

Statistical learning from big data

Where it works...



Marco Baroni

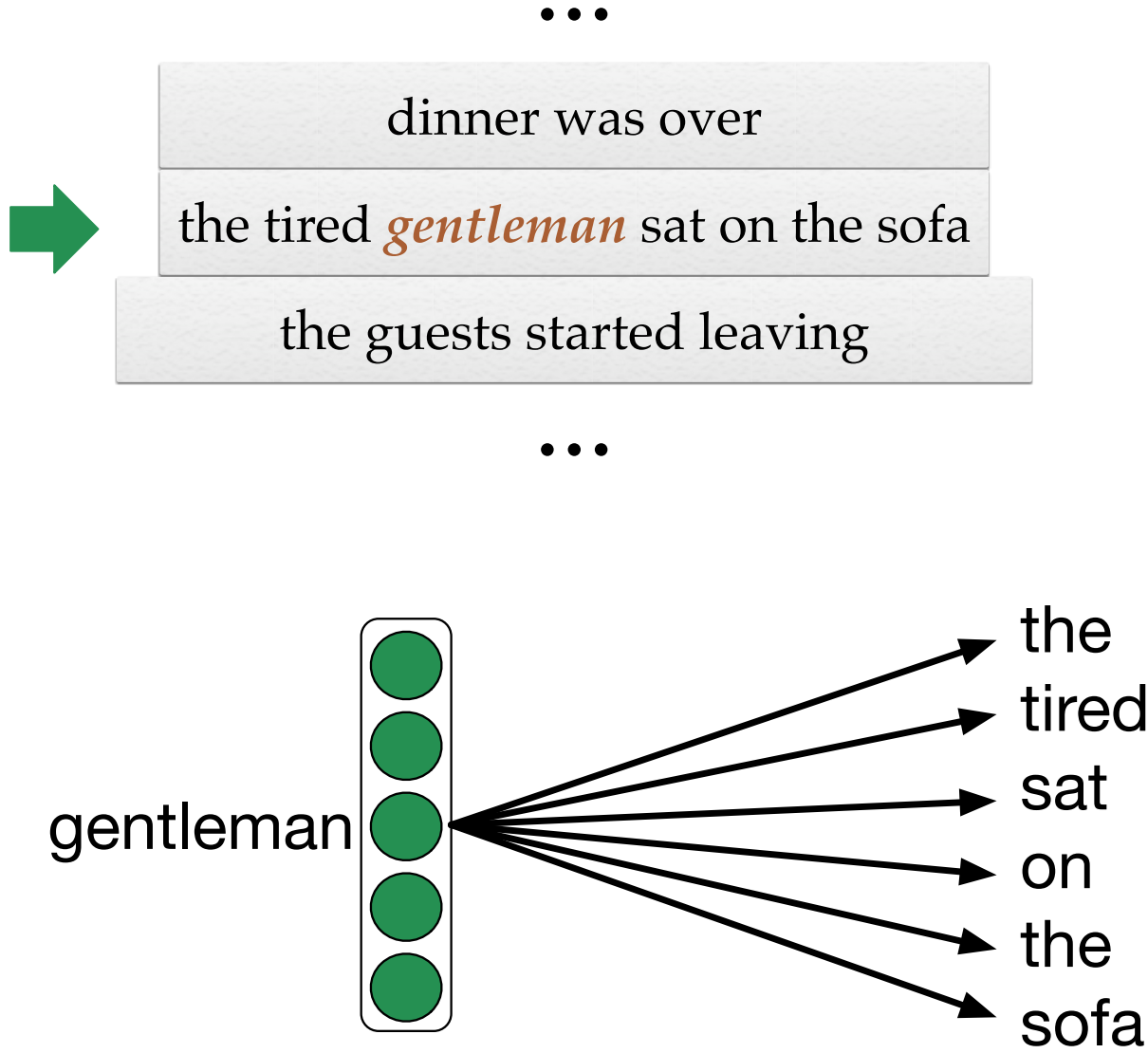
Facebook Artificial Intelligence Research

Near Synonymy

(from Boleda and Erk 2015)

- man:
 - +human +male +adult ±married
- bachelor:
 - +human +male +adult +married
- gentleman, lad, chap, dude, bloke, guy:
 - +human +male +adult ±?????

