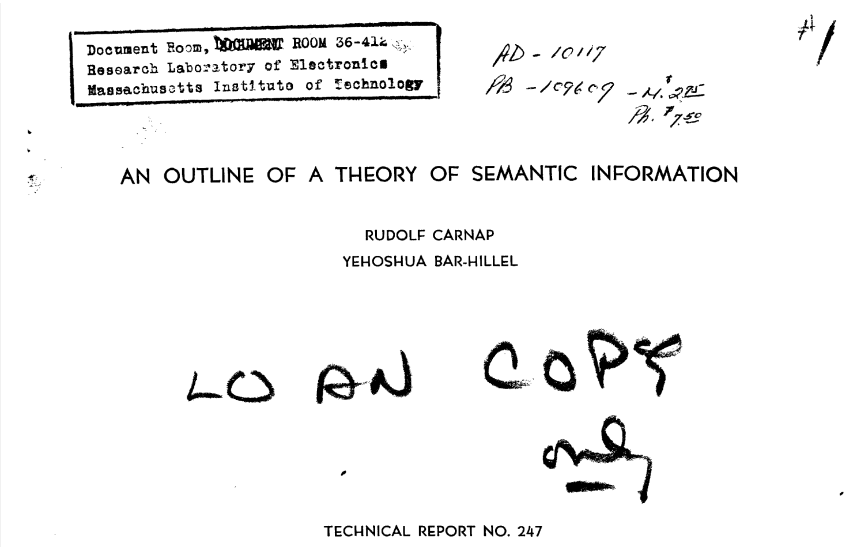
An incorrect formalism: mixing logical information and semantic information

An Outline of a Theory of Semantic Information

Spaces of Semantic Information

- The Carnap - Bar Hillel case
- More general language

Background



Document Room, DOCUMENT ROOM 36-412
Research Laboratory of Electronics
Massachusetts Institute of Technology

AD - 10117
PB - 109609

AN OUTLINE OF A THEORY OF SEMANTIC INFORMATION

RUDOLF CARNAP
YEHOSHUA BAR-HILLEL

LOAN COPY

TECHNICAL REPORT NO. 247

- C&B-H introduced a propositional Boolean language.
- A language $\mathscr{L}_n^{\pi}$ with n subjects which have π different attributes $i = 1, \ldots, \pi$ taking respectively π_i possible values.

- State-descriptions associate to each subject one of the π_i possible values, $\forall i$.
- We can form many more propositions from state-descriptions by considering predicates using
  - Disjunction $\vee$
  - Negation $\neg$
- As the language here is Boolean propositional, it means that we get

$$\chi = \left(\prod_{i=1}^{\pi} \pi_i \right)^n$$

state-descriptions and a total of 2^{χ} possible propositions.

Outline

An incorrect formalism: mixing logical information and semantic information

An Outline of a Theory of Semantic Information

Spaces of Semantic Information

- The Carnap - Bar Hillel case
- More general language

The Language in C&B-H

	M, Y	M, O	F, Y	F, O
1.	a, b, c	-	-	-
2.	-	a, b, c	-	-
3.	-	-	a, b, c	-
4.	-	-	-	a, b, c
5.	a, b	c	-	-
6.	a, c	b	-	-
7.	b, c	a	-	-
8.	a, b	-	c	-
9.	a, c	-	b	-
10.	b, c	-	a	-
11.	a, b	-	-	c
12.	a, c	-	-	b
13.	b, c	-	-	a
14.	c	a, b	-	-
15.	b	a, c	-	-
16.	a	b, c	-	-
17.	-	a, b	c	-
18.	-	a, c	b	-
19.	-	b, c	a	-
20.	-	a, b	-	c
21.	-	a, c	-	b
22.	-	b, c	-	a
23.	c	-	a, b	-
24.	b	-	a, c	-
25.	a	-	b, c	-
26.	-	c	a, b	-
27.	-	b	a, c	-
28.	-	a	b, c	-

a, b, c are subjects

M, F for Male, Female

Y, O for Young, Old

- **State-descriptions:**
  - $(a \in YM) \wedge (b \in YF) \wedge (c \in OM)$
  - $\ldots$
- **Combinatorics:**
  - $\bigvee_{i \in I} \mathscr{P}_i$ where $\mathscr{P}_i$ are state-descriptions.
- **Spaces and shapes:**
  - Extension of the language $\mathscr{L}$
  - "All subjects have the same attributes" (considering the action of permutation group as in 1..4).

