

TENSOR DECOMPOSITIONS FOR TEMPORAL KNOWLEDGE BASE COMPLETION

FaceBook Research

ICLR 2020

目录

Contents

- 研究背景
- 相关研究
- 论文方法
- 论文贡献
- 实验效果
- 结论和启发

研究背景

Background

- 知识图谱表示学习 (Knowledge Graph Embedding)
 - 将知识图谱中的实体和关系表示为低维向量空间中的向量， 实体和关系的复杂信息蕴含在向量中
 - 转移距离模型，TransE、TransH、TransR、TransD等
 - 语义匹配模型，RESCAL、DistMult、Complex等
 - 神经网络模型，ConvE、Hyper等
- 链路预测 (Link Prediction)
 - 根据知识图谱中已存在的事实去预测缺失事实的任务
 - 基于张量分解的TuckER模型在链路预测任务上取得了最先进的结果
- 上述模型都没有为三元组加入时间 (Temporal) 信息，无法准确回答诸如 (美国, 总统, ?, 2020年) 此类查询

相关研究

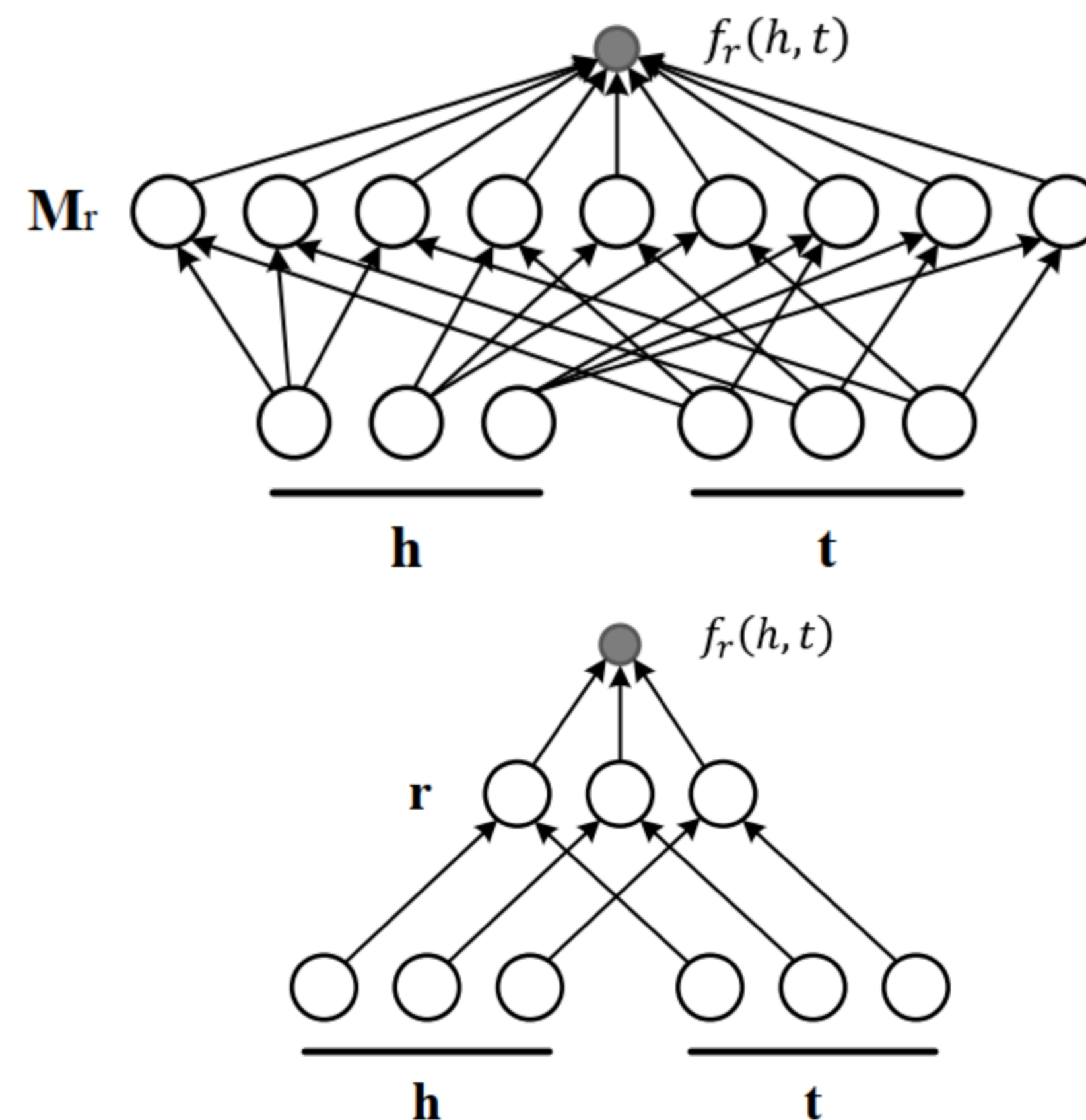
Relevant Research

- **RESCAL模型**

- 评分函数 $f_r(h, t) = h^T M_r t$
- 实体和关系的信息可以进行深层次交互，为强表达模型
- 随着关系矩阵维度的增加，复杂度会很高
- 容易过拟合

- **DistMult模型**

- RESCAL的一种特殊情况，放松对关系矩阵的约束，把关系矩阵利用对角矩阵表示
- 评分函数为 $f_r(h, t) = h^T \text{diag}(M_r) t$
- 不能对不对称关系建模



相关研究

Relevant Research

- **Complex模型**