Corpus-based distributed representations



Men in distributed semantic space

- **man:** woman, gentleman, gray-haired, boy, person
- **gentleman:** gentlewoman, Hunsden, Lestrade, Utterson, Scotchman
- **lad:** boy, bloke, scouser, lass, youngster
- **bloke:** chap, guy, ***, ***, fella
- **chap:** bloke, guy, lad, fella, man
- **dude:** freakin', woah, dorky, dumbass, stoopid
- **guy:** bloke, chap, doofus, dude, fella
- **bachelor:** bachelor's, master's, doctorate, majoring, degree

Multimodal learning

Visual Dialog



A cat drinking water out of a coffee mug.



White and red

What color is the mug?



No, something is there can't tell what it is

Are there any pictures on it?



Yes, they are

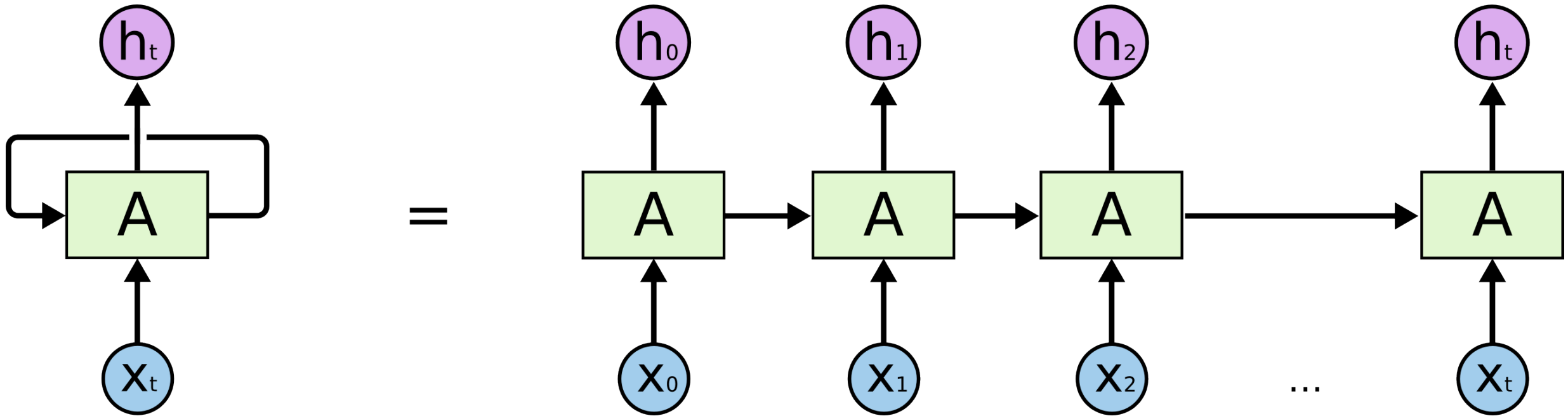
Is the mug and cat on a table?



Yes, magazines, books, toaster and basket, and a plate

Are there other items on the table?

No assumption about structure



<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

The agreement test

Linzen et al. 2016

RNN Accuracy

- The **key is** on the table **99%**
- The **length** of the forewings **is...** **90%**
- The **landmarks** *that this article lists here are...* **90%**
- The **landmarks** *this article lists here are...* **75%**

Statistical learning from big data

Where it doesn't...

Easy learning from explicit information

A **samovar** (**Russian**: самовар, IPA: [səməˈvar] listen[ⓘ]); literally "self-boiler", **Azerbaijani**: samovar, **Turkish**: semaver) is a heated metal container traditionally used to heat and boil water in and around **Russia** as well as in other countries: **Central Europe**, **South-Eastern Europe**, **Eastern Europe**, **Iran**, **Kashmir** and the **Middle-East**. Since the heated water is typically used to make **tea**, many samovars have a ring-shaped attachment (**Russian**: конфорка) around the chimney to hold and heat a **teapot** filled with tea concentrate.^[1]

Though traditionally heated with **coal** or **charcoal**, many newer samovars use electricity to heat water in a manner similar to an **electric water boiler**. Antique samovars are often prized for their beautiful workmanship.

In western countries, some types of **electric water boilers** fulfil the same function of the Samovar, being used to prepare hot water for drinks in a similar fashion.

