Longitudinal *word2vec*

CoNexDat [1] · Universidad de Buenos Aires [2] · CONICET [3] · ITBA [4]

Carlos Selmo[1,4]
Julián F. Martinez[1,2,3]
Mariano G. Beiró[1,2,3]
J. Ignacio Alvarez-Hamelin[1,2,3]

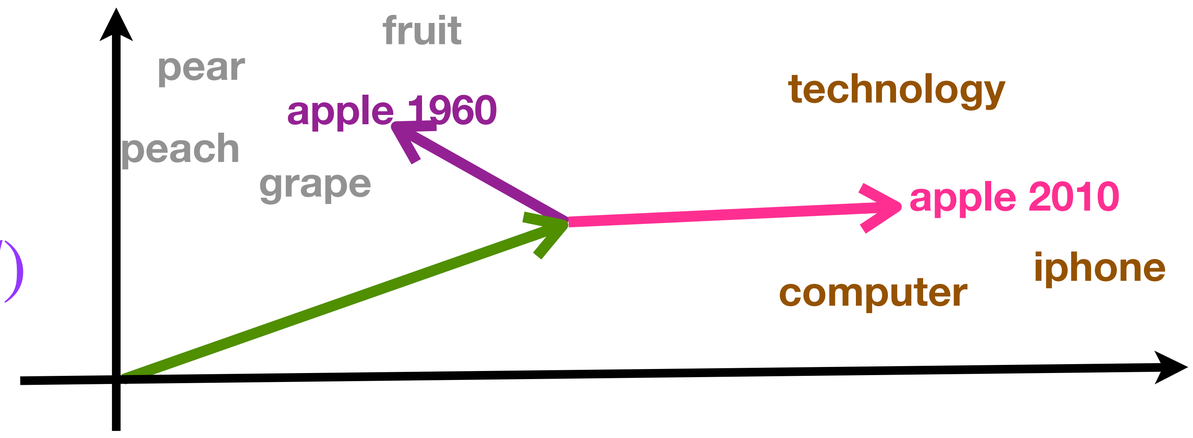
Trans-Atlantic Platform — Social Sciences and Humanities — Digging into Data Challenge

OpLaDyn project

Introduction

How to make a **word embedding** where a word has a relative position *according to* its meaning on every **year** ?

$$\overrightarrow{word}_{year} = \overrightarrow{common}(word) + \vec{\delta}_{year}(word)$$



Dataset

The New York Times newspaper has fully digitalized articles from 1970 to nowadays. We selected articles from 1990 to 2016.

Model

A [Neural Network](#) is trained by maximising the pseudo-likelihood:

$$P_{year}\left(w^O \mid w^I\right)$$

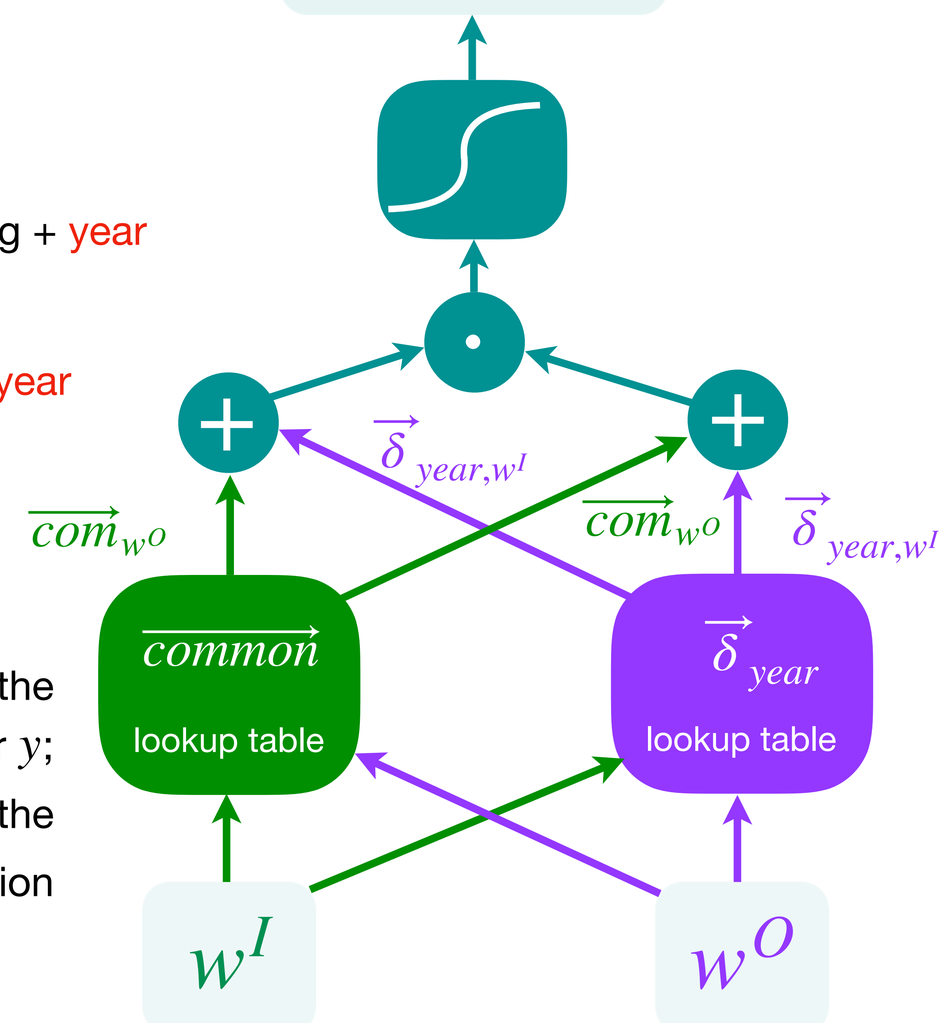
$$\mathscr{L} = \sum_{y=1} [\mathscr{L}^y_{pos} + \mathscr{L}^y_{neg} + \mathscr{L}^y_{reg}]$$

