Categorical Represtaion Learning

Presenter : Amin Hatami
Winter 2023

Category Theory

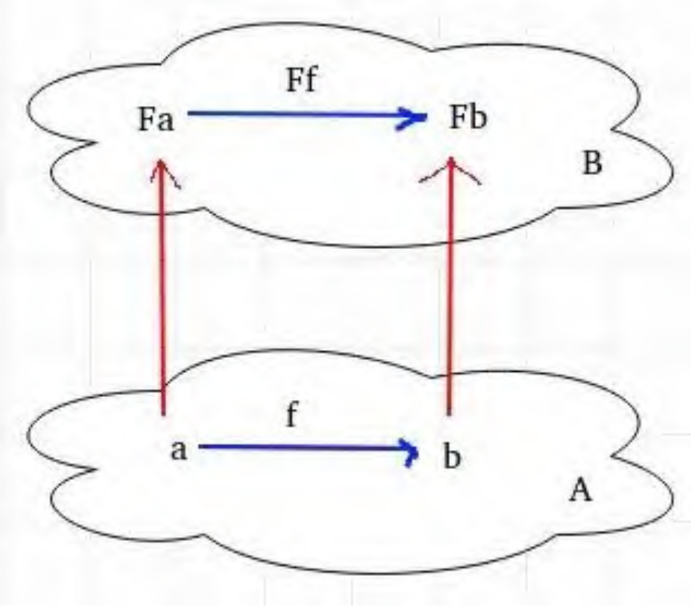
- a toolset for describing the general abstract structures and their relations
- was a great revolution of mathematics in the 20th century
- takes a bird's eye view of mathematics and provides a sweeping vista of the terrain
  - details become invisible
  - It Can spot patterns that were impossible to detect from ground level

Mathematical Landscape

Credit:Martin Kuppe

Definitions

- **Category**
  - a Collection of **objects**
- ***morphism***
  - A Structure Preserving **map** between two **objects** in an abstract category
  - For algebraic structures, they are usually the **Homomorphisms**
- **Functor**
  - a mapping between **categories**
  - considered as morphisms in categories of categories



Category Theory Perspective

- Relationships are everything
  - Objects can only defined through their **interrelations**
  - But Graph Neural Networks did not place relationships in the first place

- Categorical Representation Learning
  - Directly **learns** the representation of relations as **feature matrices**

Learning Steps

1. **Mine the categorical structure from data**

   enable the machine to extract the representation of objects and morphisms from data.

2. **Align the categorical structures between datasets**

   establish a functor between categories based on the learned categorical representations