Contents [\[hide\]](#)

- [1 Description](#)
- [2 History](#)
- [3 Russia](#)
- [4 Iran](#)



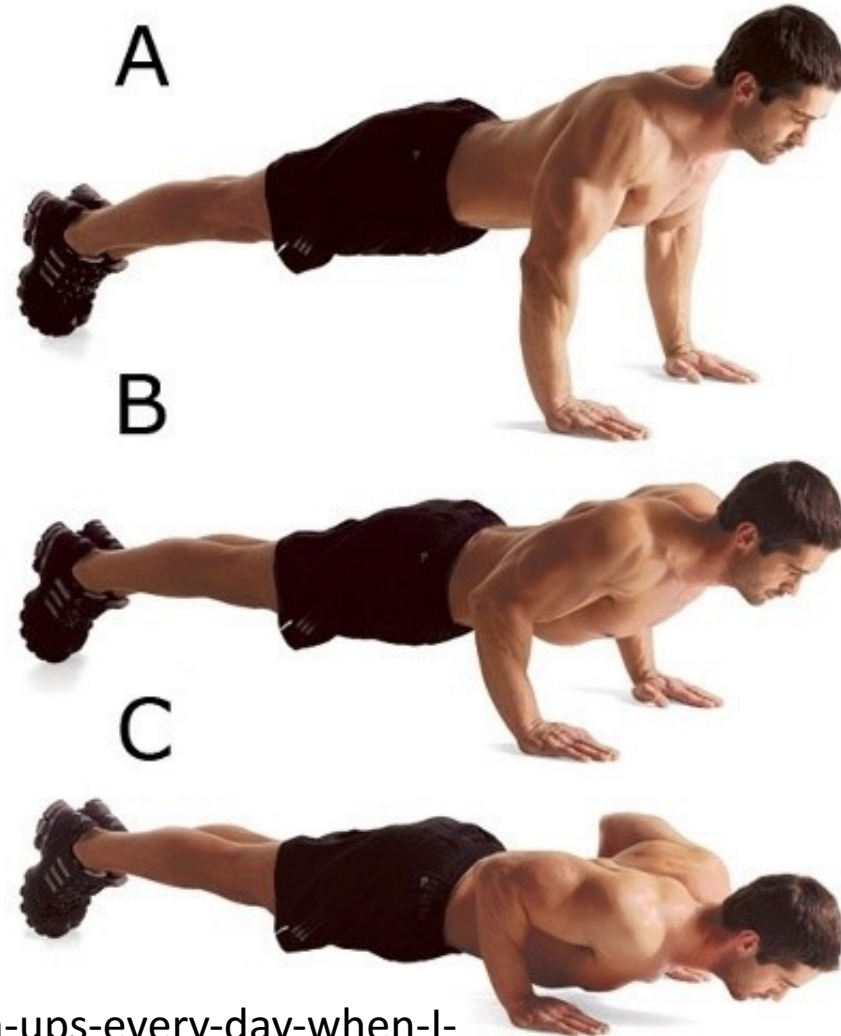
Russian samovar



Easy learning from explicit information

- Make sure you saute the vegetables before you add them to the mix
- How do you saute them?
- Pre-heat a pan with just a thin surface of oil over it... when it's hot, add the vegetables, stir, and after 30 seconds or so set to medium heat, then keep stirring until they look ready...
- Great, thanks for the advice!

Easy learning from explicit information



<https://www.quora.com/Can-I-do-push-ups-every-day-when-I-wake-up-If-so-how-many-sets-and-reps>

Life is a continuous stream of different tasks

- Interact with objects
 - categorize them
 - predict where they will be next
 - grasp them and manipulate them
 - ...
- Communicate
 - ask for directions
 - engage in social/economic transactions
 - write to somebody
 - read a text
 - translate a text
 - ...
- ...

Small skills and their composition

- Motor skills
- Segmenting perceptual input into meaningful discrete units
- Naive physics
- Storing/retrieving information from long-term memory
- Symbol manipulation
- Using a specific tool
- Counting
- The languages you speak
- ...

The small data challenge

- Repeat(101,2)
- Reverse(001)
- Concatenate(101,1111)
- Reverse(Repeat(1000,1))
- Concatenate(Reverse(010101),Repeat(001100,3))