- 将DistMult扩展到复数域，同一个实体的主客体向量是共轭的
- 实数向量之间的点积计算是具有交换性的，使用点积不能处理非对称的关系
- 共轭向量使得点积可以在不对称关系中用起来，保留点积的效率优势，即空间和时间复杂度均为线性
- 得分函数不再具有对称性

$$\begin{aligned}\phi(r, s, o; \Theta) &= \text{Re}(\langle w_r, e_s, \bar{e}_o \rangle) \\ &= \text{Re}\left(\sum_{k=1}^K w_{rk} e_{sk} \bar{e}_{ok}\right) \\ &= \langle \text{Re}(w_r), \text{Re}(e_s), \text{Re}(e_o) \rangle \\ &\quad + \langle \text{Re}(w_r), \text{Im}(e_s), \text{Im}(e_o) \rangle \\ &\quad + \langle \text{Im}(w_r), \text{Re}(e_s), \text{Im}(e_o) \rangle \\ &\quad - \langle \text{Im}(w_r), \text{Im}(e_s), \text{Re}(e_o) \rangle\end{aligned}$$

相关研究

Relevant Research

- TuckER模型

- 将三阶张量分解为一组矩阵和一个更小的核心张量

- $\phi(e_s, r, e_o) = \mathcal{W} \times_1 \mathbf{e}_s \times_2 \mathbf{w}_r \times_3 \mathbf{e}_o$

- 通过训练，知识存储在核心张量中，而不是编码到嵌入中
- 核心张量的参数数量只依赖于实体和关系嵌入维度
- DistMult和ComplEx可以看作TuckER的特例