3. **Discover hierarchical structures with tensor categories**

   combine the categorical representation learning and functorial learning

An Example

*For a **language dataset**, each object is given as a word;*

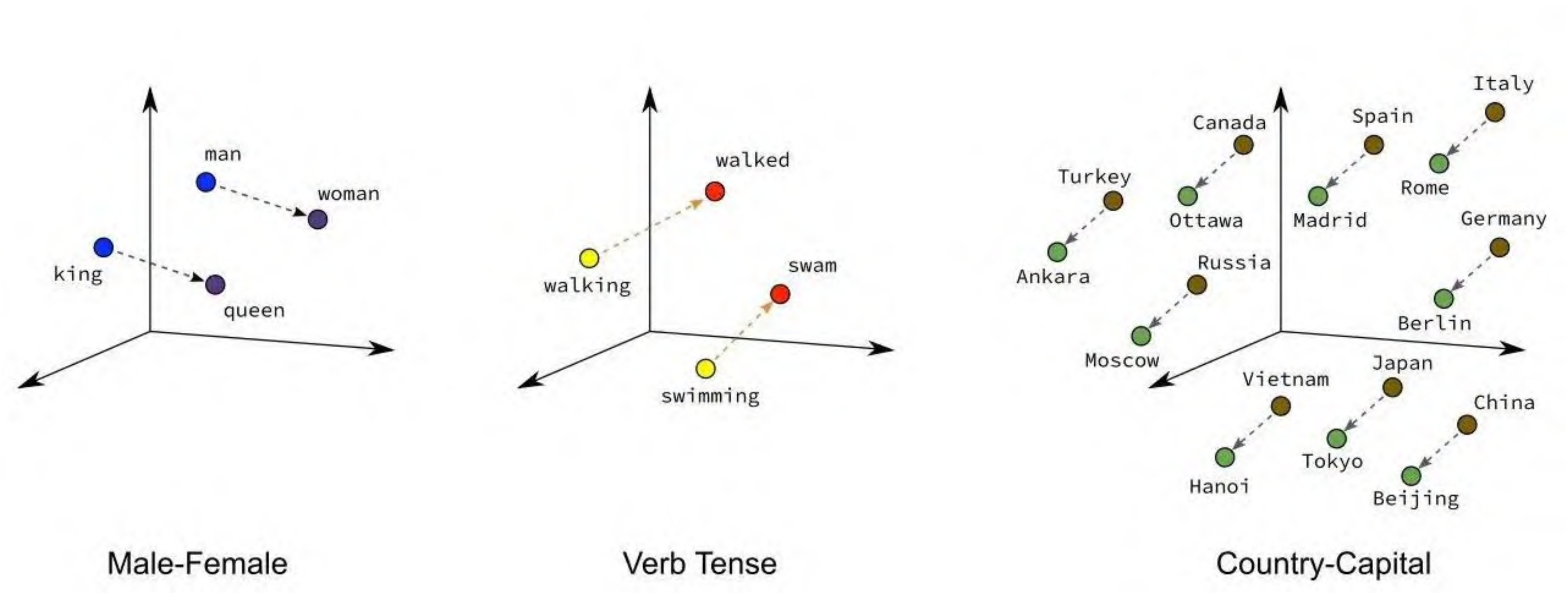
*represented as a word vector, via the **word-vector mapping functor***

Each morphism is a relation between two words, represented by a matrix

For instance:

$$\mathtt{bright} \xrightarrow{\mathtt{antonym}} \mathtt{dark} : v_{\mathtt{dark}} = M_{\mathtt{antonym}} v_{\mathtt{bright}}$$

Word Embeddings Example



Source : Crash Course by Google Developers/ CC BY 4.0

Functorial Learning

- $\mathcal{F}$: functor
- $\mathcal{V}_{\mathcal{F}}$: transformation
- a : Object
- v_a : vector embedding of each object
- $v_{\mathcal{F}(a)}$: vector embedding of the corresponding object in the target category

$$v_{\mathcal{F}(a)} = V_{\mathcal{F}} v_a$$

Matrix Embedding

- M_f : Matrix embedding of each morphism
- $M_{\mathcal{F}(f)}$: Matrix embedding of the corresponding Morphism in the target category

$$M_{\mathcal{F}(f)} V_{\mathcal{F}} = V_{\mathcal{F}} M_f$$

$$\begin{tikzcd} \mathcal{C} \arrow[d, "\mathcal{F}"] \\ \mathcal{D} \end{tikzcd}$$

$$\begin{tikzcd} v_a \arrow[r, "M_f"] \arrow[d, "V_{\mathcal{F}}"] & v_b \arrow[d, "V_{\mathcal{F}}"] \\ V_{\mathcal{F}} v_a \arrow[r, "M_{\mathcal{F}(f)}"] & V_{\mathcal{F}} v_b \end{tikzcd}$$

Learning Embedding from Statistics.

- the object and morphism embeddings are learned from the **concurrence statistics**

- the probability p(a, b) that a pair of objects (a,b) **occurs** together in the **same** composite object.

