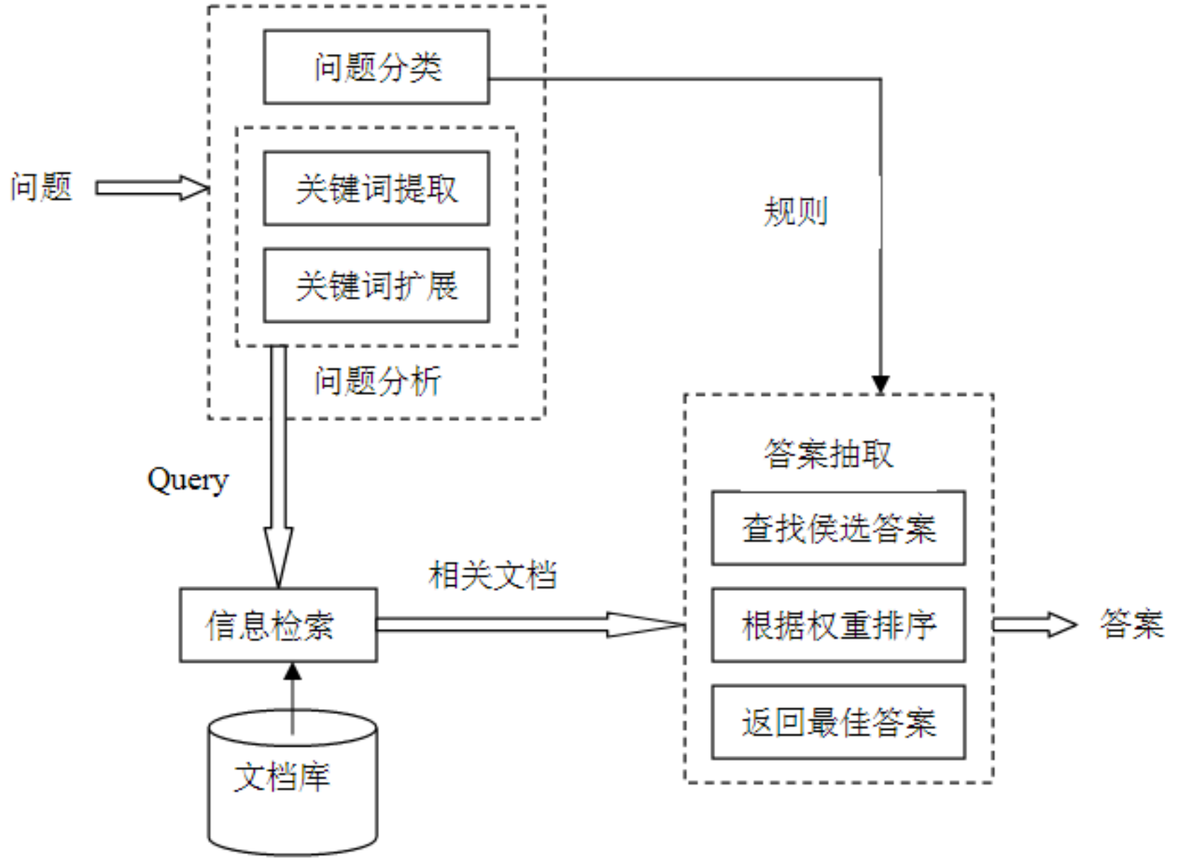
Question Answering + Visual Question Answering

玖强

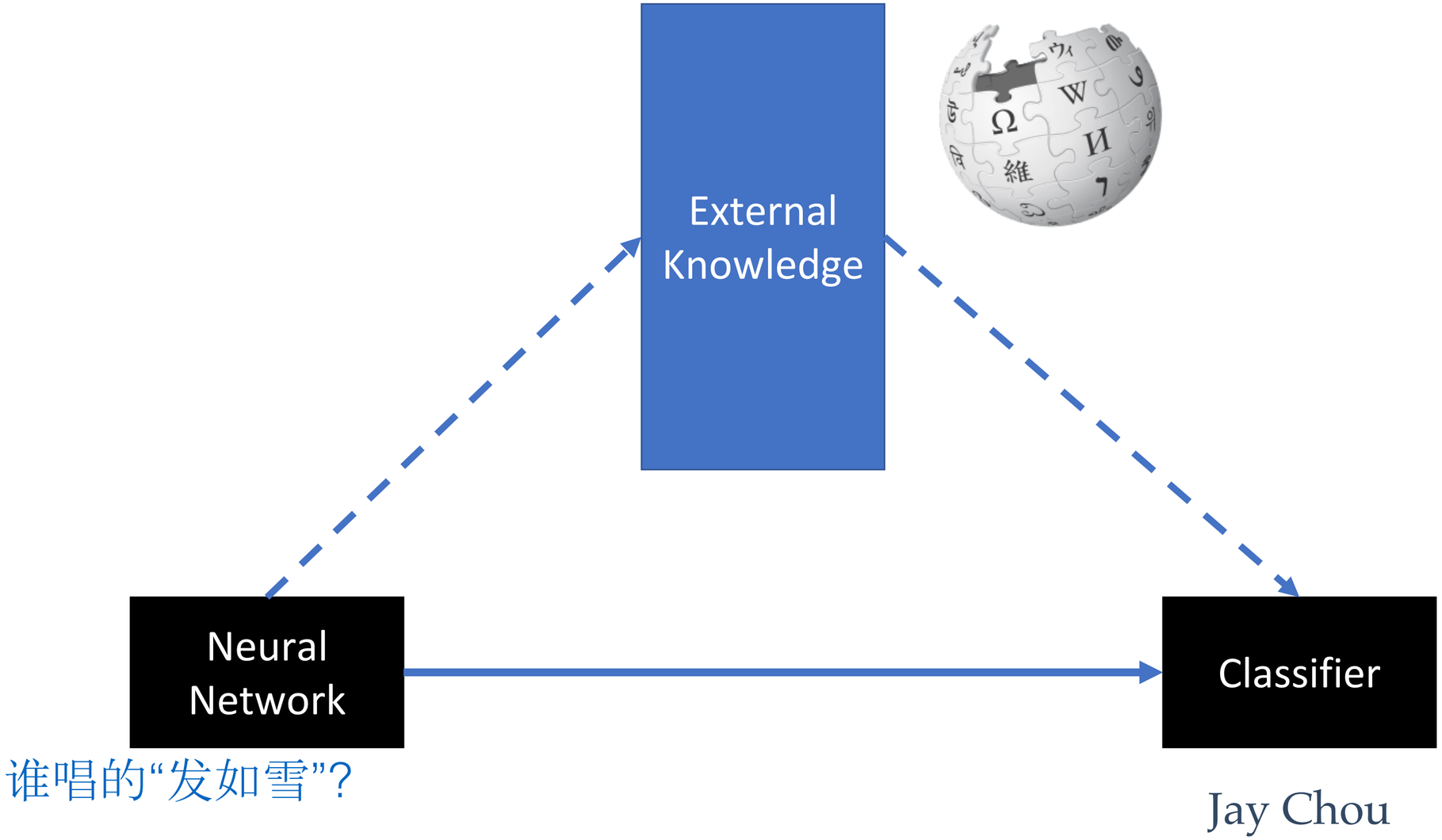
什么是Question answering (QA)

谁唱的“发如雪”?



Jay Chou

用神经网络来替代那些复杂的模块？



搜索急需一场变革

Google

IDC统计2018年数据量是多少

All News Images Videos Maps More Settings Tools

About 586,000 results (0.68 seconds)

IDC 公布2017 年全球智能手机出货量：首次出现下滑，但将在今年恢复 ...
www.ifanr.com/988307 ▾ Translate this page
2017 年全球智能手机出货量统计报告出炉，为有史以来首次出现下滑。 ... 因此，在2018 年，其他操作平台的智能手机出货量将可能会继续呈现出下滑的趋势，甚至将可能会出现仅剩下Android 和iOS 的「双寡头」局面。 不过对于 ... 根据IDC 在本月初公布统计数据显示（均为四舍五入数值），2017 年中国智能手机市场共出货4.4 亿台。

不光是中国IDC数据显示全球智能手机出货量均下滑-科技频道-金融界
finance.jrj.com.cn › 科技频道 ▾ Translate this page
Feb 4, 2018 - 因此在没有更革命性的产品支持下，恐怕这个趋势会延续到2018年。 值得一提的是，IDC统计出的数据，除了苹果和三星遥遥领先之外，前五大手机公司当中已经有三家都来自中国了，华为、小米和OPPO分别成为出货量排名第三第四和第五的企业。 受到市场整体情况的影响，前五大智能手机企业有四家都出现了不同 ...

2018年，如何处置您的物联网数据__财经头条 - 新浪
t.cj.sina.com.cn/.../ec5e47bb020004le7?cre... ▾ Translate this page
Mar 1, 2018 - 获得物联网数据意味着我们将看到更精细的设计、工作流程和战略。根据IDC统计，到2020年，物联网数据将占到全球登记数据的10%。然而，物联网...

2018年，数据中心存储的8大变革方向_IDC国内资讯_中国IDC圈
news.idcquan.com › 新闻资讯 › 国内资讯 ▾ Translate this page
Jan 17, 2018 - SSD技术的发展进步将成为2018年影响企业存储的主要趋势之一。 ... 我们看到，随着SSD固态硬盘对数据中心存储的性能所带来的提升，企业已经从硬盘转向固态硬盘的采用。 在新的2018 .... 再加上压缩技术的采用，可以将网络吞吐量与数据压缩的数据量相提并论，并且存储设备可以在1U设备中有效地支持5PB.
Missing: 统计

Bai du 百度

IDC统计2018年数据量是多少 百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约127,000个 搜索工具

2018年数据中心市场会有哪些新的变化?
2018年3月1日 - 那么2018年数据中心市场会有哪些新的变化呢? ...按照IDC的统计数据,到2020年将拥有超过500亿的终端...谷歌的数据中心用电量已经被节省了几个百分点,...
baijiahao.baidu.com/s?... - 百度快照

2018年全球互联网数据中心市场格局及市场规模情况分析(图) - 中国...
2018年2月9日 - 字体大小:大 中 小2018-02-09 13:49 来源:中国报告网据IDC统计,全球互联网数据中心(IDC)数量在2017年将达到870万座,主要集中在美国、欧洲、日本、...
free.chinabaogao.com/i... - 百度快照

2018年中国IDC市场规模及增长率预测【图】_中国产业信息网
2017年1月22日 - 国内IDC 市场伴随着互联网发展而迅速发展,一方面互联网行业客户由于自身业务发展的需要,对数据中心资源需求旺盛;另一方面 4G、云计算、大数据等网络架...
www.chyxx.com/industry... V3 - 百度快照

IDC统计:2018年全球AR/VR消费额翻倍,将达178亿美元_搜狐科技_搜狐网

2017年12月1日 - 根据国际数据公司(IDC)发布的最新报告预计,2018年全球市场的增强现实(AR)和虚拟现实(VR)消费总额将达178亿美元。 这份《全球半年度增强现实和虚拟现实...
www.sohu.com/a/2077295... - 百度快照

IDC:全球PC第一季度出货量达6038万台,惠普份额持续领先

4天前 - IDC在昨天公布了2018年第一季度全球PC出货量报告,据IDC的统计数据显示今年第一季度全球PC的出货量为6038.3万台,和2017年第一季度相比略有下跌,此前...
baijiahao.baidu.com/s?... - 百度快照

交互方式的转变需要信息服务模式的转变

问答系统是下一代搜索引擎的基本形态

知乎- 发现更大的世界
https://www.zhihu.com/ ▾ Translate this page
注册即代表你同意《知乎协议》注册机构号. 已有帐号？登录. 下载知乎App. 知乎专栏圆桌发现移动应用联系我们来知乎工作注册机构号. © 2018 知乎京ICP 证110745 号京公网安备11010802010035 号出版物经营许可证 · 侵权举报网上有害信息举报专区儿童色情信息举报专区违法和不良信息举报：010-82716601. 诚信网站示范企业.

Results from zhihu.com

发现
编辑推荐 - 神器 - 机器学习 - ...

哪些瞬间让你发现贫穷限制了你
因有朋友在Facebook总部工作，上周中有机会作为visitor身份游览了 ...

话题广场
运动 - 游戏 - 单机游戏 - 艺术 - 汽车 - 投资 - 电影 - ...

知乎- 知乎
知乎是中文互联网最大的知识社交平台，拥有认真、专业和友善的独 ...

Quora - A place to share knowledge and better understand the world.
https://www.quora.com/ ▾
Quora is a place to gain and share knowledge. It's a platform to ask questions and connect with people who contribute unique insights and quality answers. This empowers people to learn from each other and to better understand the world.

People also ask
What is the meaning of Quora?
How do I get rid of Quora?
How do I answer anonymously on Quora?
How do you ask a question on Quora?
Feedback

IBM Watson

- 沃森(Watson)：2011年，IBM研发的超级计算机“沃森”在美国知识竞赛节目《危险边缘Jeopardy!》中上演“人机问答大战”，战胜类选手Ken和Brad

辅助医疗

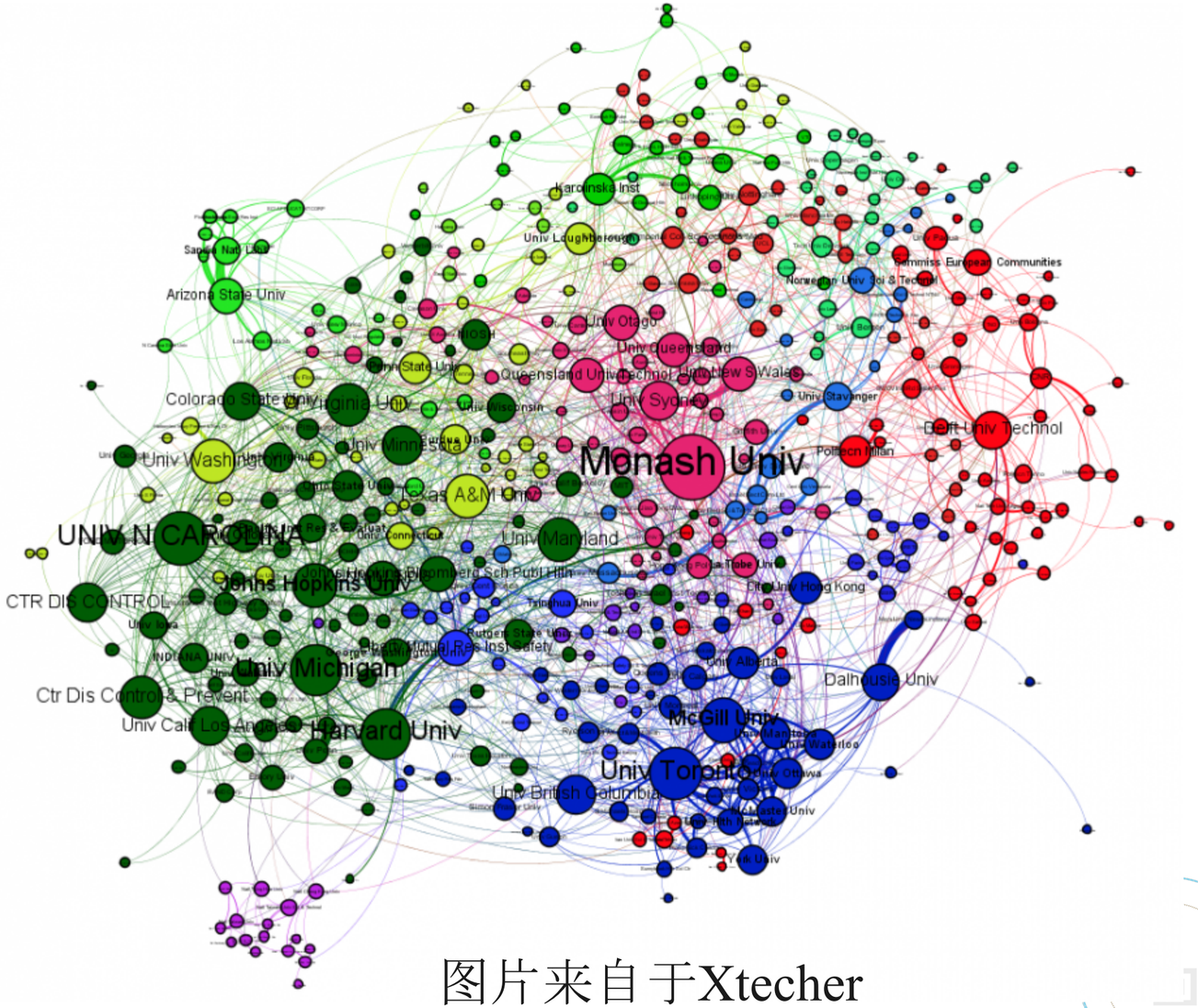
金融辅助决策

企业服务

大规模知识图谱

- 知识图谱就是把所有不同种类的信息（Heterogeneous Information）连接在一起而得到的一个关系网络。知识图谱是基于现有数据的再加工，包括关系数据库中的结构化数据、文本或XML中的非结构化或半结构化数据、客户数据、领域本体知识以及外部知识，通过各种数据挖掘、信息抽取和知识融合技术形成一个统一的全局的知识库。