Galois group of the Language

- There exists a **Galois group** G of the language, generated by the permutation of the n subjects, the permutations of the values of each attribute and the permutations of the attributes that have the same number of possible values.
- The group of subjects permutation is $\mathfrak{S}_3$, the transposition of values are $\sigma_A = (A_1 A_2)$ and $\sigma_G = (G_1 G_2)$, and the four exchanges of attributes are $\sigma = (A_1 G_1)(A_2 G_2)$, $\kappa = (A_1 G_1 A_2 G_2)$, $\kappa^3 = \kappa^{-1} = (A_1 G_2 A_2 G_1)$, and $\tau = (A_1 G_2)(A_2 G_1)$.
- The group generated by $\sigma, \sigma_A, \sigma_G$ is of order 8; it is the dihedral group D_4 of all the isometries of the square with vertices $A_1 G_1, A_1 G_2, A_2 G_2, A_2 G_1$. The stabilizer of a vertex is a cyclic group C_2, of type σ or τ, the stabilizer of an edge is of type σ_A or σ_G, noted C_2^A or C_2^G.
- Thus the Galois group of the language is $G = \mathfrak{S}_3 \times D_4$.
- Here the language $\mathscr{L}$ is a **sheaf** over the category G, which plays the role of the fiber $\mathscr{F}$ and $\mathscr{C}$ has only one object U_0.

Example [Carnap & Bar-Hillel, 1952]: The orbits

- **Type I:** One orbit of four elements, where a, b, c have the same gender and age. The stabilizer of each element is $\mathfrak{S}_3 \times C_2$, or order 12.
- **Type II:** One orbit of 24 elements made by a pair of equal subjects and one that differs from them by one attribute only. The stabilizer being the $\mathfrak{S}_2$ of the pair of subjects.
- **Type III:** One orbit of 12 elements made by a pair of equal subjects and one that differs from them by the two attributes. The stabilizer being the product $\mathfrak{S}_2 \times C_2$, where C_2 stabilizes the characteristic of the pair, which is the same as stabilizing the character of the different subject.
- **Type IV:** One last orbit of 24 elements, where the three subjects are different, then two of them differ by one attribute and differ from the last one by the two attributes. The stabilizer is the stabilizer C_2 of the missing pair of values of the attributes.
- Each type corresponds to a well formed sentence in natural languages:
  - type *I* is translated by "all the subjects have the same attributes";
  - type *II* by "all the subjects have the same attributes except one which differs by only one aspect";
  - type *III* "one subject is opposite to all the others";
  - type *IV* "all the subjects are distinguished by at least one attribute".

Example [Carnap & Bar-Hillel, 1952]: Spaces of information

- **Groupoid** formed by the elements of each orbit (objects).
- The space of information must have a form attached to G, but it also must take into account the structure of its elements. A natural choice is the connected groupoid with objects the elements $e \in T$ and with isomorphisms their stabilizers.
- A groupoid is related to a space through homotopy (fundamental groupoid) when only non-trivial $1D-$components exist.
- $\infty-$groupoid is the general case (related to the **homotopy hypothesis**).