Pointwise Mutual Information

$$PMI(a,b) = \log(\frac{P(a,b)}{P(\mathrm{a})\mathrm{P}(\mathrm{b})})$$

- If the objects were independent:
  - p(a, b) = p(a)p(b) = 0 → PMI(a,b)=0

Singular Value Decomposition

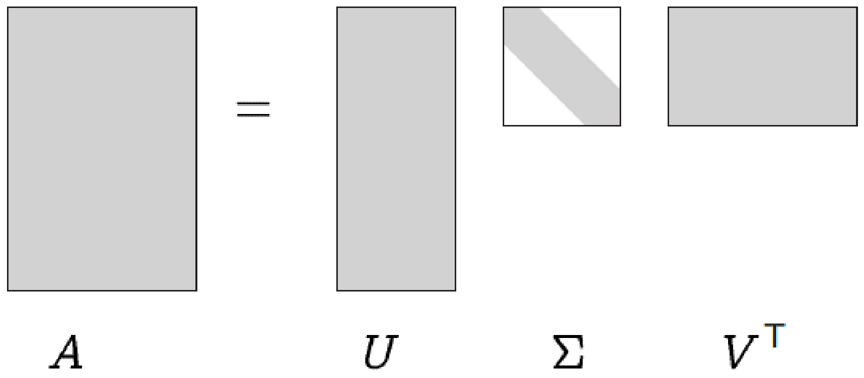
U:rows corresponding to original but m columns represents a dimension in a new latent space, such that

- r column vectors are orthogonal to each other
- Columns are ordered by the amount of variance in the dataset each new dimension accounts for

Σ: diagonal *r* x *r* matrix of **singular values** expressing the importance of each dimension.

$\boldsymbol{V^T}$: columns corresponding to original but r rows corresponding to singular values

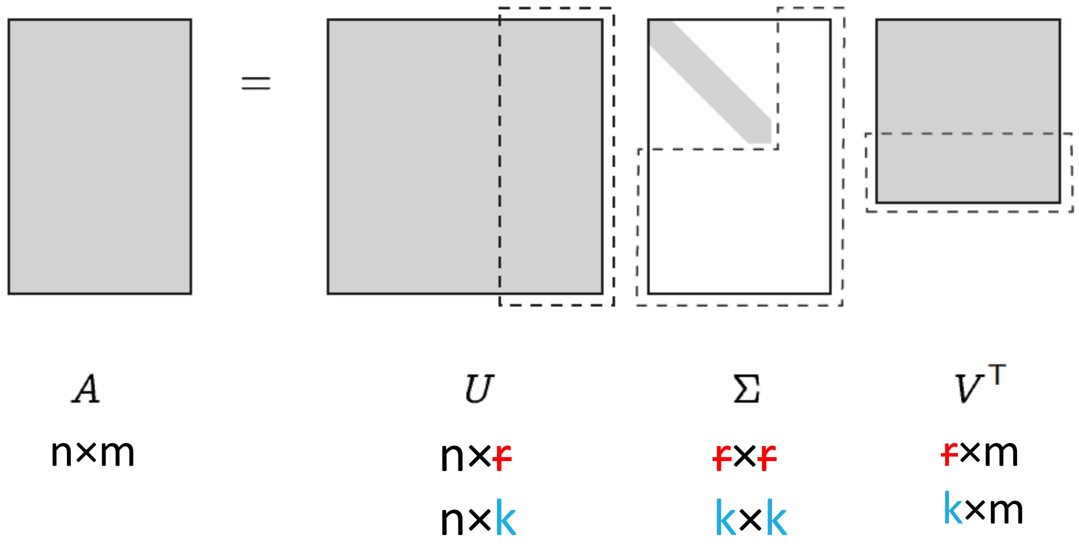
$$A = U\Sigma V^{\mathsf{T}} = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathsf{T}}$$



SVD applied to term-document matrix

- If instead of keeping all r dimensions, we just keep the top k singular values.
- The result is a **least-squares approximation** to the original matrix A
- But instead of multiplying, we'll just make use of U.
- Each row of U:
  - k-dimensional vector Representing word W