- 知识图谱提供了从“关系”的角度去分析问题的能力，可被看作是一张巨大的图，图中的节点表示实体或概念，而图中的边则由属性或关系构成。

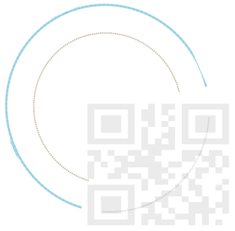


图片来自于Xtecher

OUTLINE

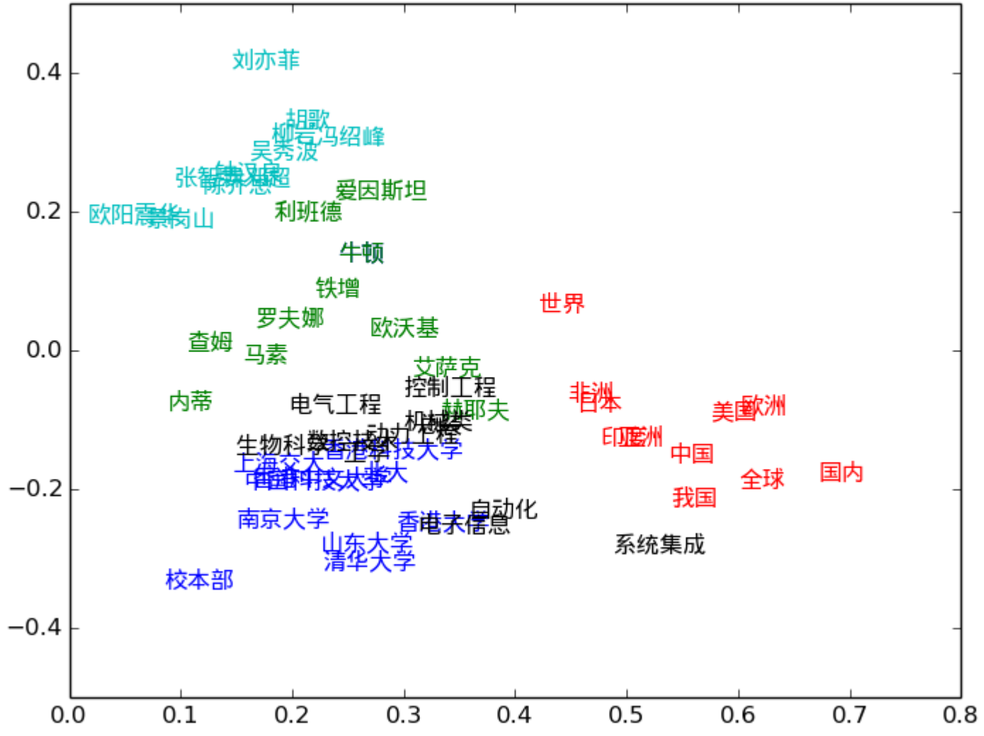
- 回顾深度学习+ NLP 基本概念
- Factoid QA (事实问答)
- Knowledge base QA (知识库问答)
- Visual QA (基于视觉的问答)
- 展望

Neural Networks for QA



回顾

- 我们可以将words用low-dimensional向量来表示(Embeddings)
- e.g., 世界-> [0.23, 1.3, -0.3, ..., 0.43]



word2vec Parameter Learning Explained

一个SENTENCE如何压缩成一个向量？

- **我们如何将那些word embeddings压缩成一个向量，而这个向量能够表示问题的意思？**

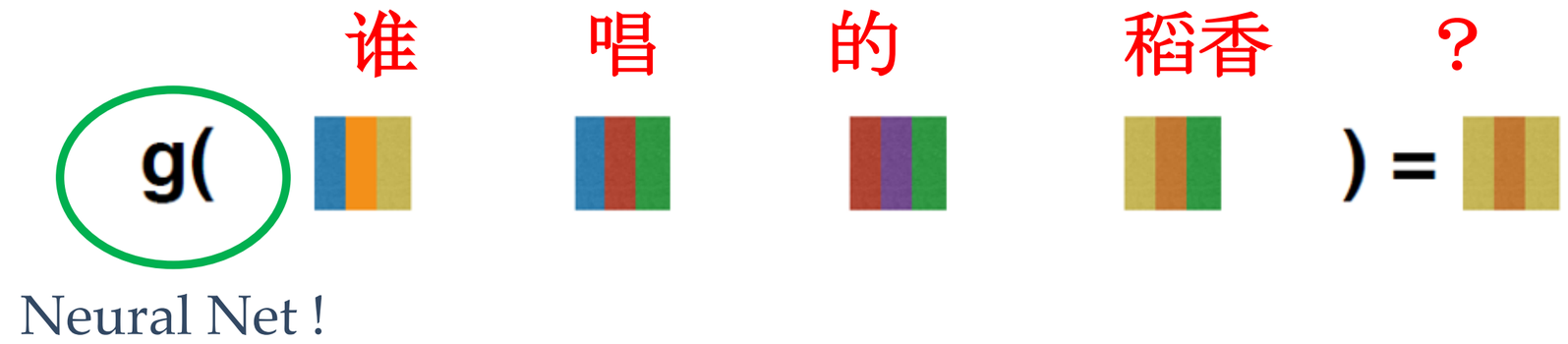
谁　　唱　　的　　稻香　　？

计算问题/SENTENCE的VECTOR

- 我们如何将那些word embeddings压缩成一个向量，而这个向量能够表示问题的意思？

计算问题（SENTENCE）的VECTOR

- 我们如何将那些word embeddings压缩成一个向量，而这个向量能够表示问题的意思？



Recurrent Neural Networks

谁

C1

唱

C2

的

C3

稻香

C4

Recurrent Neural Networks

$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix})$$

谁 唱 的 稻香

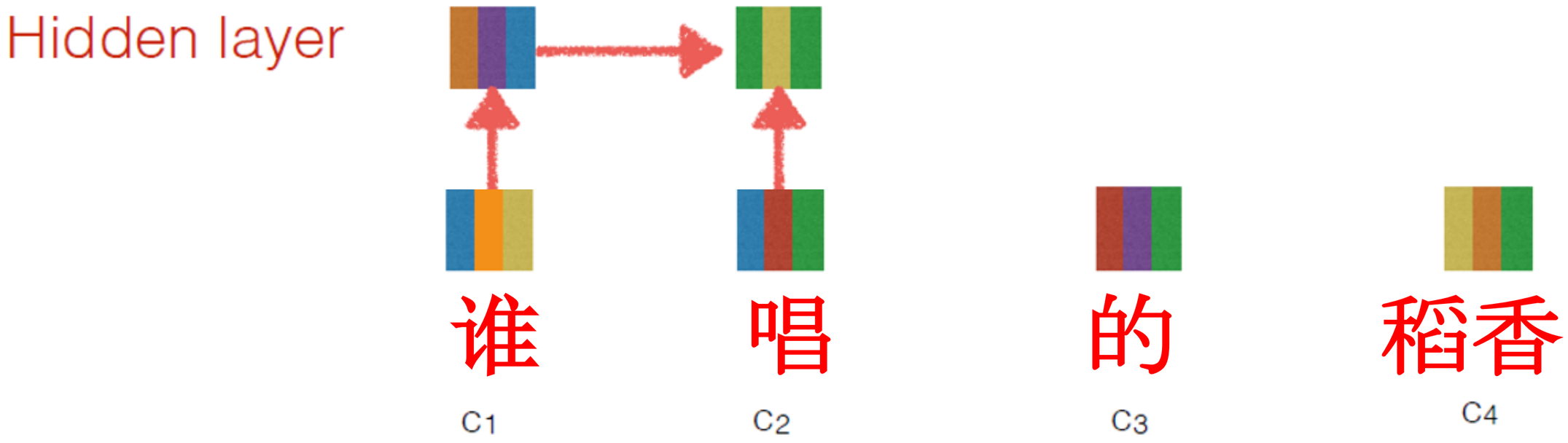
c1 c2 c3 c4

Recurrent Neural Networks

$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix})$$

Recurrent Neural Networks

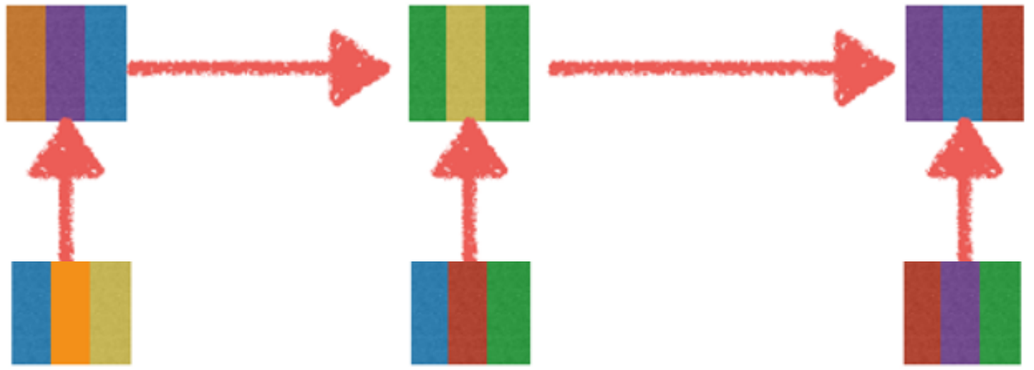
$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix}) \quad h_2 = f(W \begin{bmatrix} h_1 \\ c_2 \end{bmatrix})$$



Recurrent Neural Networks

$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix}) \quad h_2 = f(W \begin{bmatrix} h_1 \\ c_2 \end{bmatrix}) \quad h_3 = f(W \begin{bmatrix} h_2 \\ c_3 \end{bmatrix})$$

Hidden layer



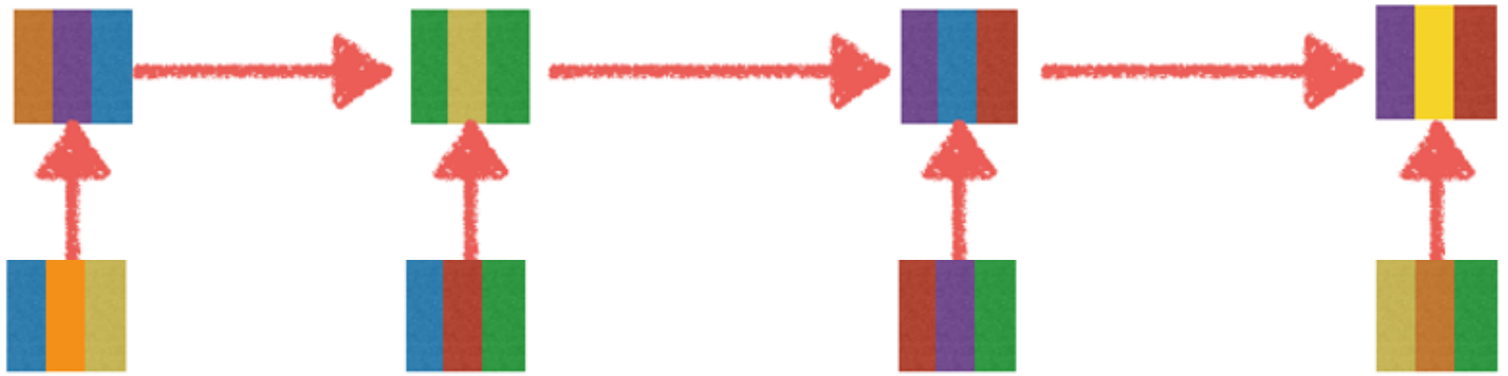
谁 唱 的 稻香

c1 c2 c3 c4

Recurrent Neural Networks

$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix}) \quad h_2 = f(W \begin{bmatrix} h_1 \\ c_2 \end{bmatrix}) \quad h_3 = f(W \begin{bmatrix} h_2 \\ c_3 \end{bmatrix}) \quad h_4 = f(W \begin{bmatrix} h_3 \\ c_4 \end{bmatrix})$$

Hidden layer



谁 唱 的 稻香

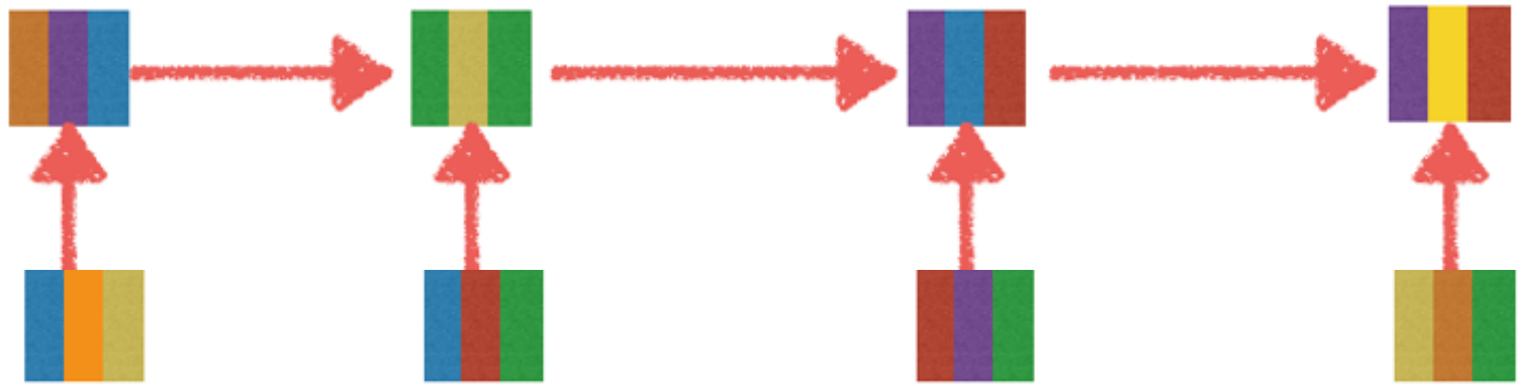
c1 c2 c3 c4

Recurrent Neural Networks

softmax: predict answer

$$h_1 = f(W \begin{bmatrix} \cdots \\ c_1 \end{bmatrix}) \quad h_2 = f(W \begin{bmatrix} h_1 \\ c_2 \end{bmatrix}) \quad h_3 = f(W \begin{bmatrix} h_2 \\ c_3 \end{bmatrix}) \quad h_4 = f(W \begin{bmatrix} h_3 \\ c_4 \end{bmatrix})$$

Hidden layer



谁 唱 的 稻香

c_1 c_2 c_3 c_4

Recurrent Neural Networks

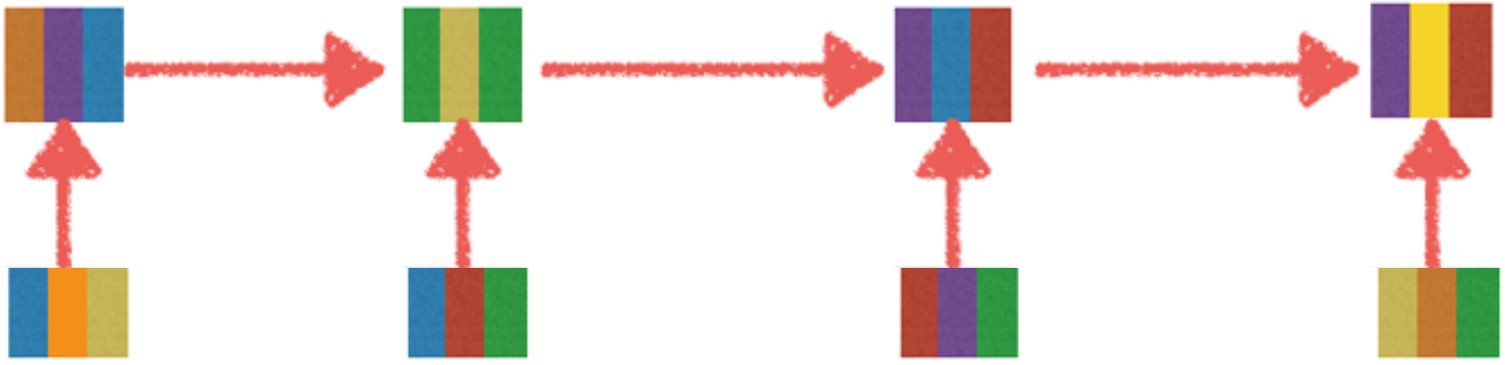
More complex variants:
LSTMs, GRUs

周杰伦

softmax: predict answer

$$h_1 = f(W\begin{bmatrix}\cdots\\ c_1\end{bmatrix}) \quad h_2 = f(W\begin{bmatrix}h_1\\ c_2\end{bmatrix}) \quad h_3 = f(W\begin{bmatrix}h_2\\ c_3\end{bmatrix}) \quad h_4 = f(W\begin{bmatrix}h_3\\ c_4\end{bmatrix})$$