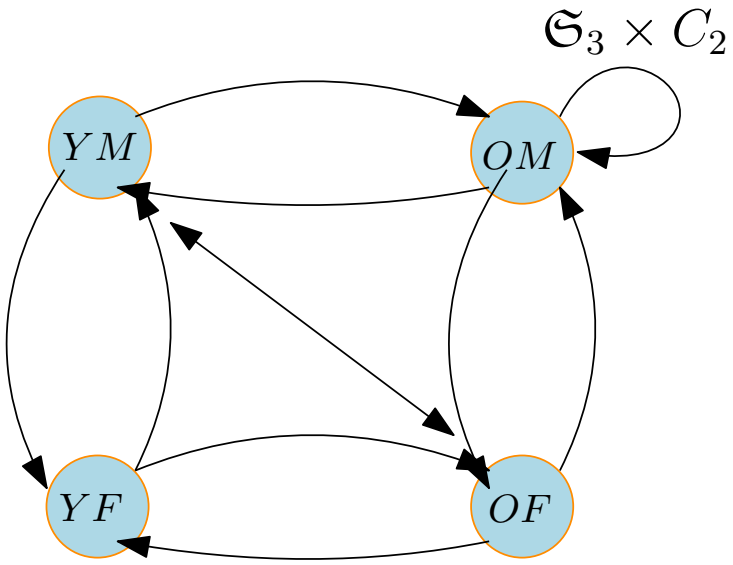
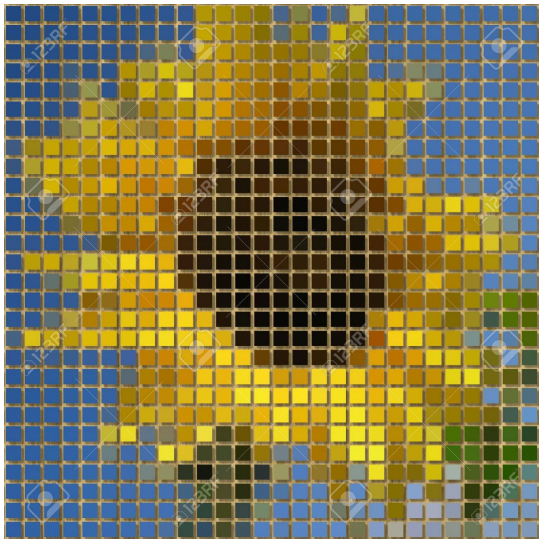


Figure: Groupoid of information for type I

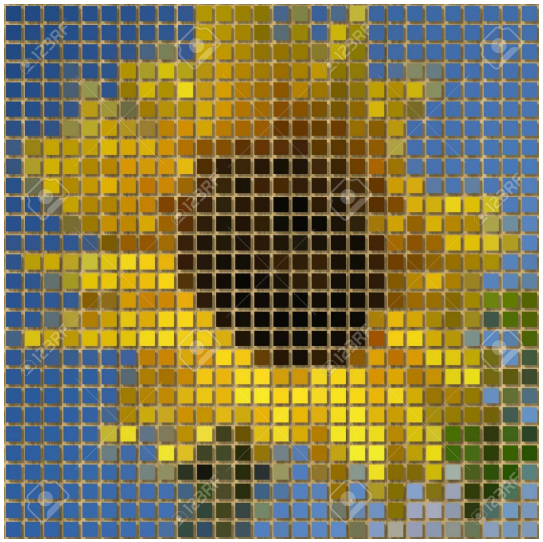
The language of learning of a DNN



- Image : Pixels described using a propositional language. Each of them (p_{ij}) has 3 attributes (RGB) taking values among 256 possible grey levels. A typical atomic proposition is :

$$[(Rp_{11} = 153) \wedge (Gp_{11} = 45) \wedge (Bp_{11} = 202)] \wedge [(Rp_{12} \cdots) \cdots] \wedge$$

The language of learning of a DNN

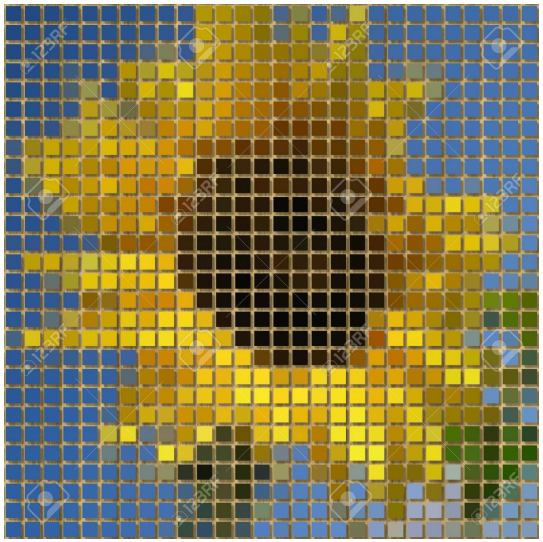


- Image : Pixels described using a propositional language. Each of them (p_{ij}) has 3 attributes (RGB) taking values among 256 possible grey levels. A typical atomic proposition is :

$$[(Rp_{11} = 153) \wedge (Gp_{11} = 45) \wedge (Bp_{11} = 202)] \wedge [(Rp_{12} \cdots) \cdots] \wedge$$

- Already at the input layer of a DNN, a category appears: A rectangle $\subset \mathbb{N}^2$, which is a poset.
- Then a fibration is given over this poset with the structural properties of what we want to detect in the image (groups or even groupoids).