✓ We Can apply SVD to PMI word-word matrices



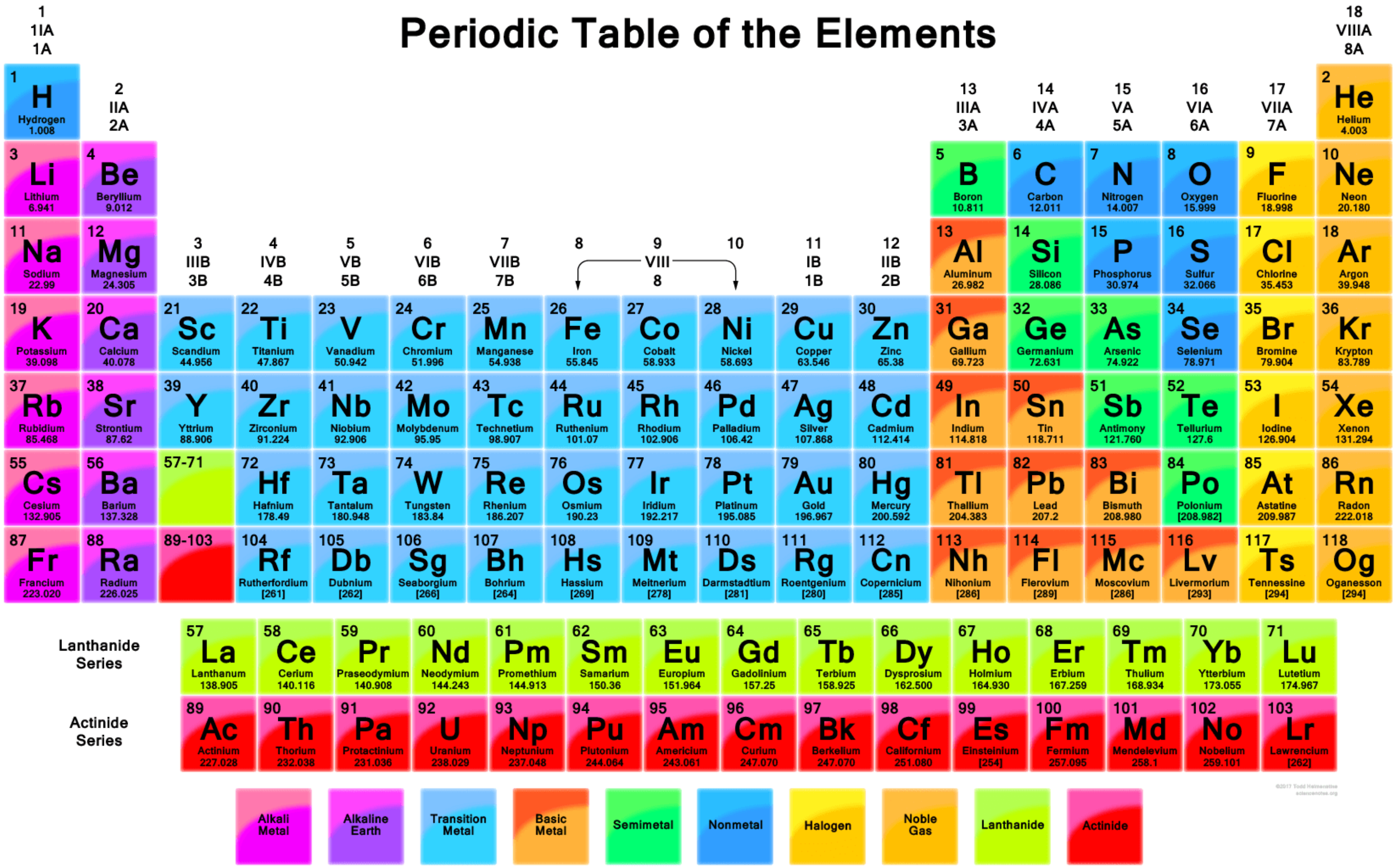
Preliminary Results

Learning Chemical Compounds

- Data Set
  - Contains 61023 inorganic compounds
  - Covers 89 elements in periodic Table

- Modeling as a category containing
  - Elements as fundamental objects
  - Compounds as composite objects