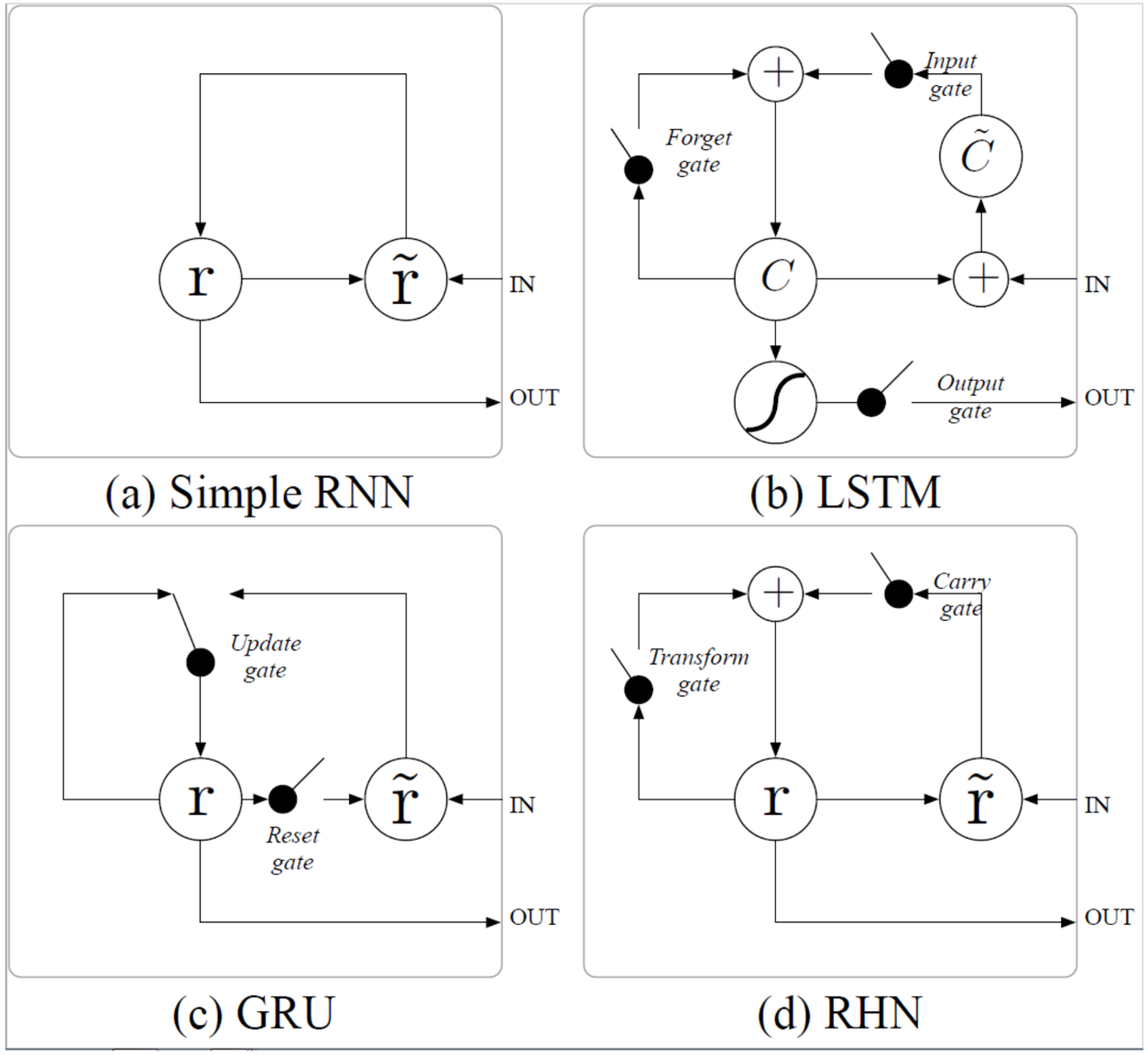
Hidden layer



谁 唱 的 稻香

c1 c2 c3 c4

Recurrent Neural Networks



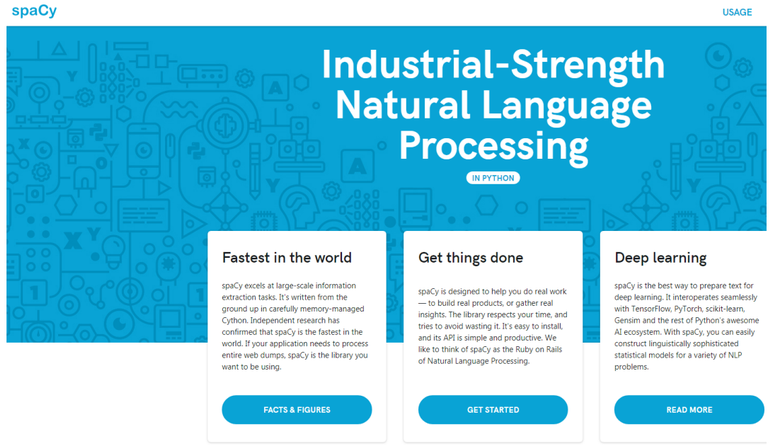
The unreasonable effectiveness of the forget gate

Jos van der Westhuizen, Joan Lasenby

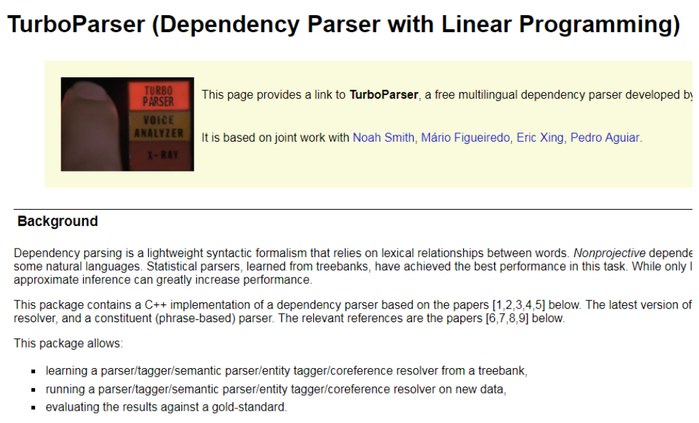
(Submitted on 13 Apr 2018)

Recurrent Neural Networks

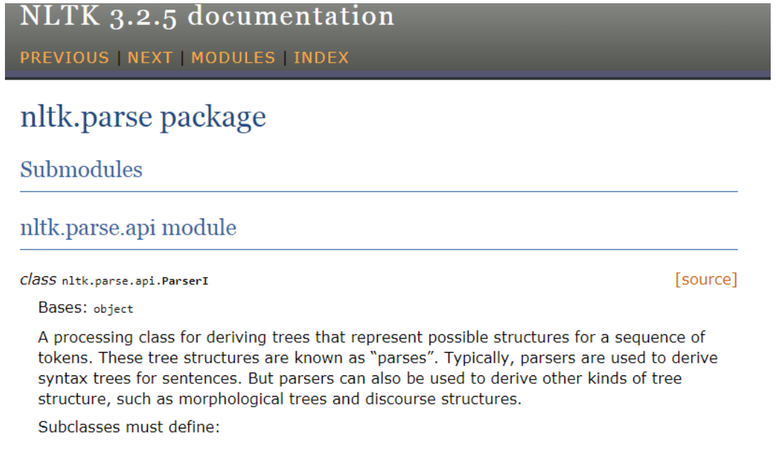
这里，g 还可以依赖于输入句子的parse tree
我们可以将parse tree带入到句子建模中，常用的工具有哪些？



https://spacy.io/



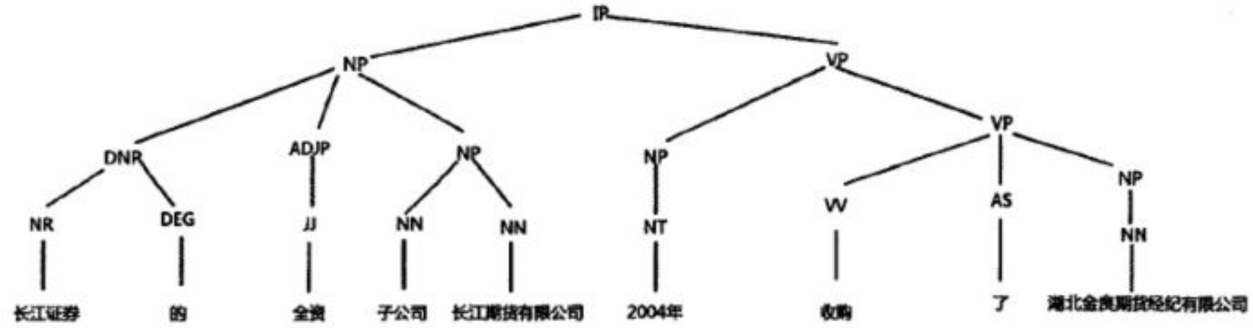
http://www.cs.cmu.edu/~ark/TurboParser/

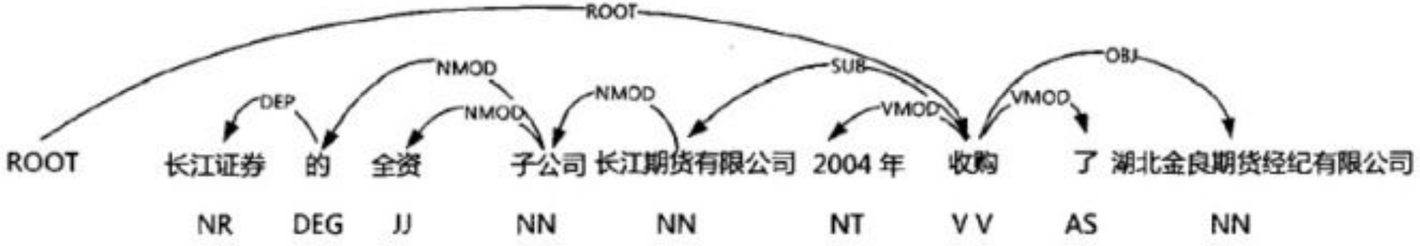


http://www.nltk.org/api/nltk.parse.html

Constituent parsing & dependency parsing

- 句法分析是对句子进行分析非常重要的部分，主要包括constituency parsing(成分句法分析)和dependency parsing(依存句法分析)，两者具有非常大的差异;
- 一个成分解析树将一段文本转化为短语，树中的非叶子结点是短语的类型，而叶子结点是句子中的word，边是没有标记的



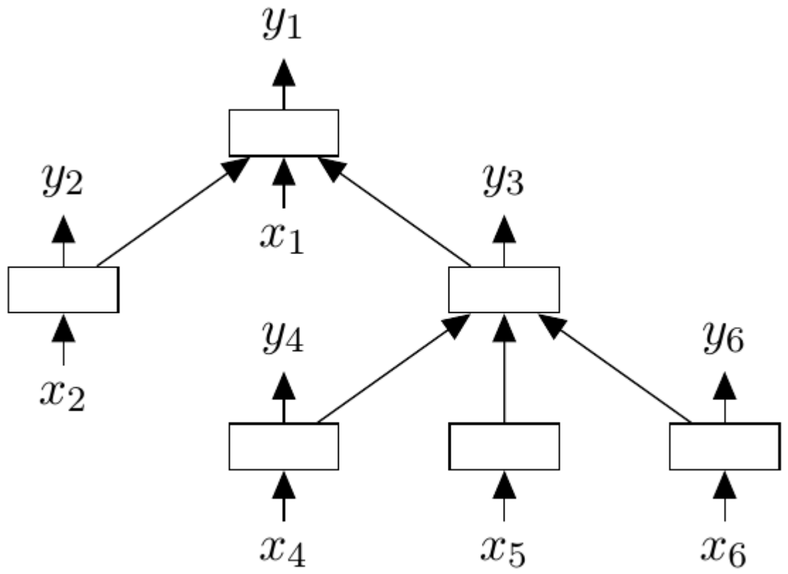


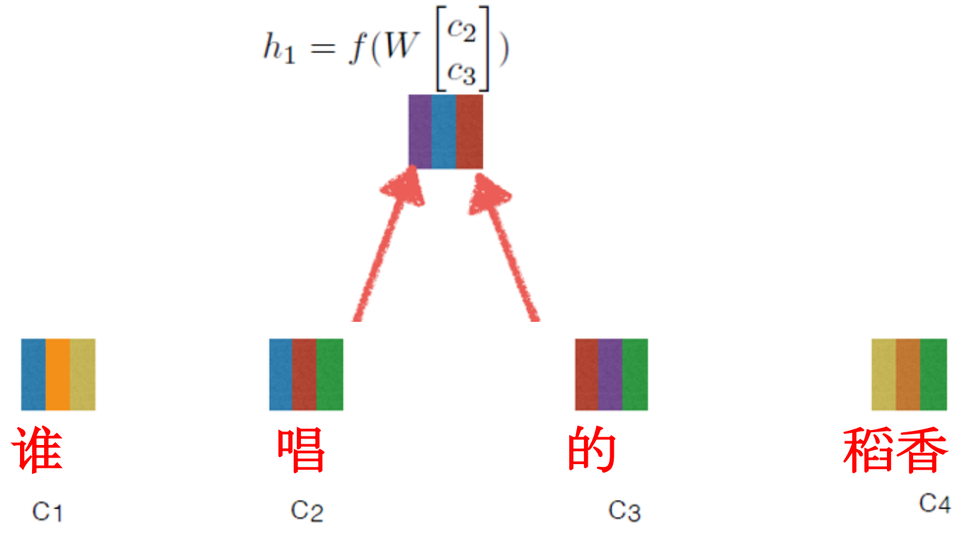
一个依存解析将word按照他们的关系连接起来，每个节点代表一个word，边用关系来进行表示

Recurrent Neural Networks

这里，g 还可以依赖于输入句子的parse tree

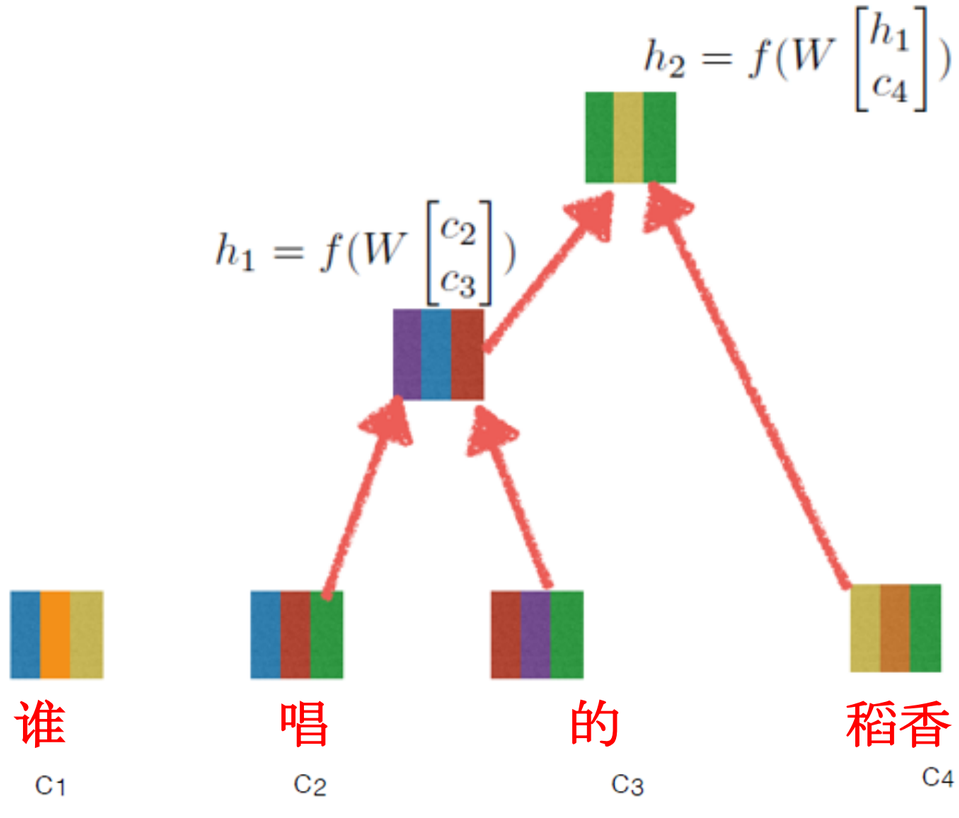
- Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks." arXiv preprint arXiv:1503.00075 (2015).
- Zhang, Xingxing, Liang Lu, and Mirella Lapata. "Top-down tree long short-term memory networks." arXiv preprint arXiv:1511.00060 (2015).