The language of learning of a DNN



- Image : Pixels described using a propositional language. Each of them (p_{ij}) has 3 attributes (RGB) taking values among 256 possible grey levels. A typical atomic proposition is :

$$[(Rp_{11} = 153) \wedge (Gp_{11} = 45) \wedge (Bp_{11} = 202)] \wedge [(Rp_{12} \cdots) \cdots] \wedge$$

- Already at the input layer of a DNN, a category appears: A rectangle $\subset \mathbb{N}^2$, which is a poset.
- Then a fibration is given over this poset with the structural properties of what we want to detect in the image (groups or even groupoids).
- It defines a kind of **Galois theory** where the maximum ambiguity is achieved at the output layer (many images are seen as flower images).

Part V

Semantic Source/Channel Coding

Outline

Sampling vs. Invariants

Semantic Source

Semantic Channel

What is the semantics corresponding to a language?

- Semantics can be given in a purely set-theoretic way : **pure data** or in a more human-like way: **invariant-based**.

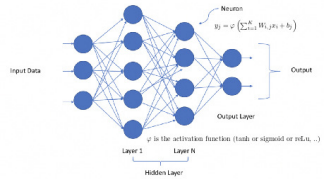


Figure: Pure data-based learning

Figure: Learning through invariants and equivalence

- Structural properties (good for generalization): basically generate a whole orbit from few data (strict minimum ! 🙂) that involve spaces of information.
- Training data.

Outline

Sampling vs. Invariants

Semantic Source

Semantic Channel

A semantic source

- A source (image, video, sound file, ...) and a priori knowledge which are associated to a base category $\mathscr{F}$ taking into account
- Languages that are *"justified"* for this source are objects of the classifying topos of presheaves over $\mathscr{F}$.
- In a more concrete way, it means that possible languages correspond to those which can say something about .
- A space of information H can show the "**shape**" of a given language.
- A semantic source encoder transforms a source into elements of a language.
- "*Minimize*" the language complexity.