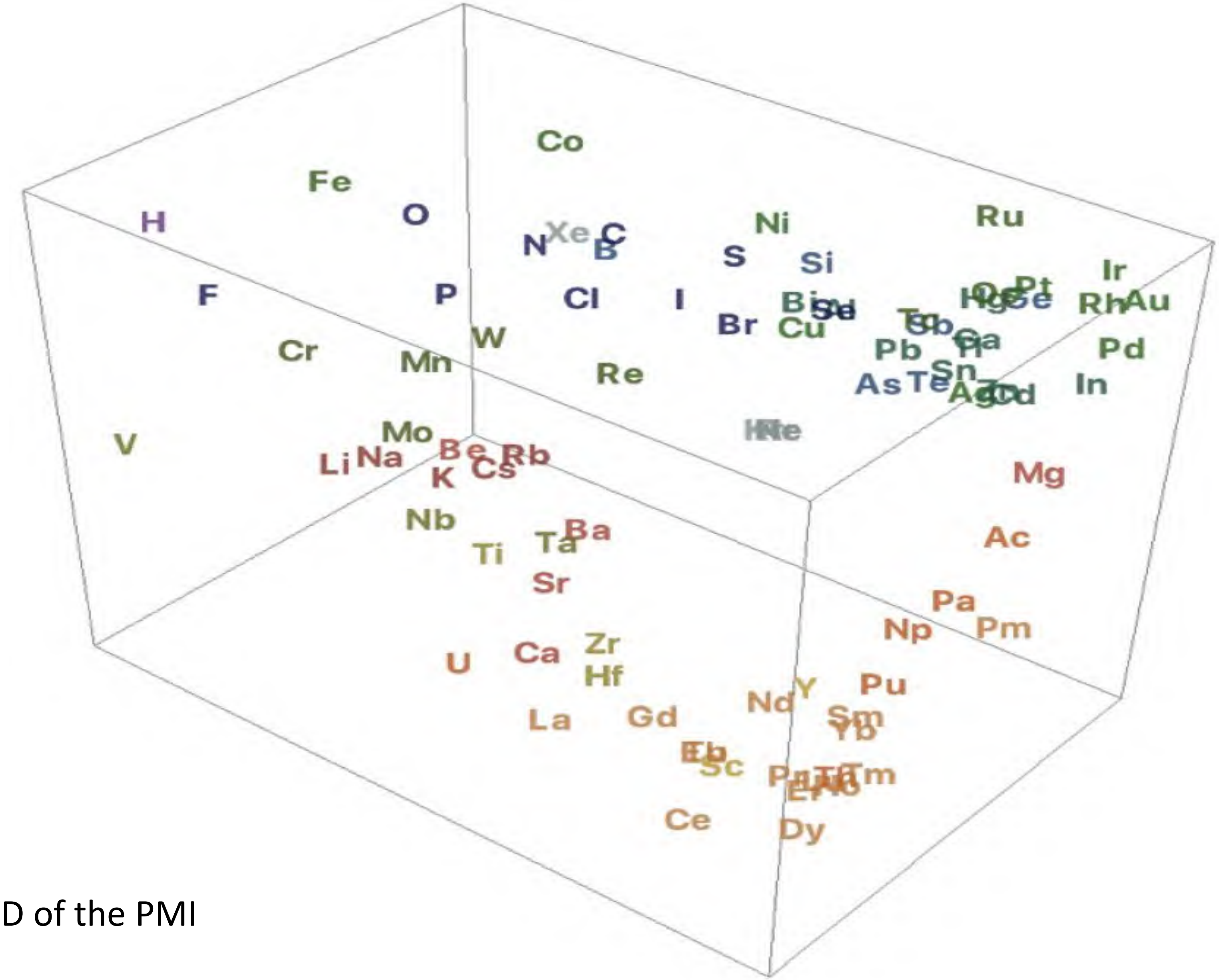
- concurrence of two elements in a compound
  - Is due to the underlying relations → Morphisms can emerge from that