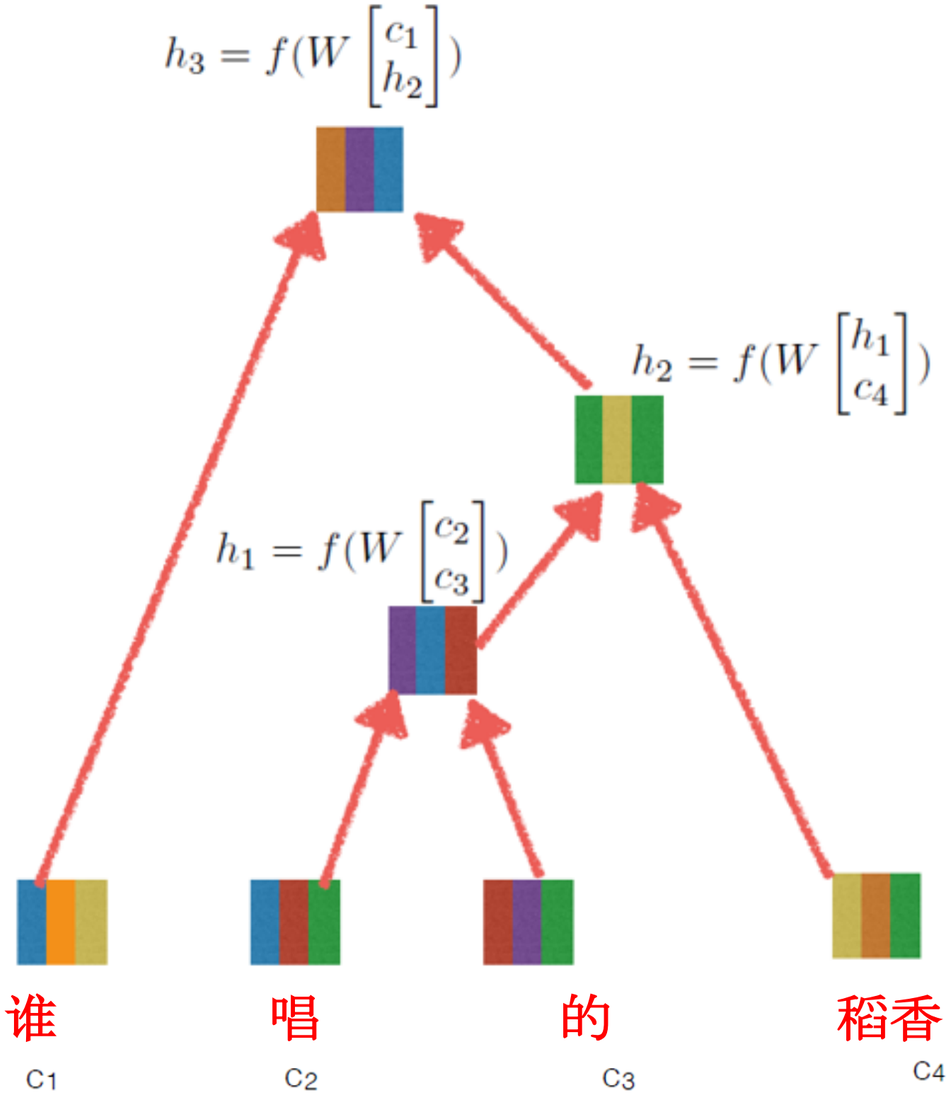
Recurrent Neural Networks

这里，g 还可以依赖于输入句子的parse tree
我们可以将parse tree带入到句子建模中，常用的工具有哪些？



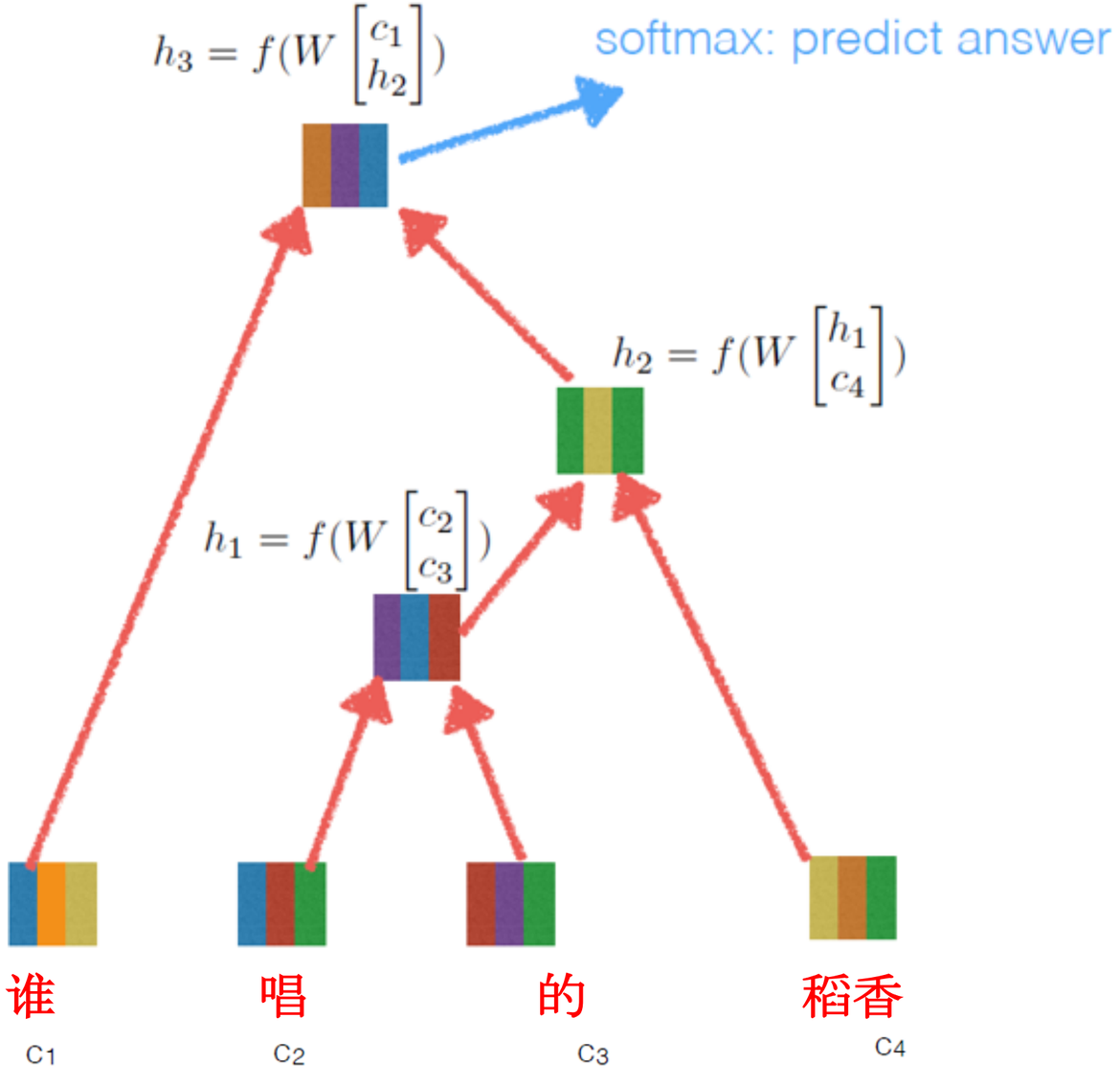
Recurrent Neural Networks

这里，g 还可以依赖于输入句子的parse tree

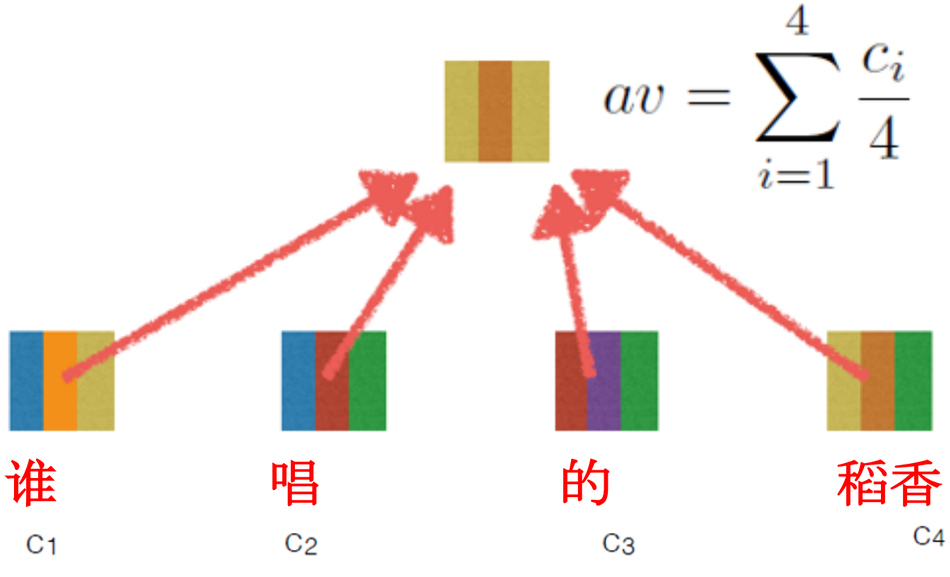


Recurrent Neural Networks

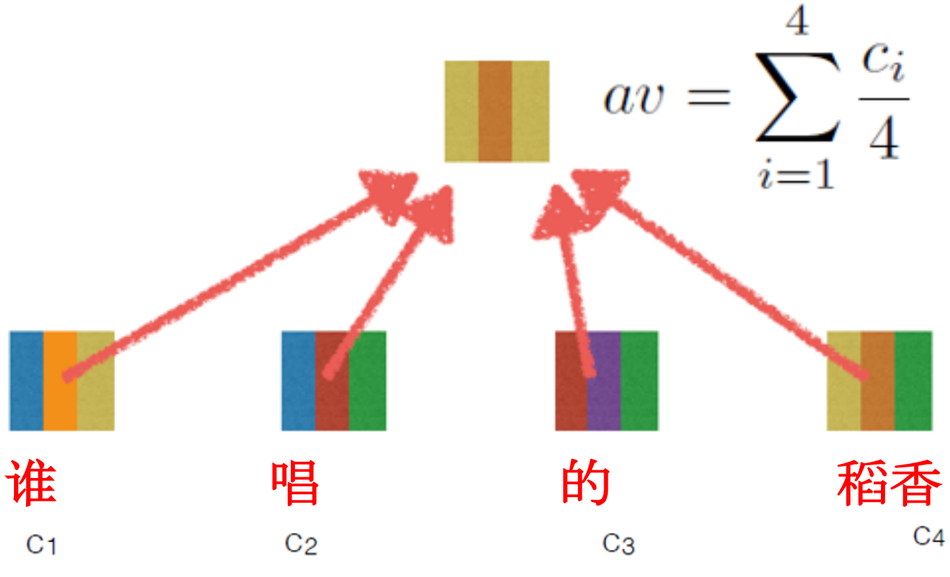
这里，g 还可以依赖于输入句子的parse tree



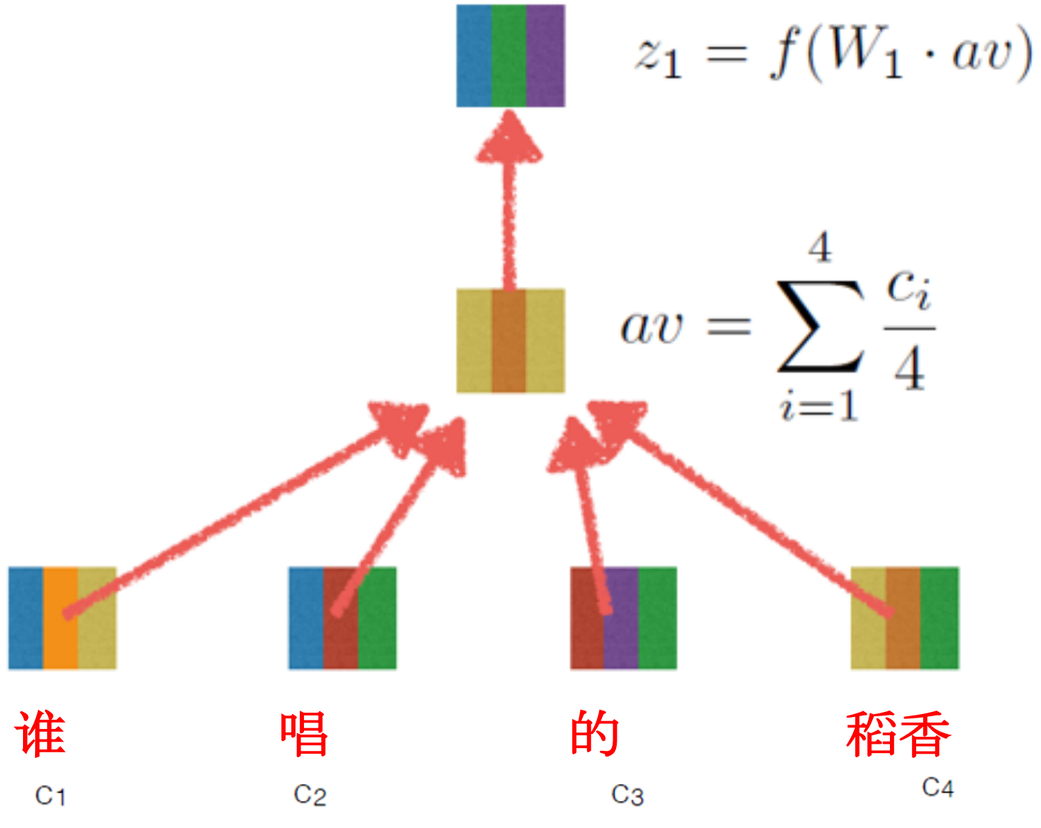
Deep Averaging Networks



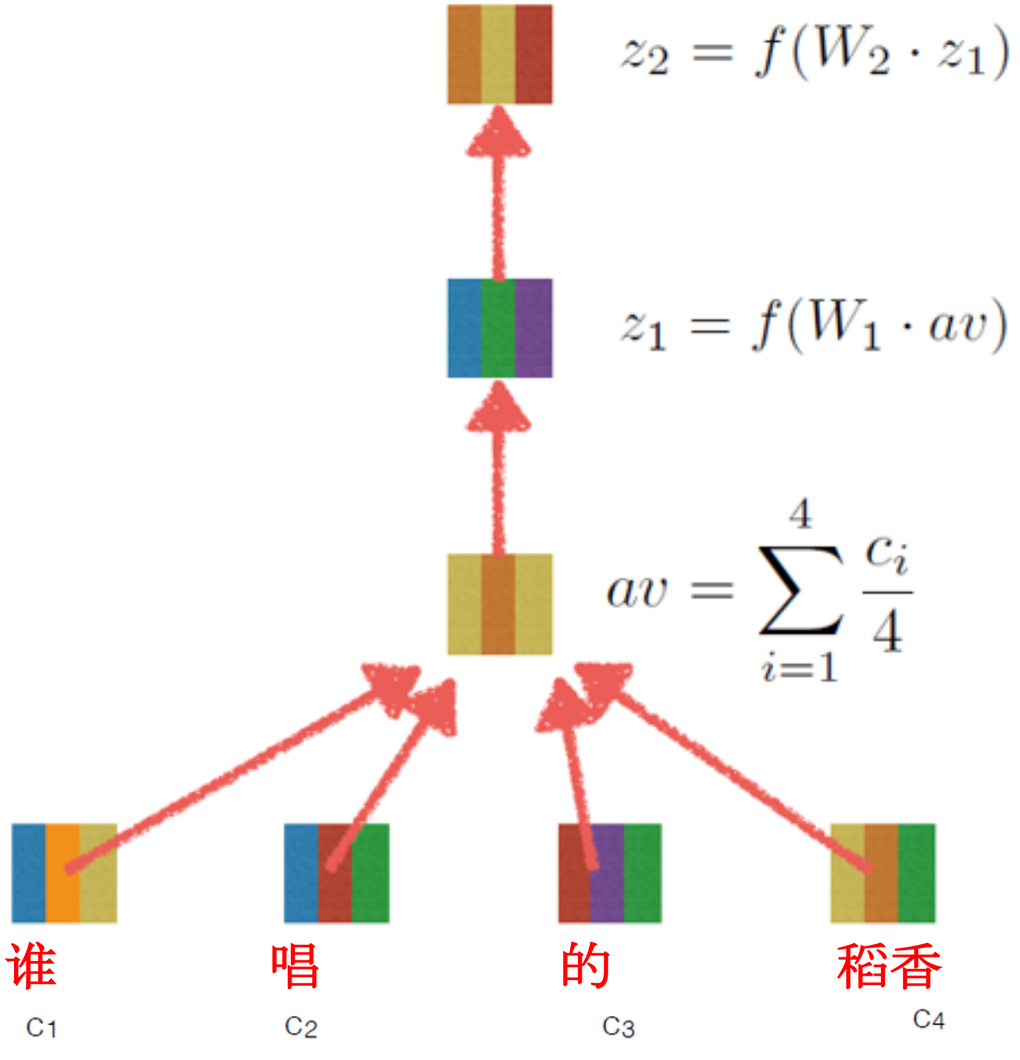
Deep Averaging Networks



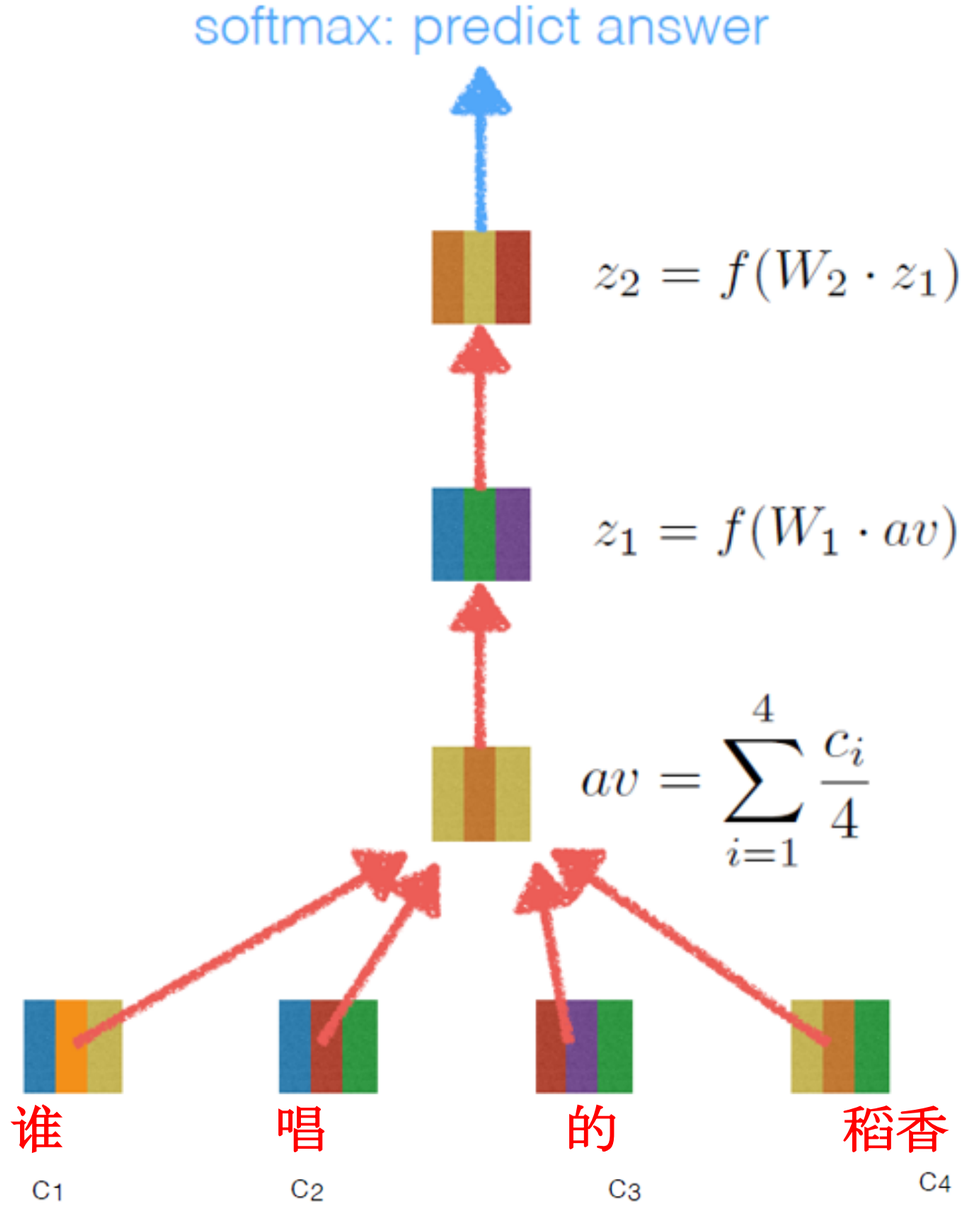
Deep Averaging Networks



Deep Averaging Networks



Deep Averaging Networks



Softmax Answer Classification

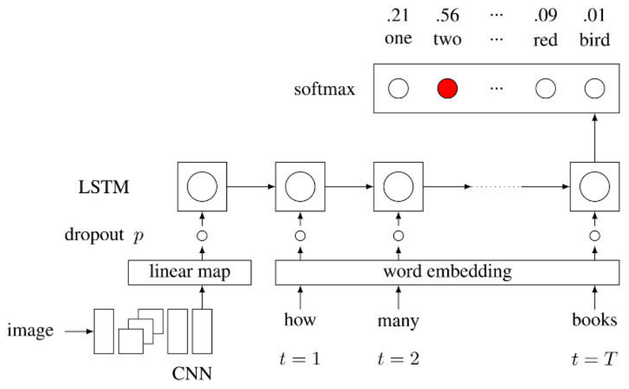
- Multinomial logistic regression

$$\hat{y}_p = \text{softmax}(W_{ans} \cdot h_q)$$

$$\text{softmax}(q) = \frac{\exp q}{\sum_{j=1}^{k} \exp q_j}$$

- Output is a distribution over a finite set of answers (如果single answer, 那么qa问题就变成了分类问题了！)

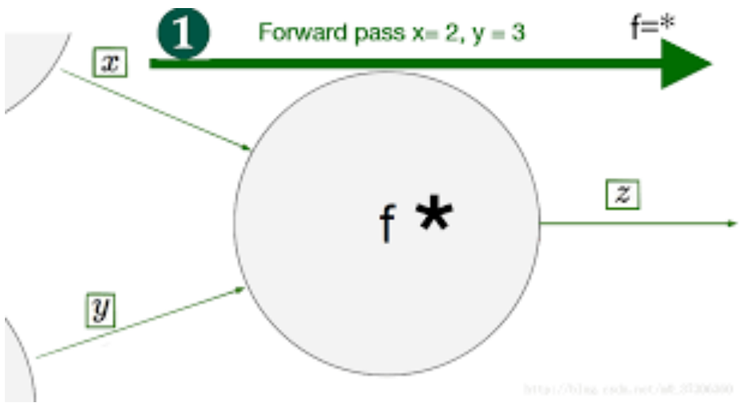
Q: Where are the pink flowers? A: On wall

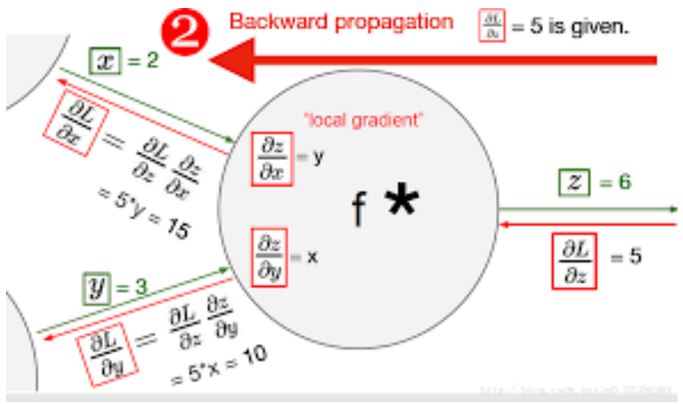


VIS+LSTM model from Ren et al., 2015

如何训练模型？

- 给定输入QA问题-答案对，我们可以用backpropagation进行训练
- In theory, use the chain rule to compute partial derivatives of the error function with respect to every parameter
- In practice, use PyTorch and never have to compute any derivatives by hand!





