Do all singers sing? Determining mandatory attributes in knowledge bases

Jonathan Lajus[1,2,3]
[1]LTCI, [2]Télécom ParisTech, [3]Université Paris-Saclay

Goal

Given a class and an attribute in a knowledge base, determine whether all instances of the class have the attribute in the real world, e.g. every Singer should sing:

$$Singer(a) \Rightarrow \exists b.sings(a, b)$$

Mine such statements from the data.

Difficulty: Confidence of the rule: 20%

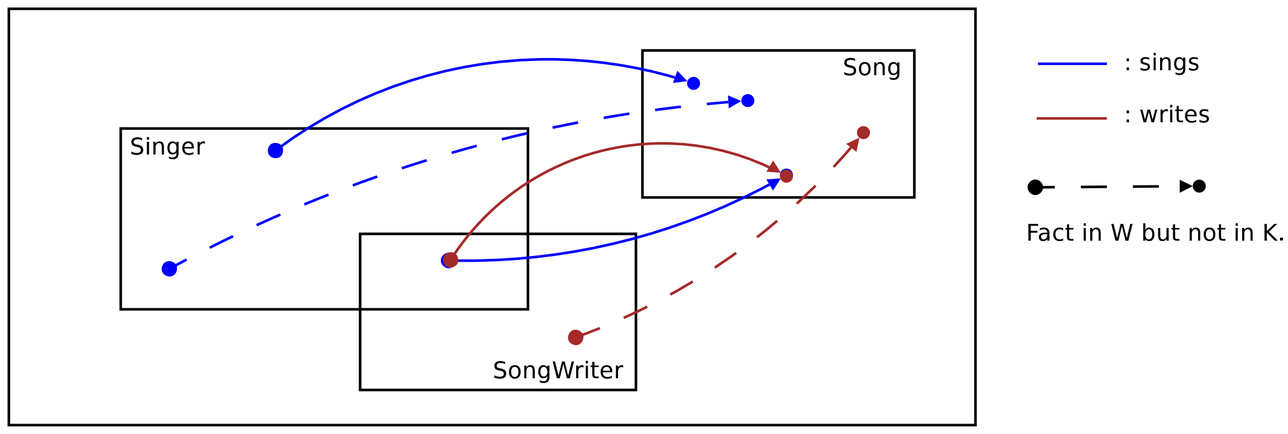
Motivation: Identify where the Knowledge Base (KB) is incomplete.

Assumptions on a KB K

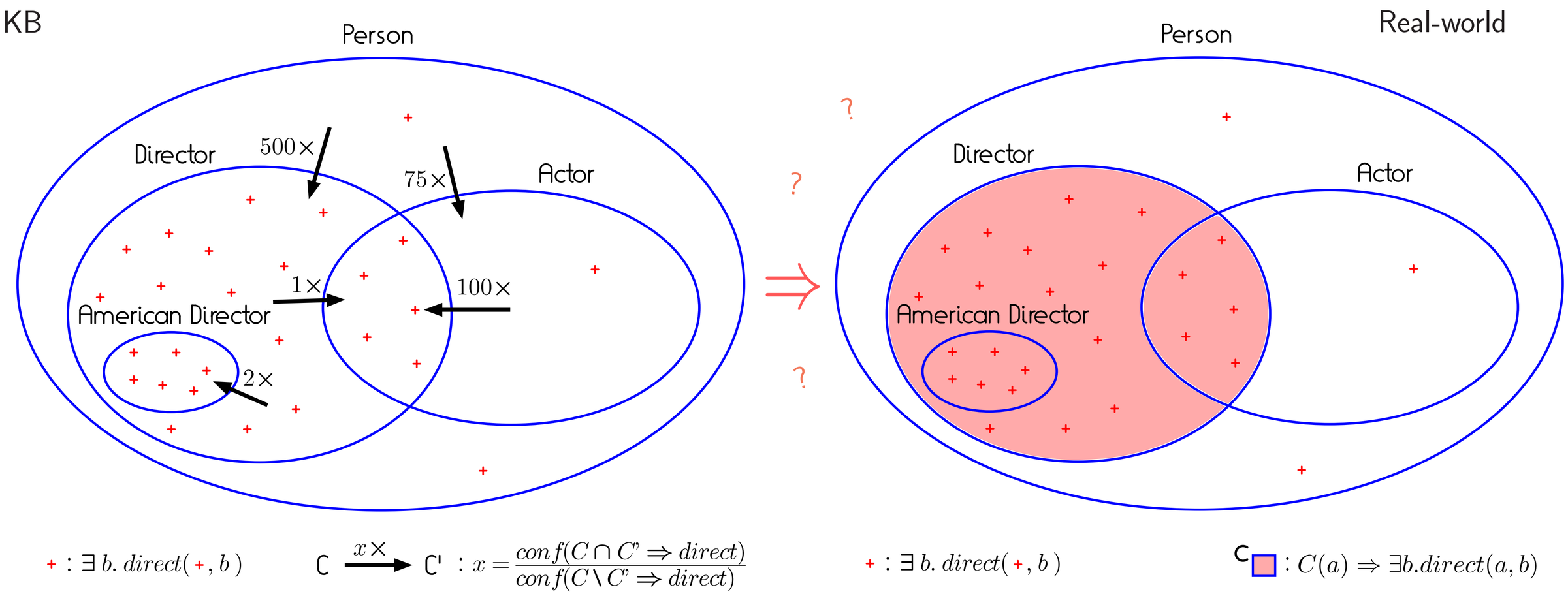
- K is correct and incomplete
- Typing is correct and complete
- K is a **random sample** of the real-world (W)

K can be considered as W with randomly deleted facts.

From W to K:



Approach: tracking the bias



Rationale of the approach

Thanks to the *random sampling assumption* the distribution of attributes should be preserved (with high probability).

Necessary condition

If an attribute is obligatory for a class C, the density of this attribute in C should be the same in every intersection of C with other classes, e.g:

The ratio of all class-internal arrows should be around 1.

In our example

We can conclude here:

- $\neg(Person(a) \Rightarrow_W \exists b.direct(a, b))$,
- $\neg(Actor(a) \Rightarrow_W \exists b.direct(a, b))$.
- *Director* may be solution. (Depending on the threshold)
- *American director* may be solution.

Conclusion

All YAGO classes could not be depicted in our example, but the results are consistent.

- *Promising approach.*
- Can be applied to conjunction of atoms.
- In practice, the ratio is never exactly 1:
  - What is a good threshold ?
  - Should we use a threshold or another classifier such as the MinCut algorithm ?
  - What if a ratio is very low ?

Thank you