Periodic Table of the Elements
1 1IA 1A
2 IIA 2A
3 IIIB 3B
4 IVB 4B
5 VB 5B
6 VIB 6B
7 VIIB 7B
8
9 VIII 8
10
11 IB 1B
12 IIB 2B
13 IIIA 3A
14 IVA 4A
15 VA 5A
16 VIA 6A
17 VIIA 7A
18 VIIIA 8A
1 H Hydrogen 1.008
2 He Helium 4.003
3 Li Lithium 6.941
4 Be Beryllium 9.012
5 B Boron 10.811
6 C Carbon 12.011
7 N Nitrogen 14.007
8 O Oxygen 15.999
9 F Fluorine 18.998
10 Ne Neon 20.180
11 Na Sodium 22.99
12 Mg Magnesium 24.305
13 Al Aluminum 26.982
14 Si Silicon 28.086
15 P Phosphorus 30.974
16 S Sulfur 32.066
17 Cl Chlorine 35.453
18 Ar Argon 39.948
19 K Potassium 39.098
20 Ca Calcium 40.078
21 Sc Scandium 44.956
22 Ti Titanium 47.867
23 V Vanadium 50.942
24 Cr Chromium 51.996
25 Mn Manganese 54.938
26 Fe Iron 55.845
27 Co Cobalt 58.933
28 Ni Nickel 58.693
29 Cu Copper 63.546
30 Zn Zinc 65.38
31 Ga Gallium 69.723
32 Ge Germanium 72.631
33 As Arsenic 74.922
34 Se Selenium 78.971
35 Br Bromine 79.904
36 Kr Krypton 83.789
37 Rb Rubidium 85.468
38 Sr Strontium 87.62
39 Y Yttrium 88.906
40 Zr Zirconium 91.224
41 Nb Niobium 92.906
42 Mo Molybdenum 95.95
43 Tc Technetium 98.907
44 Ru Ruthenium 101.07
45 Rh Rhodium 102.906
46 Pd Palladium 106.42
47 Ag Silver 107.868
48 Cd Cadmium 112.414
49 In Indium 114.818
50 Sn Tin 118.711
51 Sb Antimony 121.760
52 Te Tellurium 127.6
53 I Iodine 126.904
54 Xe Xenon 131.294
55 Cs Cesium 132.905
56 Ba Barium 137.328
57-71
72 Hf Hafnium 178.49
73 Ta Tantalum 180.948
74 W Tungsten 183.84
75 Re Rhenium 186.207
76 Os Osmium 190.23
77 Ir Iridium 192.217
78 Pt Platinum 195.085
79 Au Gold 196.967
80 Hg Mercury 200.592
81 Tl Thallium 204.383
82 Pb Lead 207.2
83 Bi Bismuth 208.980
84 Po Polonium [208.982]
85 At Astatine 209.987
86 Rn Radon 222.018
87 Fr Francium 223.020
88 Ra Radium 226.025
89-103
104 Rf Rutherfordium [261]
105 Db Dubnium [262]
106 Sg Seaborgium [266]
107 Bh Bohrium [264]
108 Hs Hassium [269]
109 Mt Meitnerium [278]
110 Ds Darmstadtium [281]
111 Rg Roentgenium [280]
112 Cn Copernicium [285]
113 Nh Nihonium [286]
114 Fl Flerovium [289]
115 Mc Moscovium [286]
116 Lv Livermorium [293]
117 Ts Tennessine [294]
118 Og Oganesson [294]
Lanthanide Series
57 La Lanthanum 138.905
58 Ce Cerium 140.116
59 Pr Praseodymium 140.908
60 Nd Neodymium 144.243
61 Pm Promethium 144.913
62 Sm Samarium 150.36
63 Eu Europium 151.964
64 Gd Gadolinium 157.25
65 Tb Terbium 158.925
66 Dy Dysprosium 162.500
67 Ho Holmium 164.930
68 Er Erbium 167.259
69 Tm Thulium 168.934
70 Yb Ytterbium 173.055
71 Lu Lutetium 174.967
Actinide Series
89 Ac Actinium 227.028
90 Th Thorium 232.038
91 Pa Protactinium 231.036
92 U Uranium 238.029
93 Np Neptunium 237.048
94 Pu Plutonium 244.064
95 Am Americium 243.061
96 Cm Curium 247.070
97 Bk Berkelium 247.070
98 Cf Californium 251.080
99 Es Einsteinium [254]
100 Fm Fermium 257.095
101 Md Mendelevium 258.1
102 No Nobelium 259.101
103 Lr Lawrencium [262]
Alkali Metal
Alkaline Earth
Transition Metal
Basic Metal
Semimetal
Nonmetal
Halogen
Noble Gas
Lanthanide
Actinide