Naïve Neural Approach

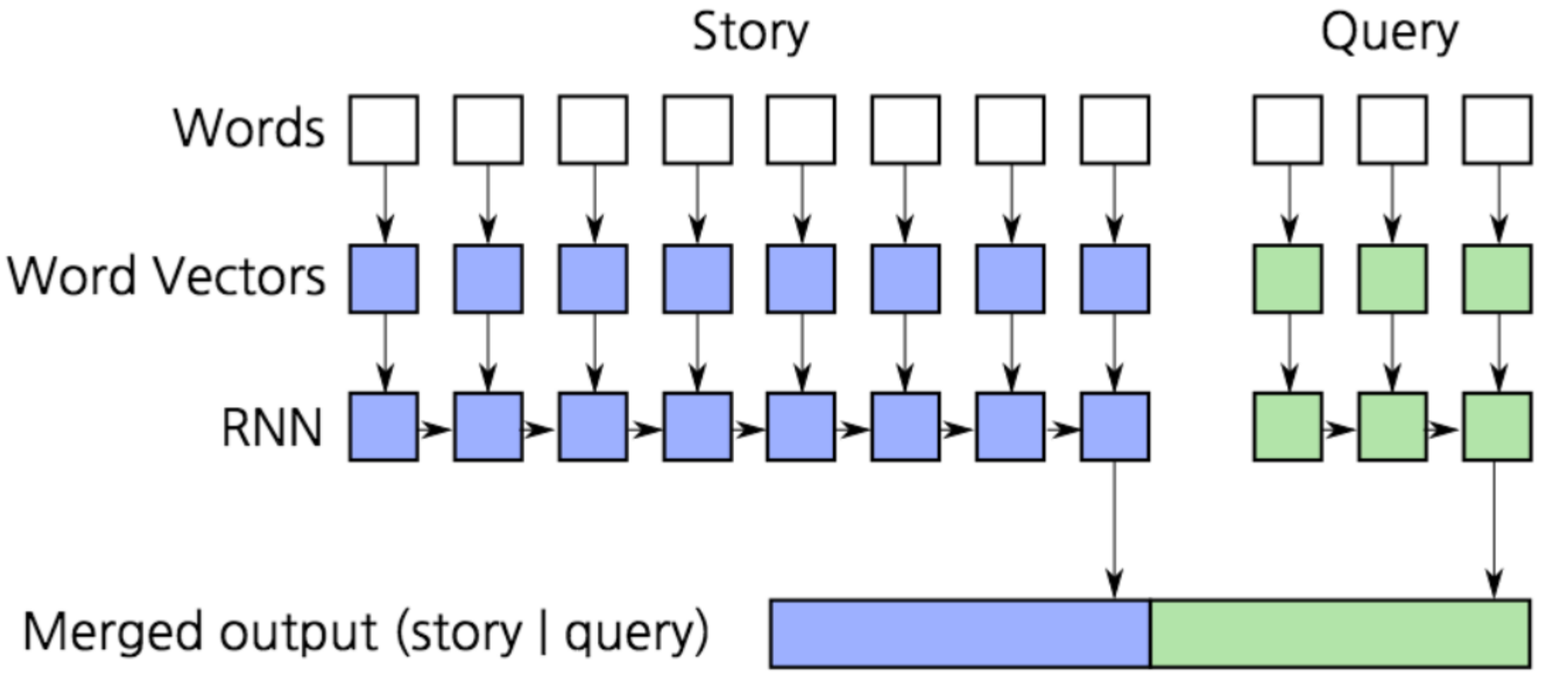


image: http://smerity.com/articles/2015/keras_qa.html

Naïve Neural Approach

https://smerity.com/articles/2015/keras_qa.html

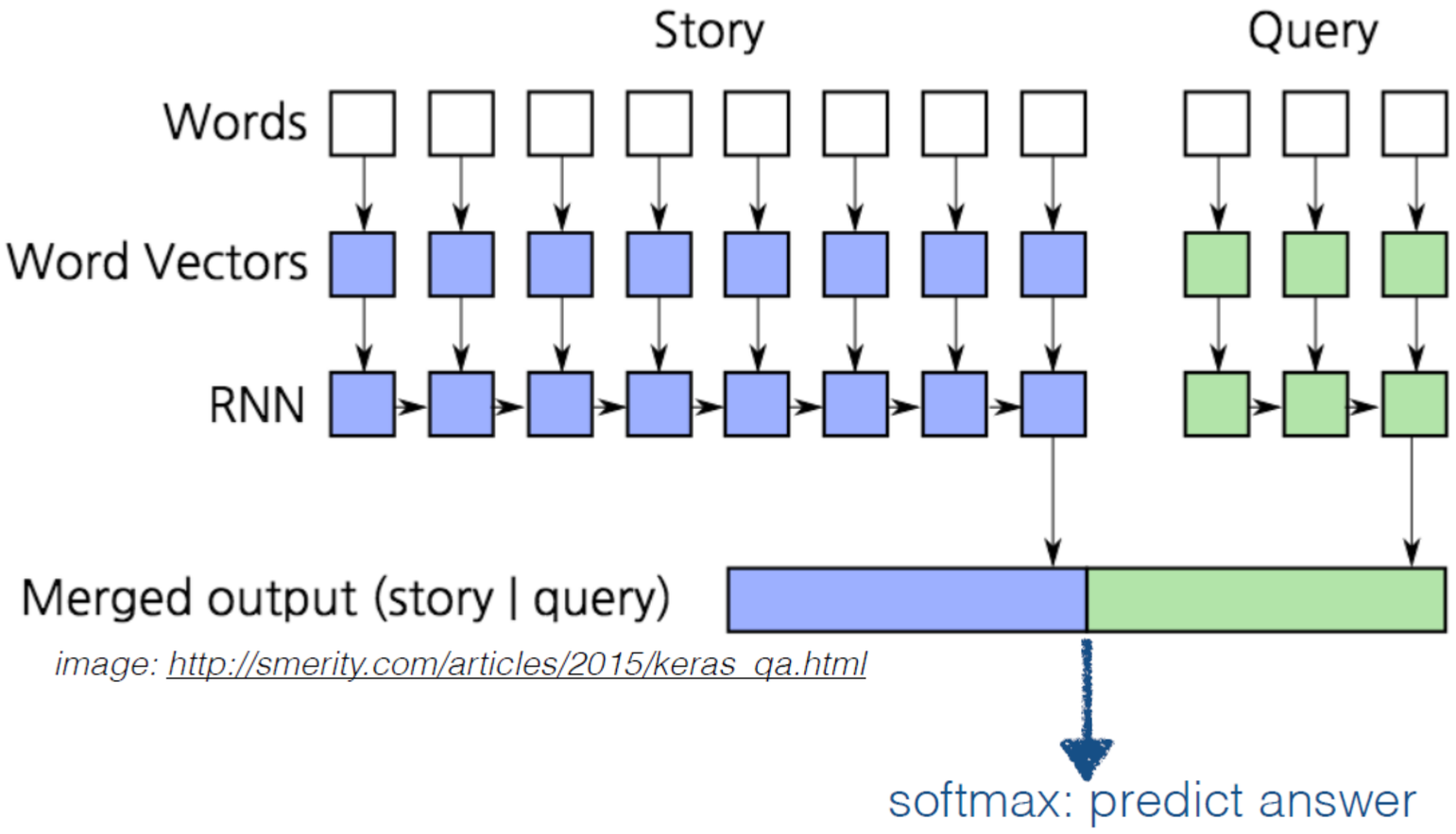


image: http://smerity.com/articles/2015/keras_qa.html

Problems

对于复杂或者很长的问题：

- RNNs/LSTMs 虽然很多改进，但是还是存在long-term dependency的问题
- 当然，可以额外增加外部memory（这里和LSTM的cell区别），这样就可以用来存储重要的信息并且进行推理

Dynamic Memory Networks

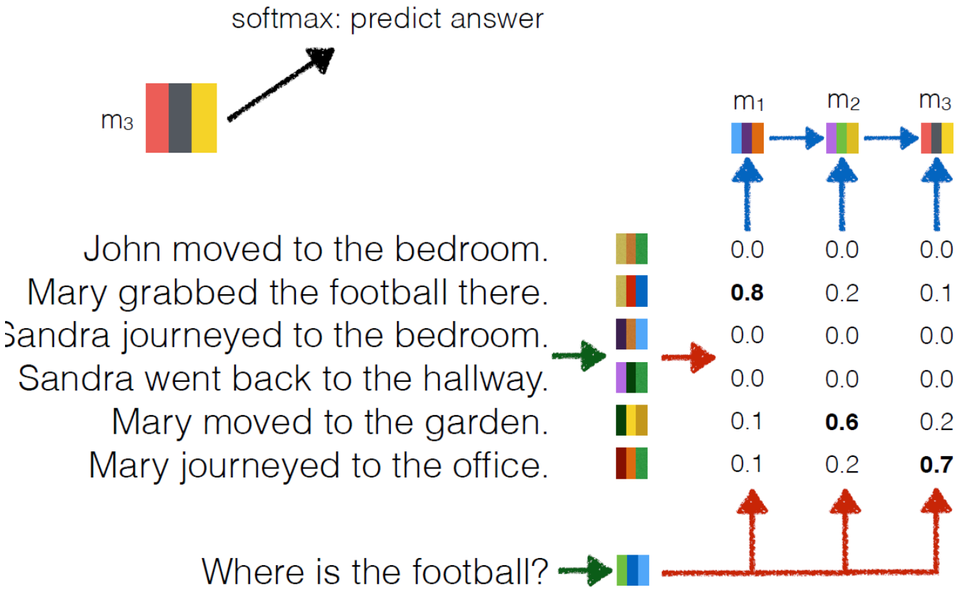
Extends simple RNNs with an *iterative attention mechanism* that focuses on one fact at a time and enables transitive reasoning

1. Compute vector $\mathbf{s_i}$ for every sentence in input and vector **q** for the question using recurrent network **A**
2. Compute an *attention score* $\mathbf{a_i}$ for every sentence

$$a_i = G(s_i, m_{t-1}, q)$$

3. Compute an *episodic memory* $\mathbf{m_t}$ by weighting each $\mathbf{s_i}$ with its corresponding $\mathbf{a_i}$ and passing them through another recurrent network **B**
4. Repeat until network **B** outputs a "finished reading" signal
5. Feed final episodic memory **m** to a softmax layer to predict the answer

Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, and Richard Socher. **Ask Me Anything: Dynamic Memory Networks for Natural Language Processing.** NIPS Deep Learning Symposium, 2015.