Outline

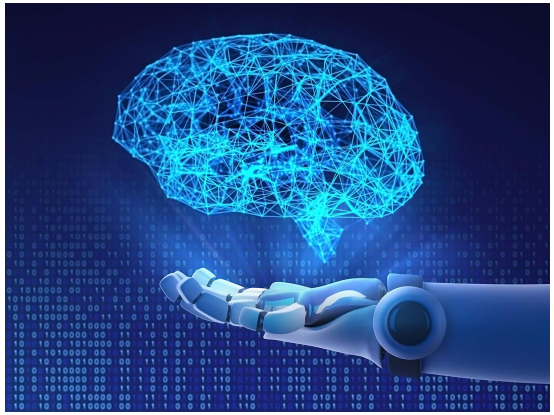
Sampling vs. Invariants

Semantic Source

Semantic Channel

A semantic channel

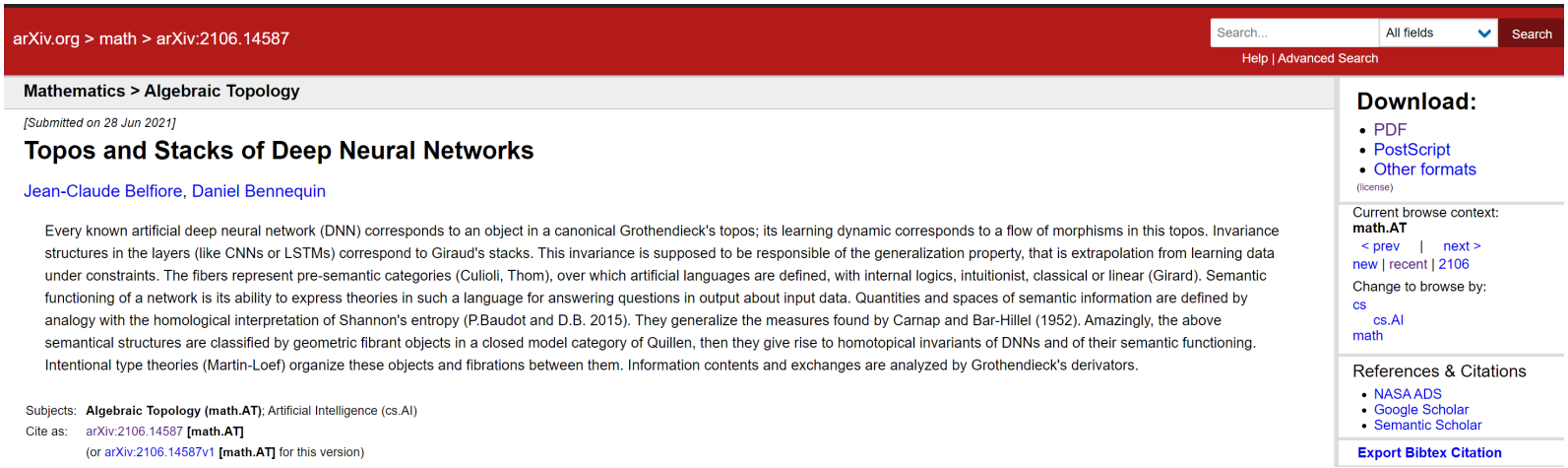
- Building a global brain with many neural modules is equivalent to build a semantic network.
- **Need of feedback:** For instance, Y asks X to help it to solve a problem, or to complete its own knowledge, then X sends an information to Y, and vice versa, and the roles of X and Y can be reversed in some other situations (R. Thom on Information).
- A space of information corresponding to $I(X;Y)$ measures what is missing to Y to understand X and vice versa.



- It means that some kind of semantic channel coding would consist in sending, apart from elements of a language, the semantic content which is missing to the receiver.
- In some sense, the less the receiver understand the transmitter, the more semantic data have to be sent.

Conclusion

- Starting a mathematical theory of semantic communication based on toposes, stacks and abstract homotopy.
- It aims at capturing structural properties.
- J.-C. Belfiore & D. Bennequin, “Topos and Stacks of Deep Neural Networks,” `https://arxiv.org/abs/2106.14587` .



arXiv.org > math > arXiv:2106.14587

Search... All fields Search

Help | Advanced Search

Mathematics > Algebraic Topology

[Submitted on 28 Jun 2021]

Topos and Stacks of Deep Neural Networks

Jean-Claude Belfiore, Daniel Bennequin

Every known artificial deep neural network (DNN) corresponds to an object in a canonical Grothendieck's topos; its learning dynamic corresponds to a flow of morphisms in this topos. Invariance structures in the layers (like CNNs or LSTMs) correspond to Giraud's stacks. This invariance is supposed to be responsible of the generalization property, that is extrapolation from learning data under constraints. The fibers represent pre-semantic categories (Culioli, Thom), over which artificial languages are defined, with internal logics, intuitionist, classical or linear (Girard). Semantic functioning of a network is its ability to express theories in such a language for answering questions in output about input data. Quantities and spaces of semantic information are defined by analogy with the homological interpretation of Shannon's entropy (P.Baudot and D.B. 2015). They generalize the measures found by Carnap and Bar-Hillel (1952). Amazingly, the above semantical structures are classified by geometric fibrant objects in a closed model category of Quillen, then they give rise to homotopical invariants of DNNs and of their semantic functioning. Intentional type theories (Martin-Loef) organize these objects and fibrations between them. Information contents and exchanges are analyzed by Grothendieck's derivators.

Subjects: **Algebraic Topology (math.AT)**; Artificial Intelligence (cs.AI)
Cite as: arXiv:2106.14587 **[math.AT]**
(or arXiv:2106.14587v1 **[math.AT]** for this version)

Download:
- PDF
- PostScript
- Other formats

(license)

Current browse context:
math.AT
< prev | next >
new | recent | 2106

Change to browse by:
cs
cs.AI
math

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

Export Bibtex Citation

- New paper in preparation : “*Searching Semantic Spaces*”