we can obtain three-dimensional vector encodings of elements

We observe that elements of similar chemical properties are close to each other, because they share similar context in the compound



Embeddings of elements by SVD of the PMI

Semisupervised Translation

- We take the data set in English and translate each element into Chinese

- The task of semi-supervised translation is to learn to translate chemical compounds from one language to another with only a few aligned samples.

- The unsupervised translation is possible since the chemical relation between elements are identical in both languages

Er2SO2	钐钒氧4
La2SiO5	镨(铜锗)2
FeClO	铈(锗钌)2
Pr2SO2	铈(钴锗)2
GdNbO4	锂镥氟4
ErNbO4	钾3铝氢6
FeSiRu2	镱2(锌锗)3
Er(NiGe)2	钐2镉铱
LiMg2Pd	铁3铑氮
Pr(MnGe)2	钡2磷氯
...	...

Results of the semisupervised translation

For each English element, the top three Chinese translations are listed.

* The gray lines are selected supervised elements

+ The row is green if the correct translation is the top candidate.

? The row is yellow if the correct translation is not the top candidate but appears within top three.

- The row is red if the correct translation does not appear even within the top-three candidates

*	K	=	钾:	0.95,	氙:	0.72,	钠:	0.69
*	Kr	=	氪:	0.90,	氟:	0.82,	锰:	0.80
*	La	=	镧:	0.94,	镝:	0.83,	铒:	0.78
+	Li	=	锂:	0.86,	钛:	0.70,	钠:	0.69
?	Lu	=	铥:	0.76,	镥:	0.75,	钪:	0.74
+	Mg	=	镁:	0.78,	钷:	0.74,	锕:	0.71
+	Mn	=	锰:	0.86,	锂:	0.75,	铁:	0.71
+	Mo	=	钼:	0.89,	氩:	0.84,	铵:	0.79
-	N	=	钼:	0.69,	钴:	0.67,	氩:	0.66
-	Na	=	氩:	0.89,	铵:	0.83,	铷:	0.81
?	Nb	=	氩:	0.79,	钆:	0.79,	铌:	0.78
-	Nd	=	镨:	0.82,	钐:	0.81,	铒:	0.79

-	Ni	=	钴:	0.73,	镤:	0.68,	钌:	0.68
+	Np	=	镎:	0.82,	钬:	0.72,	镧:	0.72
+	O	=	氧:	0.88,	氦:	0.81,	磷:	0.78
-	Os	=	钌:	0.65,	镤:	0.63,	铝:	0.57
+	P	=	磷:	0.86,	氦:	0.77,	氧:	0.76
+	Pa	=	镤:	0.85,	铥:	0.80,	镥:	0.79
+	Pb	=	铅:	0.64,	镉:	0.63,	碘:	0.63
+	Pd	=	钯:	0.71,	钌:	0.66,	铱:	0.66
+	Pm	=	钷:	0.82,	锕:	0.76,	镱:	0.75
+	Pr	=	镨:	0.84,	镝:	0.81,	铒:	0.80
+	Pt	=	铂:	0.59,	钌:	0.57,	锇:	0.53
+	Rb	=	铷:	0.73,	钬:	0.65,	氢:	0.64

References

- Artan Sheshmani, Yizhuang You, Categorical Representation Learning: Morphism is All You Need, arXiv
- Chris Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999
- Jacobson, Nathan , Basic algebra, vol. 2 (2nd ed.), Dover, 2009

Thank you for your Attention!