Dynamic Memory Networks

Extends simple RNNs with an *iterative attention mechanism* that focuses on one fact at a time and enables transitive reasoning

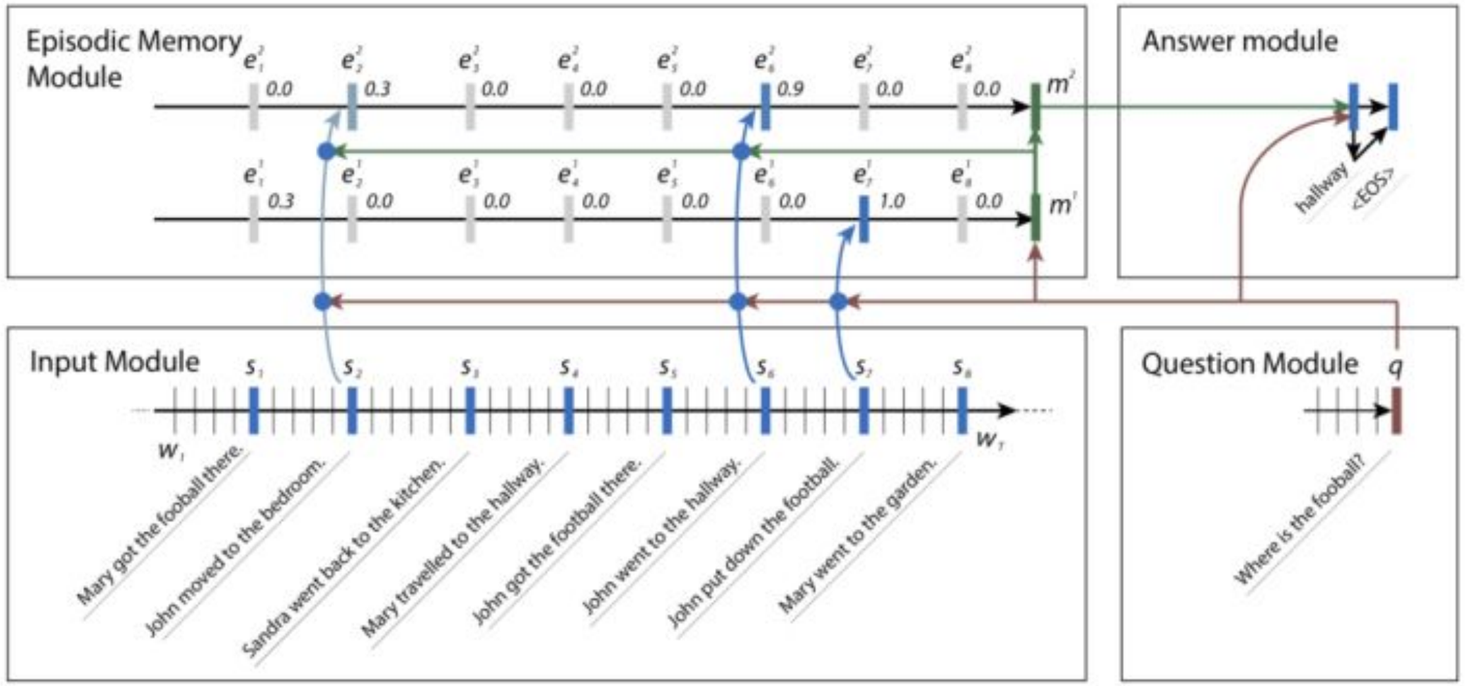


Figure 3. Real example of an input list of sentences and the attention gates that are triggered by a specific question from the bAbI tasks (Weston et al., 2015a). Gate values g_t^i are shown above the corresponding vectors. The gates change with each search over inputs. We do not draw connections for gates that are close to zero. Note that the second iteration has wrongly placed some weight in sentence 2, which makes some intuitive sense, as sentence 2 is another place John had been.

Ask Me Even More: Dynamic Memory Tensor Networks (Extended Model)

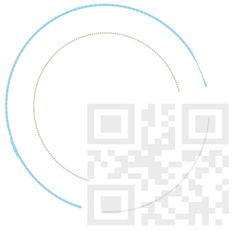
Evaluation: FB bAbi

- 20 very simple tasks (e.g., counting, basic deduction, induction, coreference)
- DMNs solve 18 out of 20 tasks with over 95% accuracy, comparable to other baselines that use
- hand-engineered features (e.g., n-grams, positional features)
- Can also be applied to many other NLP tasks (what is the sentiment of this sentence? what is this sentence's translation in French?)

The bAbI project

This page gather resources related to the bAbI project of Facebook AI Research which is organized towards the goal of automatic text understanding and reasoning. The datasets we have released consist of:

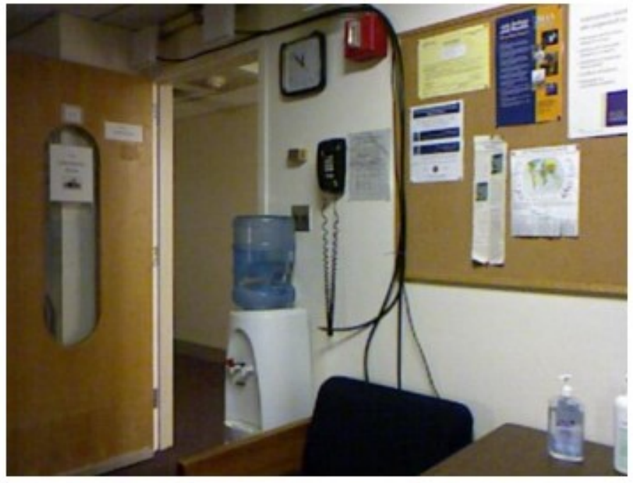
- **The (20) QA bAbI tasks**
- **The (6) dialog bAbI tasks**
- **The Children's Book Test**
- **The Movie Dialog dataset**
- **The WikiMovies dataset**
- **The Dialog-based Language Learning dataset**
- **The SimpleQuestions dataset**
- **HITL Dialogue Simulator**



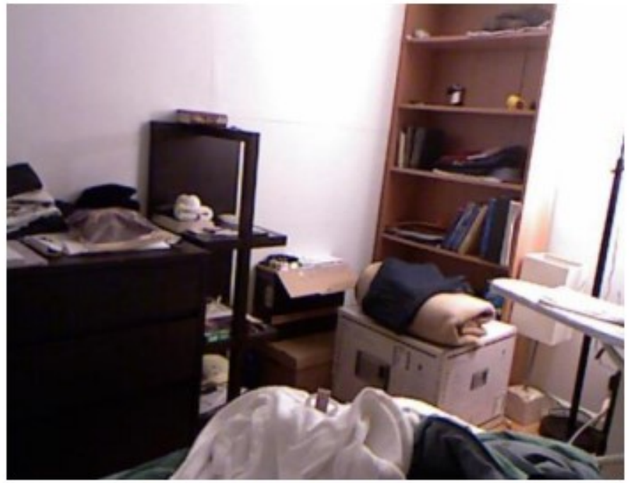
VQA

Visual Question Answering

- Given an image, and a natural language-like question, find the correct answer to it
- Training on a set of triplets (image, question, answer).
- Free-form and open-ended questions.
- Answers can be single word or multiple word.



Question: what is the largest blue object in this picture?
Ground truth: water carboy
Proposed CNN: water carboy



Question: what color is the shade of the table lamp close to the bookshelf?
Ground truth: white
Proposed CNN: white

知识共享新形式 美版知乎Quora加入视频回答功能

互联网 腾讯科技 2017-05-23 10:38 收藏 0 评论 分享

腾讯科技讯 据外媒报道，作为新晋的独角兽公司，美版知乎Quora可不想被锁死在现有的文本问答模式下，它们开始了积极的探索，本周该公司将开始测试视频回答功能，以巩固自己辛苦得来的地位。

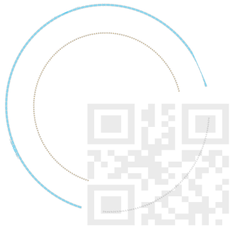
Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

Datasets

- DAQUAR(DAtaset for QUestion Answering on Real-world images) – 1450 images and 12468 questions related to them. On an average 12 words per question.

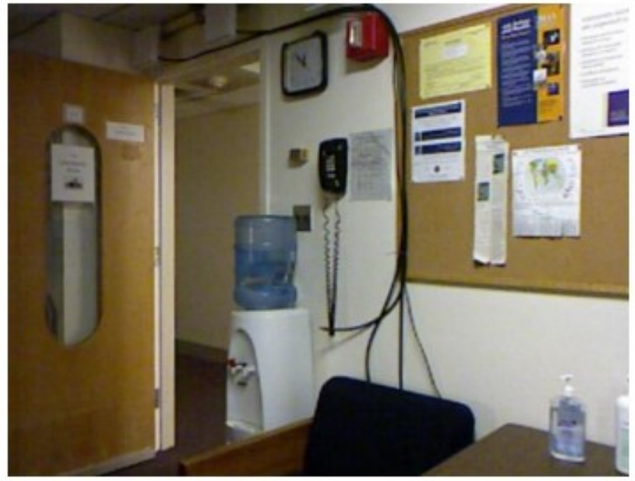
```
1   what is on the left side of the white oven on the floor and on right side of the blue armchair in the image1 ?
2   garbage_bin
3   what is on the left side of the fire extinguisher and on the right side of the chair in the image1 ?
4   table
5   what is between the the two white and black garbage bins in the image1 ?
6   chair
7   how many objects are between the fire extinguisher and the white oven on the floor in the image1 ?
8   3
9   what is the largest object in this picture in the image1 ?
10  washing_machine
```

- VQA(Visual Question Answering) dataset – 254,721 images, 764,163 questions, 9,934,119 answers
- Wu-Palmer Similarity Measure(WUPS score) is used for performance evaluation – Script by Malinowski M.

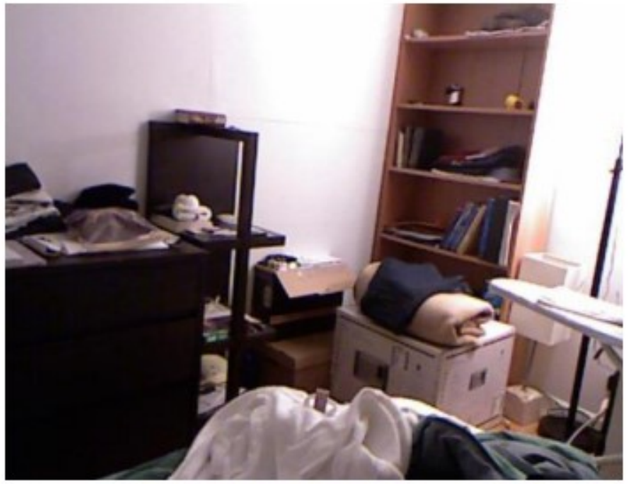


Challenges

- ❑ The output is to be conditioned on both image and language inputs.
- ❑ A better representation of the image content is essential
- ❑ Interactions between the two modalities need to appropriately modelled.



Question: what is the largest blue object in this picture?
Ground truth: water carboy
Proposed CNN: water carboy



Question: what color is the shade of the table lamp close to the bookshelf?
Ground truth: white
Proposed CNN: white

Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

Previous Approaches

- Neural-based approach - image representation from a CNN is fed to each hidden layer of a single LSTM. The LSTM then models the concatenation of question and answer.

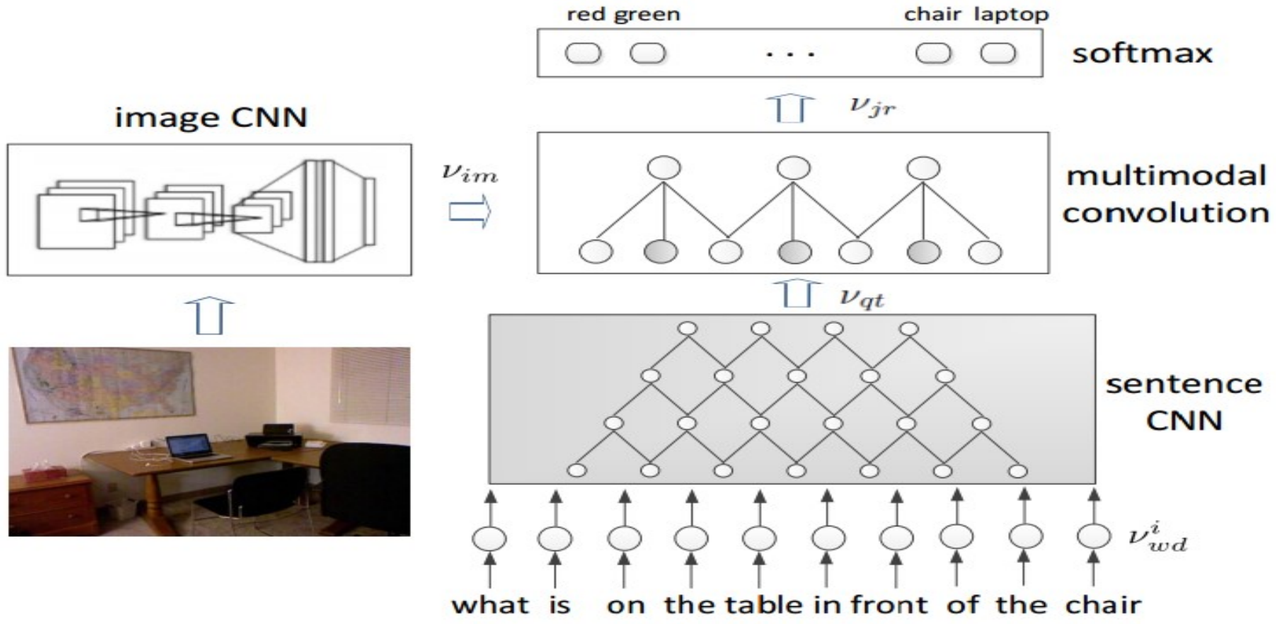
- mQA approach – 4 units - an LSTM to extract the question representation, a CNN to extract the visual representation, an LSTM for storing the linguistic context in an answer, and a fusing component to combine the information from the first three components and generate the answer.

- VIS + LSTM - Here the image is treated as a single word, and the intermediate representation of the input thus obtained is used for classification into the correct class, which is the single word answer.

- CNN approach - uses 3 CNN's - one to extract sentence representation, one for image representation, and the third is a multimodal layer to fuse the two.

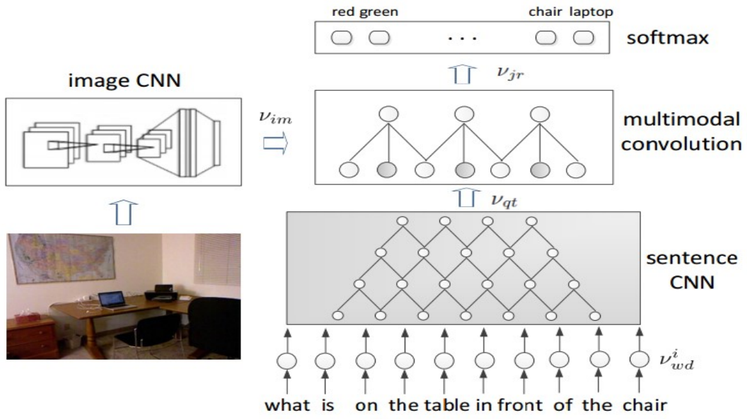
Malinowski et al. 2015
Gao et al. 2015
Kiros et al. 2015
Lin Ma et al. 2015

CNN model



Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

CNN model



E
19 weight layers
conv3-64
conv3-64
conv3-128
conv3-128
conv3-256
conv3-256
conv3-256
conv3-256
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512

$$\nu_{im} = \sigma(\boldsymbol{w_{im}}(CNN_{im}(I)) + b_{im})$$

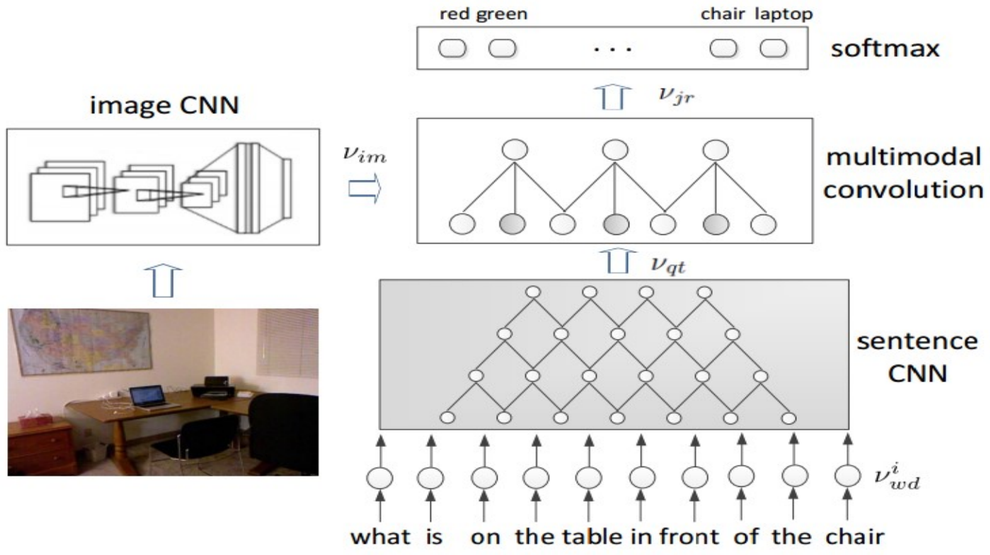
σ: Nonlinear activation function.

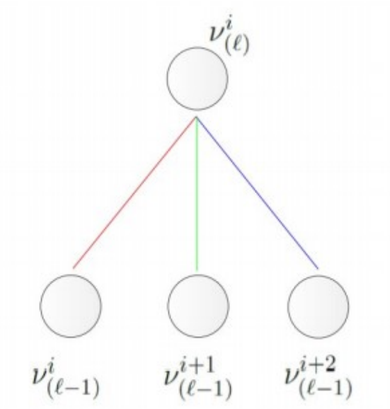
$\boldsymbol{w_{im}}|_{d\times 4096}$: Mapping matrix

CNN_{im} takes image as input and outputs a fixed length vector.

Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

Sentence CNN





1. For sequential input σ, convolution unit for feature map of type f on the l^{th} layer is
$$\nu^i_{(l,f)} \stackrel{\text{def}}{=} \sigma(\boldsymbol{w}_{(l,f)}\, \vec{\nu}^i_{(l-1)} + b_{(l,f)})$$

2. $$\vec{\nu}^i_{(l-1)} \stackrel{\text{def}}{=} \nu^i_{(l-1)} \,||\, \nu^{i+1}_{(l-1)} \,||\, \nu^{i+2}_{l-1}$$

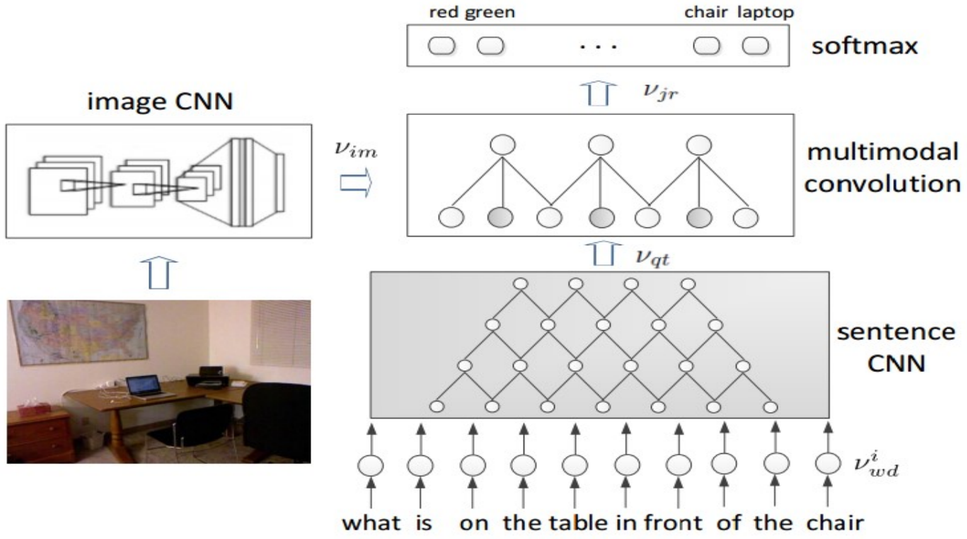
3. $$\vec{\nu}^i_{(0)} \stackrel{\text{def}}{=} \nu^i_{wd} \,||\, \nu^{i+1}_{wd} \,||\, \nu^{i+2}_{wd}$$

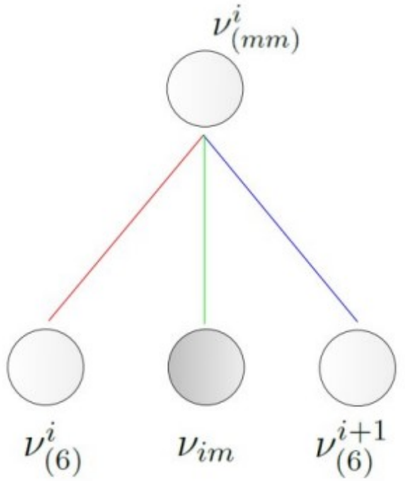
4. Max-pooling after each convolution
$$\nu^i_{(l+1,f)} = max(\nu^{2i}_{(l,f)}, \nu^{2i+1}_{(l,f)})$$

ν^i_{wd} : Skip-gram[6] word embedding of i-th question word

Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

Multimodal Convolutional Layer





input: $\nu_{qt} = [\nu^0_{(6)} ... \nu^n_{(6)}]$

Capturing the interaction between two multimodal inputs

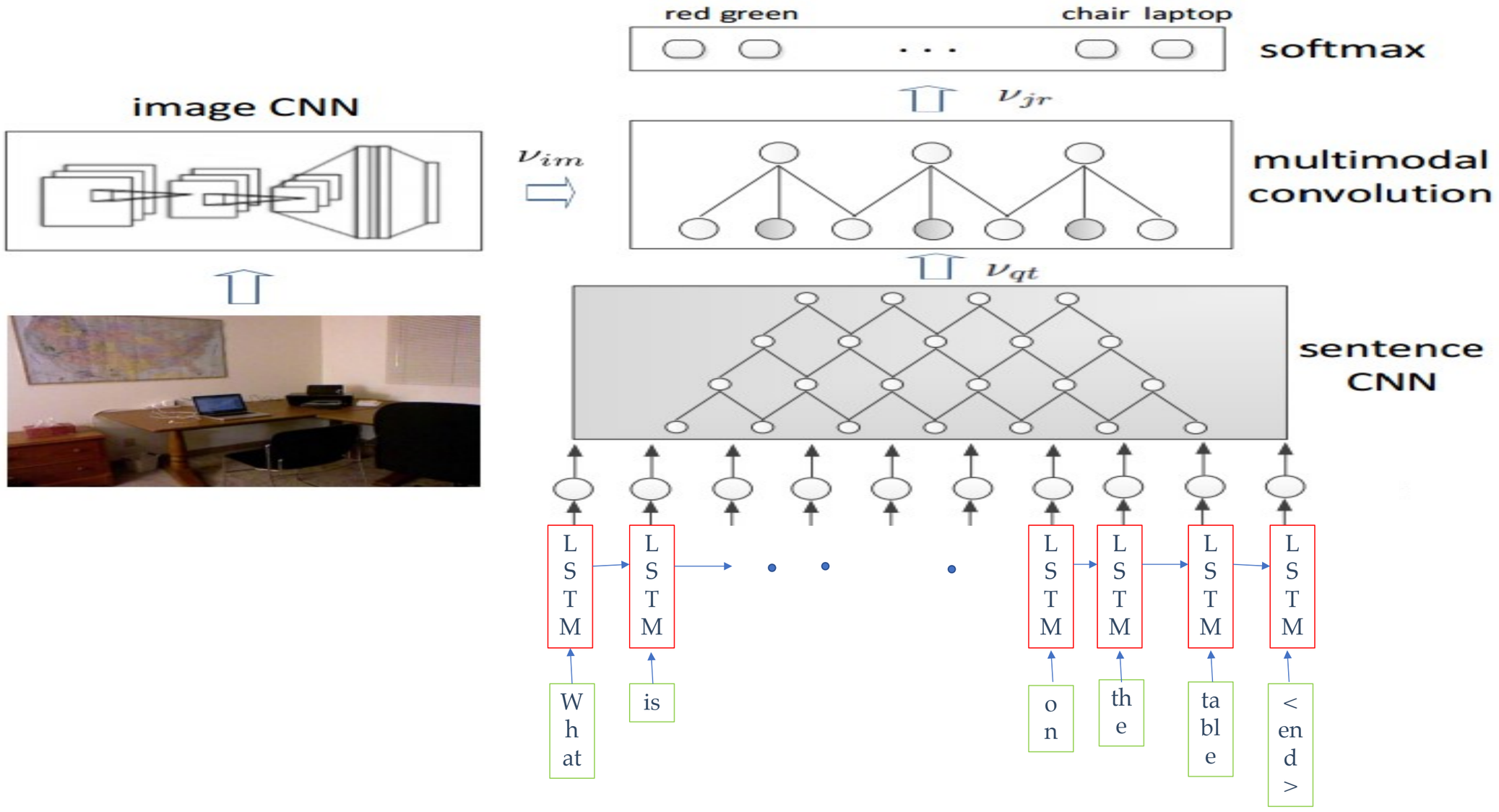
$$\vec{\nu}^i_6 = \nu^i_6 \,||\, \nu_{im} \,||\, \nu^{i+1}_6$$

$$\nu^i_{(mm,f)} = \sigma(\boldsymbol{w}_{(mm,f)} \vec{\nu}^i_{(6)} + b_{(mm,f)})$$

Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.

Proposed Modification

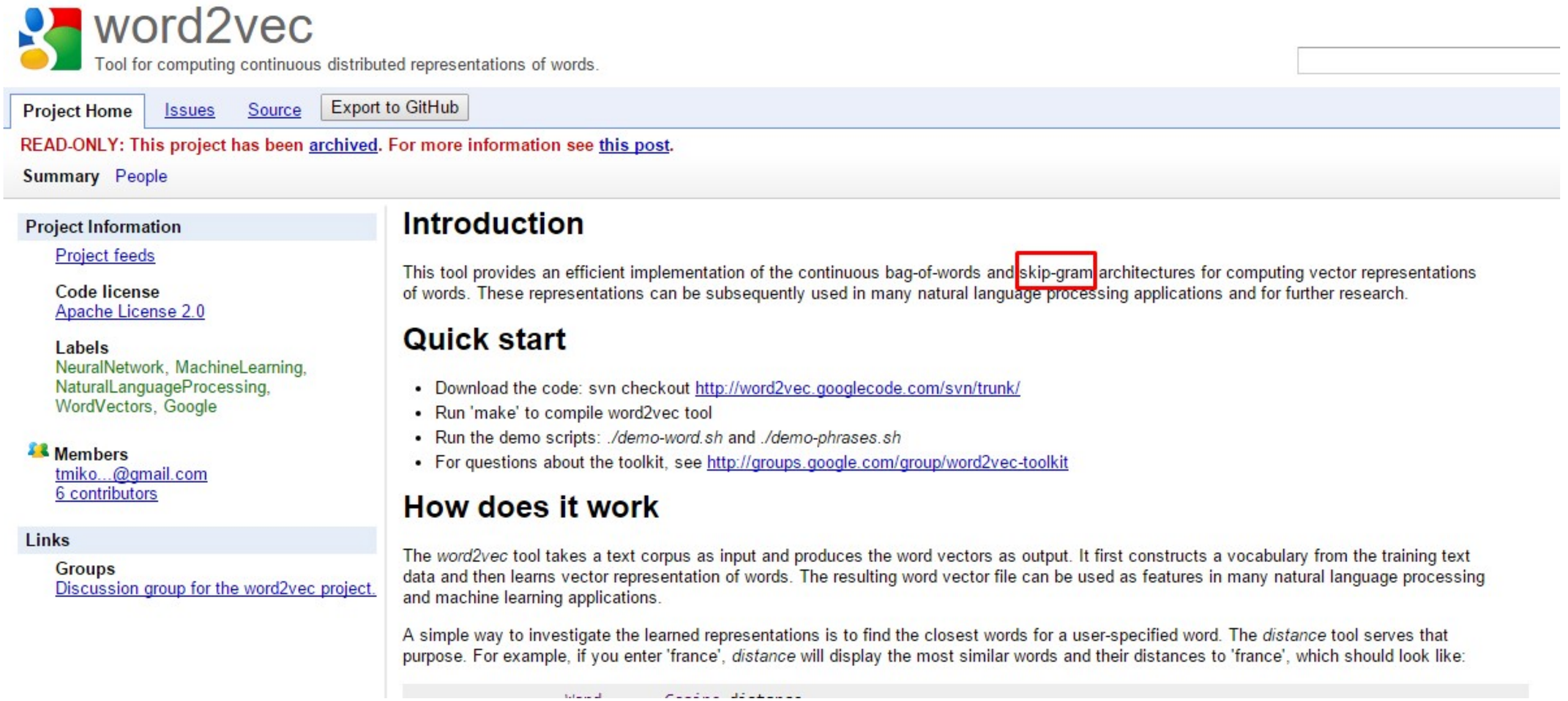
Ma, Lin, Zhengdong Lu, and Hang Li. "Learning to Answer Questions from Image Using Convolutional Neural Network." AAAI. Vol. 3. No. 7. 2016.



Input to LSTM

- Skip-gram word embeddings from the question sentence

Mikolov et al. 2013 https://code.google.com/p/word2vec/



word2vec

Tool for computing continuous distributed representations of words.

Project Home | Issues | Source | Export to GitHub

READ-ONLY: This project has been archived. For more information see this post.

Summary People

Project Information

Project feeds

Code license
Apache License 2.0

Labels
NeuralNetwork, MachineLearning, NaturalLanguageProcessing, WordVectors, Google

Members
tmiko...@gmail.com
6 contributors

Links

Groups
Discussion group for the word2vec project.

Introduction

This tool provides an efficient implementation of the continuous bag-of-words and skip-gram architectures for computing vector representations of words. These representations can be subsequently used in many natural language processing applications and for further research.

Quick start

- Download the code: svn checkout http://word2vec.googlecode.com/svn/trunk/
- Run 'make' to compile word2vec tool
- Run the demo scripts: *./demo-word.sh* and *./demo-phrases.sh*
- For questions about the toolkit, see http://groups.google.com/group/word2vec-toolkit

How does it work

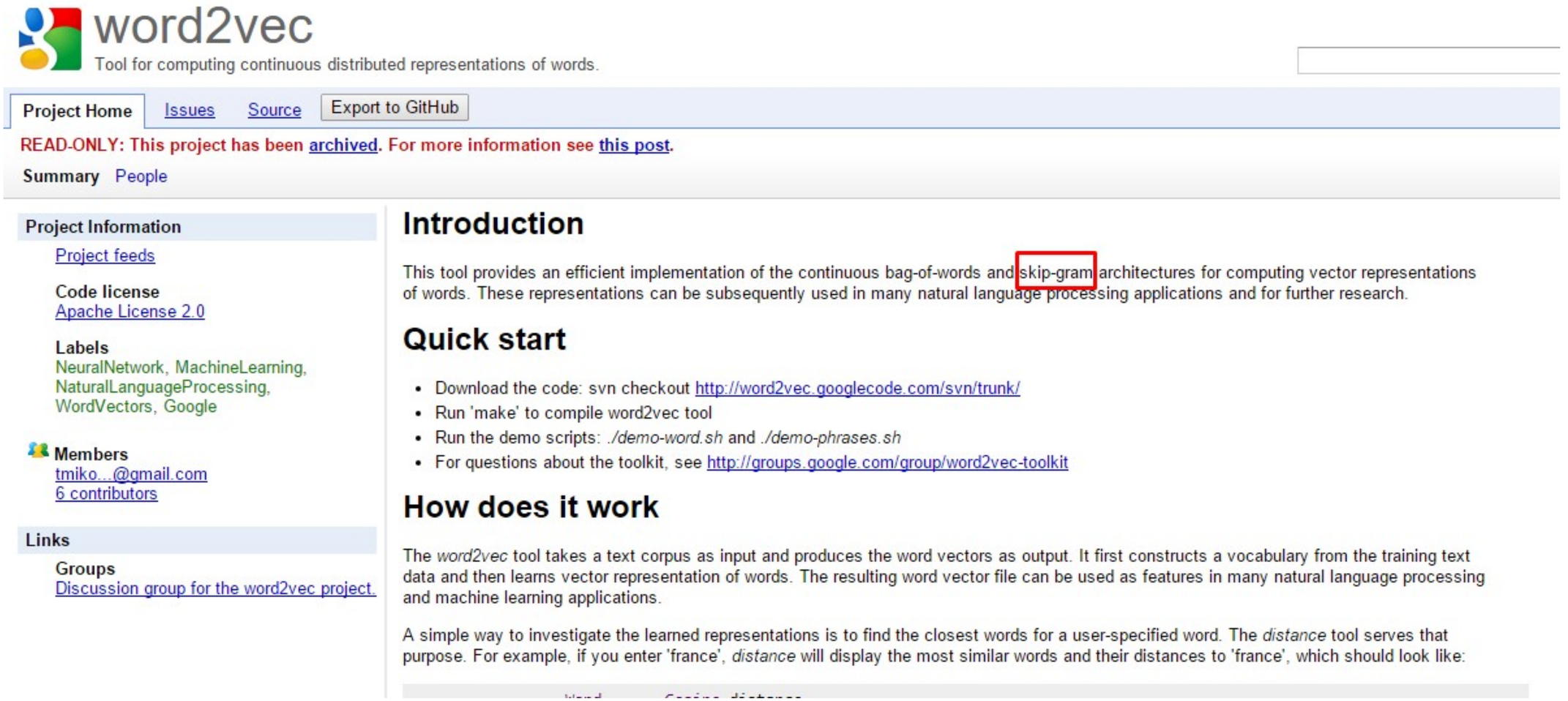
The *word2vec* tool takes a text corpus as input and produces the word vectors as output. It first constructs a vocabulary from the training text data and then learns vector representation of words. The resulting word vector file can be used as features in many natural language processing and machine learning applications.

A simple way to investigate the learned representations is to find the closest words for a user-specified word. The *distance* tool serves that purpose. For example, if you enter 'france', *distance* will display the most similar words and their distances to 'france', which should look like:

Input to LSTM

- Skip-gram word embeddings from the question sentence

Mikolov et al. 2013 https://code.google.com/p/word2vec/



word2vec

Tool for computing continuous distributed representations of words.

Project Home | Issues | Source | Export to GitHub

READ-ONLY: This project has been archived. For more information see this post.

Summary People

Project Information

Project feeds

Code license
Apache License 2.0

Labels
NeuralNetwork, MachineLearning, NaturalLanguageProcessing, WordVectors, Google

Members
tmiko...@gmail.com
6 contributors

Links

Groups
Discussion group for the word2vec project.

Introduction

This tool provides an efficient implementation of the continuous bag-of-words and skip-gram architectures for computing vector representations of words. These representations can be subsequently used in many natural language processing applications and for further research.