$$\mathscr{L}^y_{pos} = \frac{1}{T_y} \sum_{t=1}^{T_y} \sum_{-c \le j \le c, j \neq 0} \log \sigma\left(v'^{\mathrm{T}}_{y,w_y,t+j} \cdot v_{y,w_y,t}\right)$$, word2vec's positive sampling + year

$$\mathscr{L}^y_{neg} = \sum_{k=1}^{K} \mathbb{E}_{P_n(\cdot \mid y)} \left[\log \sigma\left(-v'^{\mathrm{T}}_{y,W^k_O} \cdot v'_{y,W^k_I}\right)\right]$$, word2vec's negative sampling + year

$$\mathscr{L}^y_{reg} = - \sum_{w \in V_y} \lambda_{y,w} \|\delta_{y,w}\|^2$$, **our regularization term**

where: v' is the vector representing a word w; y is the year; $-c \le j \le c$ is the word j of the context window with size c; T_y is the size of vocabulary at year y; W^k_O and W^k_I are independent random words sampled from $P_n(\cdot \mid y)$, the distribution obtained by $P_n(w \mid y) \propto U_y(w)^{\frac{3}{4}}$ with $U_y(w)$ the unigram distribution of the year y; x^{T} is the vector x transposed.



Results

A top-10 words close to **Bush 1990** along years

rank	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1	bush	bush	bush	clinton	clinton	clinton	clinton	clinton	clinton	clinton	clinton	bush	bush	bush	bush	bush	bush	bush	bush	obama	obama	obama	obama	obama	obama	obama	obama
2	bush's	pledge	president's	bush	bush	clinton's	pledge	congress	pledge	congress	administration	bush's	president's	bush's	president's	bush's	bush's	president's	bush's	obama's	obama's	obama's	boehner	kerry	president's	president's	obama's
3	administration	governor	bush's	bush's	administration	pledge	gingrich	annan	congress	governor	pledge	clinton	bush's	congress	bush's	pledge	congress	bush's	lawmakers	bush	bush	proposal	obama's	president's	obama's	obama's	president's
4	dole	bush's	governor	administration	congress	gingrich	dole's	congress	lawmakers	mccain	bush	annan	proposal	administration	clinton	administration's	president's	congress	administration	president's	bush's	boehner	president's	obama's	congress	congress	congress
5	congress	congress	presidency	clinton's	pledges	dole's	congress	pledge	gingrich	administration	congress	administration's	congress	pledge	annan	president's	mccain	resolution	mccain	clinton	pledge	bush	proposal	proposal	kerry	kerry	bush
6	pledge	proposal	pledge	president's	clinton's	congress	lawmakers	gingrich	resolution	bush's	gore	administration's	congress	administration	lawmakers	kerry	negotiators	president's	pledge	mccain's	biden	congress	bush's	clinton	lawmakers	proposal	bush	kerry
7	cheney	proposals	congress	gore	administration's	administration	negotiators	senate	annan	gore	governor	congress	administration	lawmakers	kerry	negotiators	president's	pledge	mccain's	biden	congress	bush's	clinton	lawmakers	proposal	bush	kerry
8	lawmakers	president's	clinton	congress	dole	bush	election	negotiators	gore	clinton's	resolution	proposal	administration's	annan	congress	annan	obama	sen	pledge	president's	president's	resolution	bush	congress	negotiators	pledges	bush's
9	baker's	administration's	administration's	lawmakers	gingrich	panetta	gore	albright	newt	annan	annan	lawmakers	lott's	resolution	lawmakers	delay	annan	obama's	governor	hillary	pledges	congress	administration	resolution	bipartisan	pledge	administration
10	negotiators	gorbachev	administration	compromise	allies	governor's	clinton's	bipartisan	administration	administration's	clinton's	president's	lawmakers	roh	cheney	senate	boehner	sarkozy's	administration	resolution	clinton	clinton	allies	kerry's	bush	lawmakers	proposal