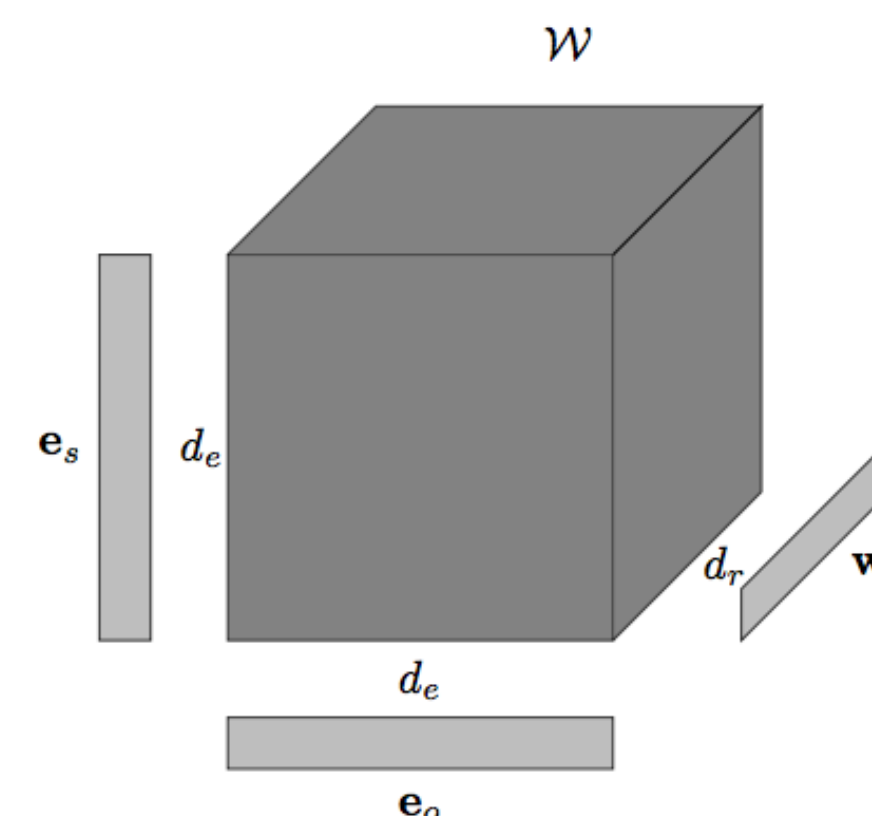
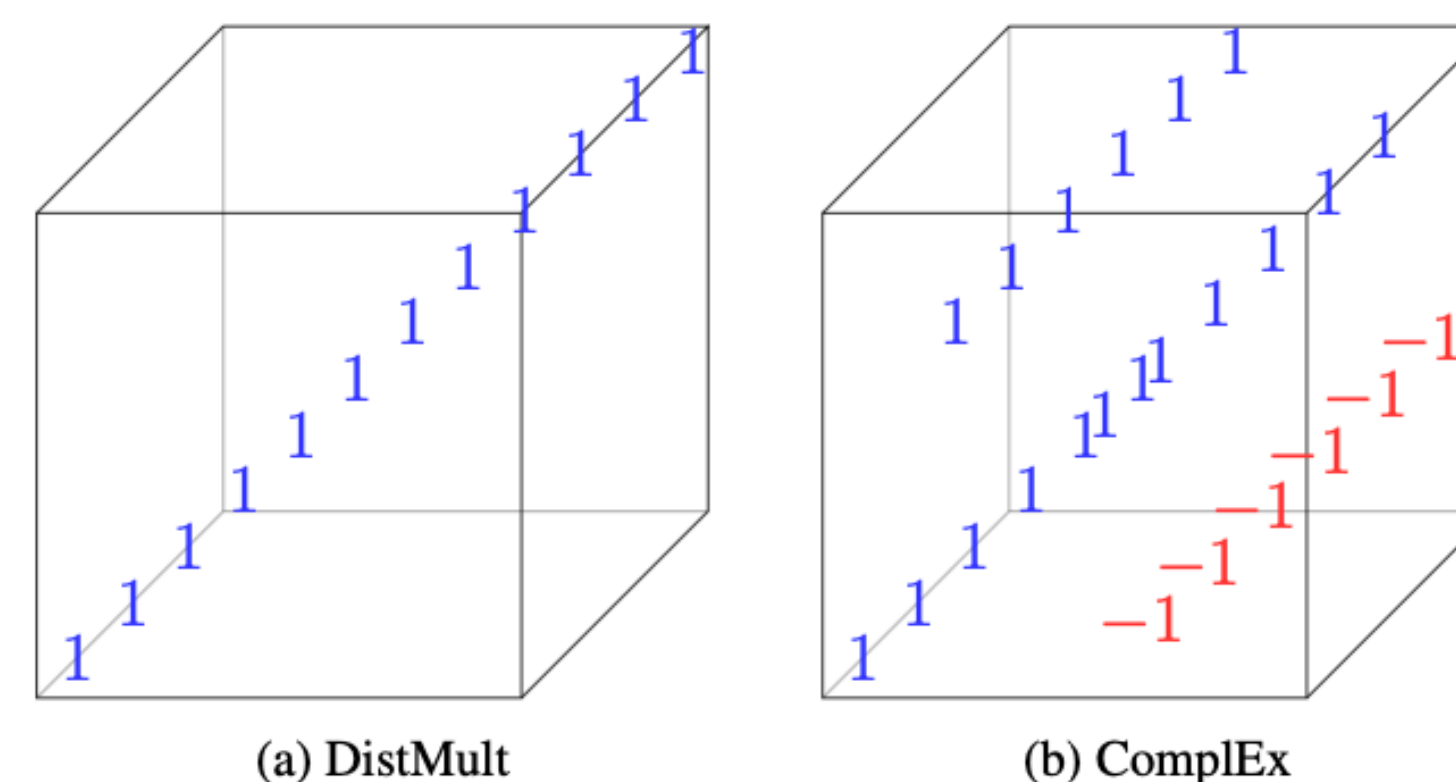


Figure 1. Visualization of the TuckER architecture.