Quick start

- Download the code: svn checkout http://word2vec.googlecode.com/svn/trunk/
- Run 'make' to compile word2vec tool
- Run the demo scripts: *./demo-word.sh* and *./demo-phrases.sh*
- For questions about the toolkit, see http://groups.google.com/group/word2vec-toolkit

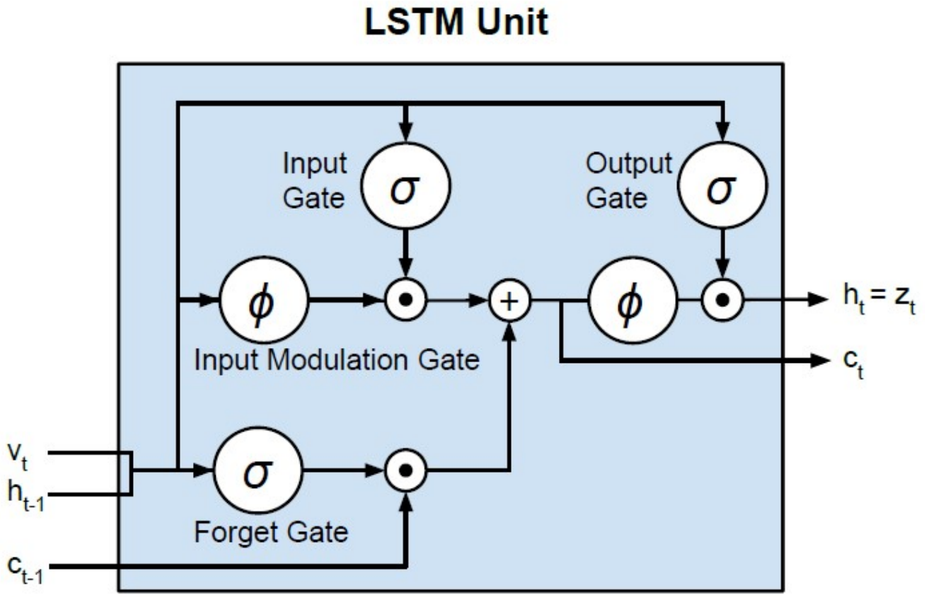
How does it work

The *word2vec* tool takes a text corpus as input and produces the word vectors as output. It first constructs a vocabulary from the training text data and then learns vector representation of words. The resulting word vector file can be used as features in many natural language processing and machine learning applications.

A simple way to investigate the learned representations is to find the closest words for a user-specified word. The *distance* tool serves that purpose. For example, if you enter 'france', *distance* will display the most similar words and their distances to 'france', which should look like:

Input to LSTM

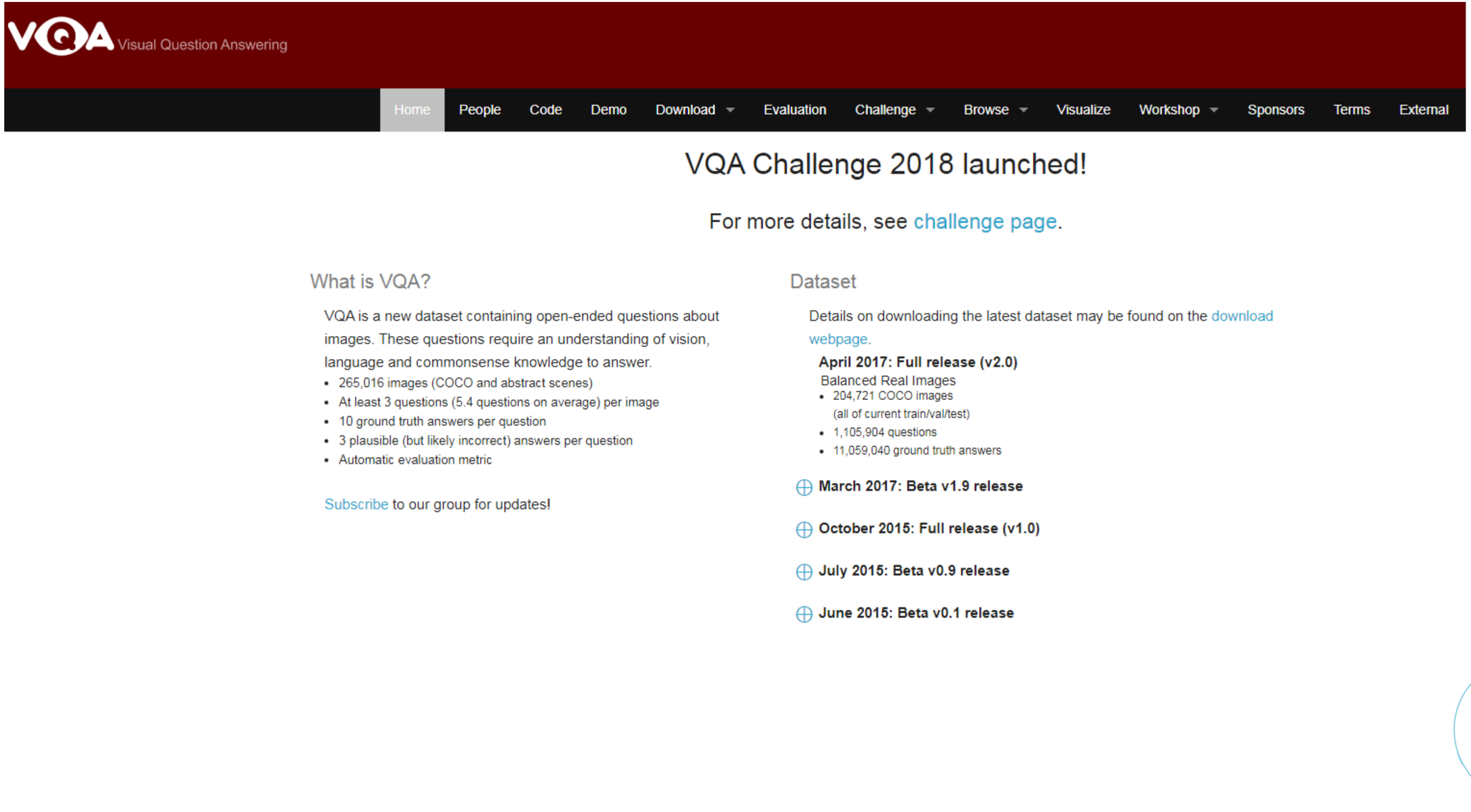
$$
\begin{aligned}
\boldsymbol{i}_t &= \sigma(W_{vi}\boldsymbol{v}_t + W_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \\
\boldsymbol{f}_t &= \sigma(W_{vf}\boldsymbol{v}_t + W_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \\
\boldsymbol{o}_t &= \sigma(W_{vo}\boldsymbol{v}_t + W_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \\
\boldsymbol{g}_t &= \phi(W_{vg}\boldsymbol{v}_t + W_{hg}\boldsymbol{h}_{t-1} + \boldsymbol{b}_g) \\
\boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{g}_t \\
\boldsymbol{h}_t &= \boldsymbol{o}_t \odot \phi(\boldsymbol{c}_t)
\end{aligned}
$$



Application

- Is this truck considered “vintage”?
- Does the road look new?
- What kind of tree is behind the truck?

Visual QA dataset



Application

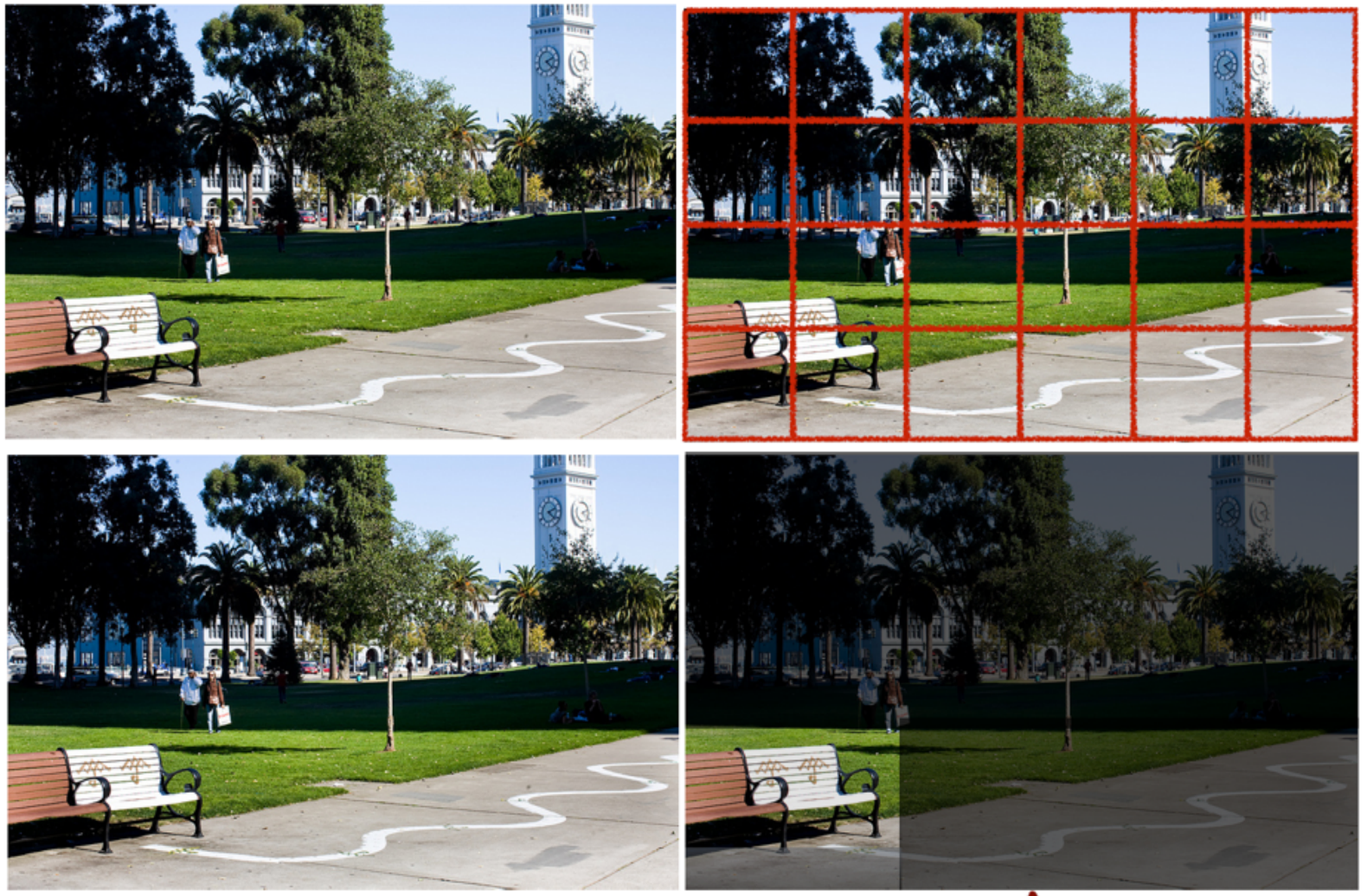
ConvNet () =

- i = *ConvNet*(image) > use an existing network trained for image classification and freeze weights
- q = *RNN*(question) > learn weights
- answer = softmax([i;q])

softmax: predict ‘truck’

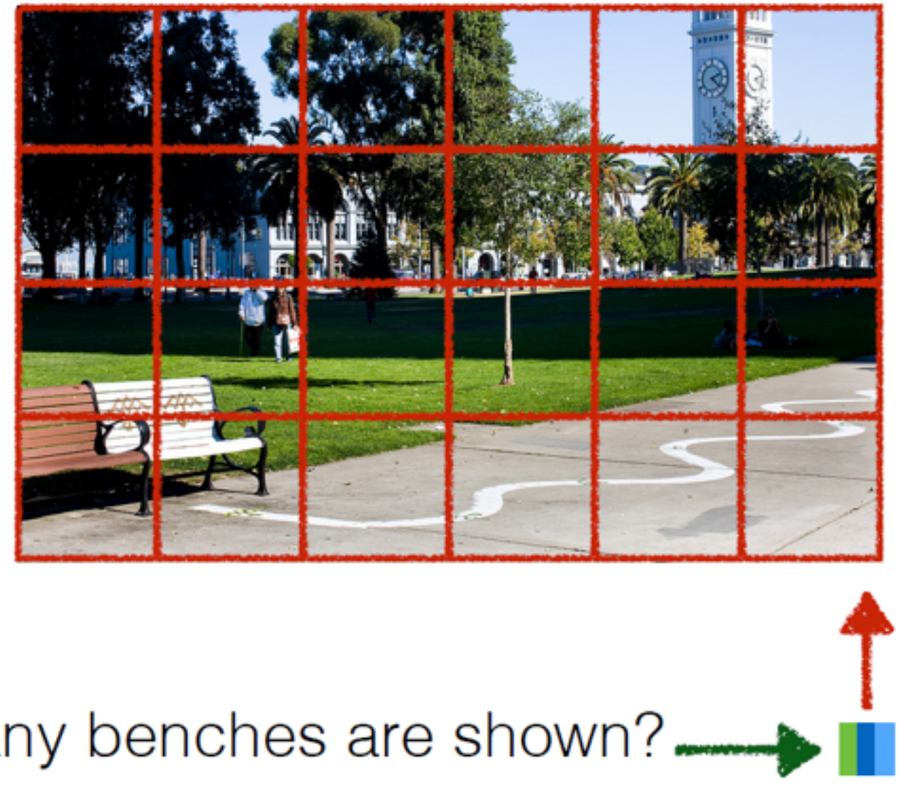
Visual Attention

Use the question representation q to determine where in the image to look



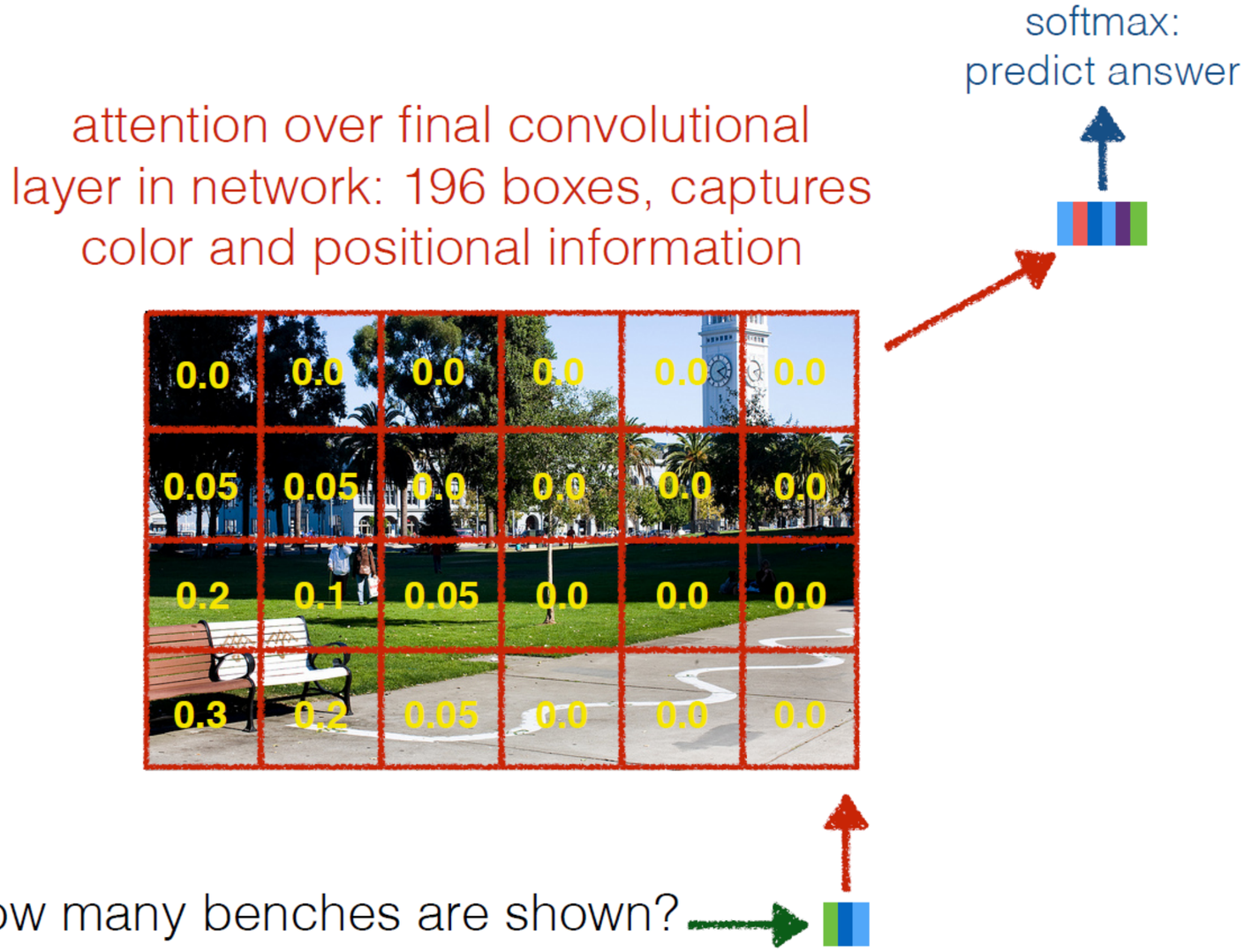
Visual Attention

Use the question representation q to determine where in the image to look



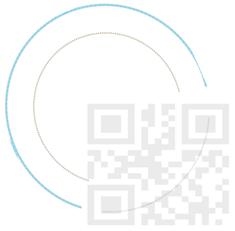
Visual Attention

Use the question representation q to determine where in the image to look



Issues

- Visual attention is more complicated than textual attention; requires many more QA pairs than are currently available
- Focusing on more than one “box” at a time is difficult for the current model; perhaps an iterative attention mechanism like the DMN’s can solve the problem



END