A top-10 words close to **Iraq 2005** along years

rank	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1	afghanistan	iraq	afghanistan	afghanistan	somalia	afghanistan	afghanistan	bosnia	bosnia	kosovo	kosovo	chechnya	afghanistan	afghanistan	iraq	iraq	afghanistan	afghanistan	afghanistan	iraq	iraq	iraq	iraq	afghanistan	afghanistan	iraq	iraq
2	kuwait	kuwait	somalia	somalia	chechnya	chechnya	bosnia	iraq	afghanistan	afghanistan	afghanistan	iraq	chechnya	kosovo	afghanistan	afghanistan	iraq	iraq	iraq	afghanistan	afghanistan	afghanistan	afghanistan	iraq	iraq	afghanistan	afghanistan
3	iraqi	algeria	kuwait	lebanon	kuwait	somalia	chechnya	troops	kuwait	bosnia	chechnya	afghanistan	kosovo	iraq	kosovo	kosovo	somalia	darfur	forces	troops	somalia	iraqi	libya	iraqi	yemen	syria	syria
4	rebel	army	balkans	iraq	iraqi	bosnia	saudi	rwanda	troops	chechnya	bosnia	bosnia	afghan	bosnia	afghan	kuwait	darfur	somalia	troops	yemen	pakistan	somalia	somalia	libya	syria	libya	forces
5	baghdad	iraqi	algeria	cambodia	troops	kuwait	arabia	peacekeeping	rwanda	iraq	iraqi	iraqi	kuwait	iraqi	chechnya	afghan	troops	libya	syria	pakistan	afghan	afghan	pakistan	pakistan	pakistan	pakistan	taliban
6	saudi	forces	bosnian	haiti	iraq	iraq	chechen	rebels	kosovo	chechen	troops	kuwait	iraqi	chechnya	troops	sudan	lebanon	sudan	sudanese	lebanon	forces	pakistan	iraqi	troops	iraqi	afghan	afghan
7	libya	cambodia	iraq	sudan	afghanistan	iraqi	libya	arabia	iran	iraqi	pakistan	chechen	iraqis	arabia	chechen	darfur	sudanese	taliban	somalia	afghan	troops	libya	afghan	somalia	kobani	iraqi	troops
8	afghan	kurdish	forces	troops	algeria	forces	baghdad	cambodia	iraqi	bosnian	iraq	kosovo	syria	iran	darfur	forces	pakistan	nato	lebanon	forces	iraqis	yemen	syria	afghan	syrian	forces	libya
9	yemen	baghdad	libya	yugoslavia	chechen	yugoslavia	iraq's	civilians	taliban	taliban	yugoslavia	balkans	iran	libya	lebanon	iraqi	israeli	lebanon	pakistan	somalia	syria	iraqis	military	sudan	troops	lebanon	iraqi
10	forces	egypt	iraqi	kuwait	rebels	algeria	troops	zaire	military	troops	chechen	syria	bosnia	forces	forces	pakistan	syria	iraqi	iran	iraqi	yemen	troops	forces	yemen	libya	somalia	pakistan

Git-hub repository for NYT 1990-2016 embedding