论文方法

Methods

- **TComplEx模型**

- 添加一个新的因子T，在ComplEx模型中引入时间嵌入

- $\hat{X}(U, V, T) = \text{Re}([U, V, \overline{U}, T]) \iff \hat{X}(U, V, T)_{i,j,k,l} = \text{Re}(\langle u_i, v_j, \overline{u_k}, t_l \rangle)$

- $\langle u_i, v_j, \overline{u_k}, t_l \rangle = \langle u_i \odot t_l, v_j, \overline{u_k} \rangle = \langle u_i, v_j \odot t_l, \overline{u_k} \rangle = \langle u_i, v_j, \overline{u_k} \odot t_l \rangle$

- **TNTComplEx模型**

- 将张量X分解为两个张量之和，一个为时间，另一个为非时间

- $\hat{X} = \text{Re}([U, V^t, \overline{U}, T] + [U, V, \overline{U}, \mathbf{1}]) \iff \hat{X}_{i,j,k,l} = \text{Re}(\langle u_i, v_j^t \odot t_l + v_j, \overline{u_k} \rangle)$

参数数量

ComplEx	$2r(E +2 P)$
TComplEx	$2r(E + T +2 P)$
TNTComplEx	$2r(E + T +4 P)$

论文方法

Methods

- 离散化时间戳范围（如通过将时间戳缩减为年份），将三元组扩展为四元组（主语，谓语，宾语，时间）
- 将连续的时间戳（如年，日及其数值属性）注入到正则化项中

- 嵌入正则项 Ω

$$\Omega^3(U, V, T; (i, j, k, l)) = \frac{1}{3} (\|u_i\|_3^3 + \|u_k\|_3^3 + \|v_k \odot t_l\|_3^3)$$
$$\Omega^3(U, V^t, V, T; (i, j, k, l)) = \frac{1}{3} (2\|u_i\|_3^3 + 2\|u_k\|_3^3 + \|v_j^t \odot t_l\|_3^3 + \|v_j\|_3^3)$$

- 时间正则项 Δ

$$\Delta^p(U, V, T; (i, j, k, l)) = \frac{1}{p} (2\|u_i\|_p^p + 2\|u_k\|_p^p + \|v_j^t\|_p^p + \|t_l\|_p^p + \|v_j\|_p^p)$$

- 时间嵌入平滑

- 相近的时间戳有更相似的表示

$$\Lambda_p(T) = \frac{1}{|T| - 1} \sum_{i=1}^{|T|-1} \|t_{i+1} - t_i\|_p^p$$

论文贡献

Contributions

- 拓展ComplEx模型，引入时间嵌入，利用张量分解解决带有时间信息的知识补全问题
- 提出了时间链路预测数据集
 - 基于Wikidata
 - 删除了对象不是实体的所有事实。
 - 筛选出度数至少为5的实体和谓词至少发生50次的实体
 - 包含432715个实体，407个谓词和1724个时间戳的数据集（仅保留年份）
 - 训练集大小700万个，验证集和测试集各5万个

实验效果

Experiments

	ICEWS14	ICEWS15-05	Yago15k
TA	0.48	0.47	0.32
DE-Simple	0.53	0.51	-
ComplEx	0.47 (0.47)	0.49 (0.49)	0.35 (0.36)
TComplEx	0.56 (0.61)	0.58 (0.66)	0.35 (0.36)
TNTComplEx	0.56 (0.62)	0.60 (0.67)	0.35 (0.37)

Table 2: Results for TA (García-Durán et al., 2018) and DE-Simple (Goel et al., 2020) are the best numbers reported in the respective papers. Our models have as many parameters as DE-Simple. Numbers in parentheses are for ranks multiplied by 10.

Reg.	MRR
No regularizer	0.62
Δ^2	0.63
Δ^3	0.63
Δ^4	0.64
Ω^3	0.65
$\Omega^3 + \Lambda_4$	0.67

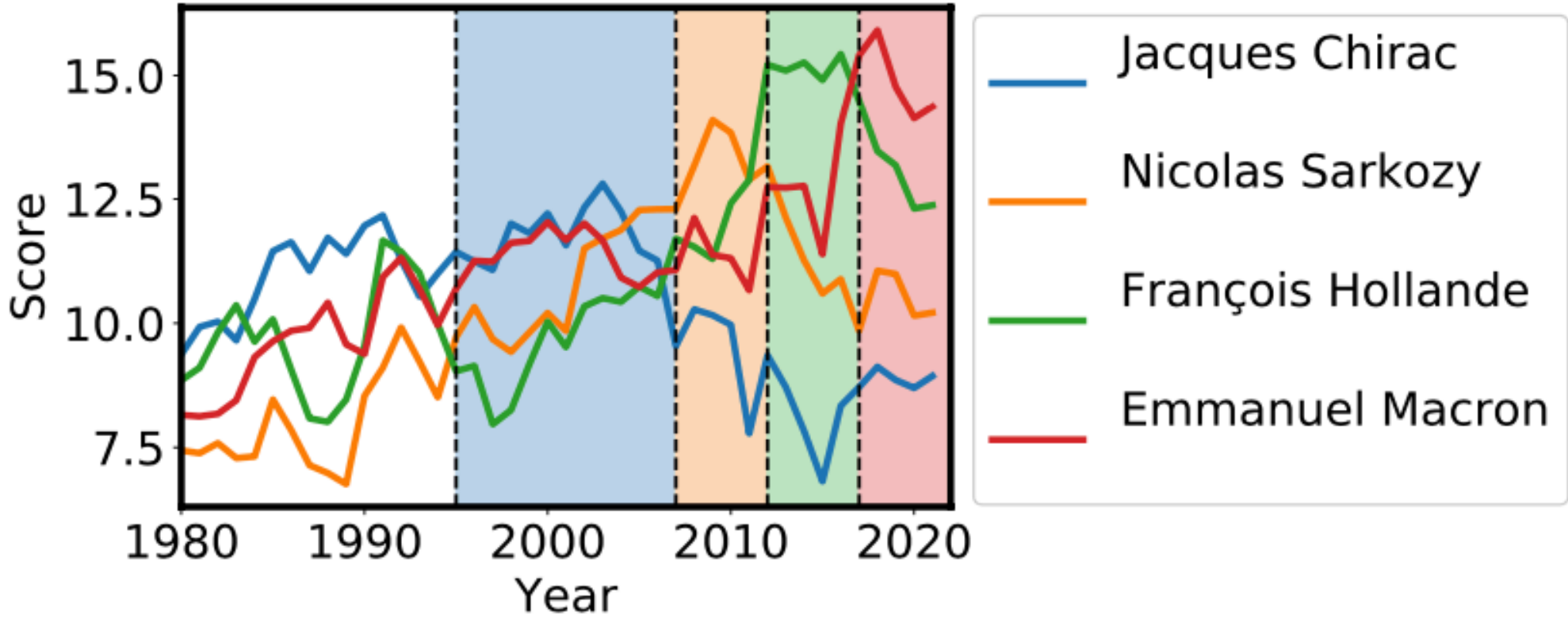
Table 3: Impact of regularizers on ICEWS05-15 for TNTComplEx.

实验效果

Experiments

	MRR	NT-MRR	T-MRR
ComplEx	0.45	0.48	0.29
TComplEx	0.42	0.45	0.30
TNTComplEx	0.44	0.47	0.32

Table 4: Results on wikidata for entity dimension $d = 100$.



总结和启发

Conclusion&Inspiration

- 提出了用于对时间知识补全TComplEx模型（都看作带有时间的事实）和TNTComplex模型（事实分解为时间事实和非时间事实两部分）
- 在没有大量时间知识的数据集中，引入时间嵌入的ComplEx模型性能与ComplEx效果相当，但前者在时间知识数据集上效果更好
 - 因为即使在Wikidata上，90%的事实没有时间数据，非时间MRR（NT-MRR）远远超过了时间MRR（T-MRR）
- TNTComplex在Wikidata上的更优的实验结果表明，对非时间谓词（关系）进行单独建模是有好处的
- 可能引入其他维度的信息，比如空间
- 新型数据集的产生对相关领域的研究有促进作用

Thank you
祝大家午餐愉快

