Application(s) of Deep Learning

Pavel Krömer[1]

2017 CZ-AT Summer School on Deep Learning and Visual Data Analysis

[1]Dept. of Computer Science,
VŠB - Technical University of Ostrava,
Ostrava, Czech Republic
pavel.kromer@vsb.cz

Introduction

Introduction

I think people need to understand that deep learning is making a lot of things, behind–the–scenes, much better. Deep learning is already working in Google search, and in image search; it allows you to image search a term like "hug".

– Geoffrey Hinton

Deep Learning in Information Retrieval

Information retrieval

The area of information retrieval (IR) is a branch of computer science dealing with storage, maintenance, and searching in large amounts of data.

An information retrieval system (IRS) is a software tool for data representation, storage and information searching.

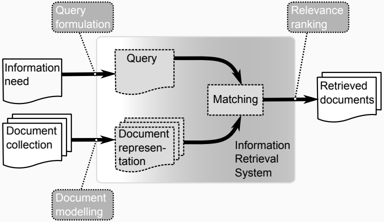
Information retrieval

The area of information retrieval (IR) is a branch of computer science dealing with storage, maintenance, and searching in large amounts of data.

An information retrieval system (IRS) is a software tool for data representation, storage and information searching.

Main tasks of an IRS include

- document modelling (indexing)
- query processing
- document – query matching
- result (relevance) ranking



Information retrieval model (cont.)

An IR model is a formal background defining the internal document representation, query language, and a document – query matching mechanism. It determines the document indexing procedure, result ordering, and other aspects of an IRS.

IR models are based e.g. on

- set theory and Boolean logic
- geometry of high-dimensional vector spaces
- probability theory
- feature–based approaches

Boolean IR model

Boolean IR model is a traditional model based on set theory, Boolean logic, and exact document – query match principle.

A document, $\boldsymbol{d}_i$, is represented as a set of indexed terms. An index of n documents with m terms is defined as

$$\boldsymbol{d}_i = (t_{i1}, t_{i2}, ..., t_{im}),\ \forall\ t_{i,j} \in \{0, 1\}, \qquad \boldsymbol{D} = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nm} \end{pmatrix} \tag{1}$$

Queries in the Boolean model are Boolean logic formulas.

The document – query matching is based on the exact match principle and divides documents into two disjunct subsets of retrieved and non-retrieved documents.

Vector space model

Vector space model (VSM) interprets documents and queries as points in a multidimensional space

$$\boldsymbol{d}_i = (t_{i1}, t_{i2}, \ldots, t_{im}),\ \forall\, t_{ij} \in \mathbb{R}, \qquad \boldsymbol{D} = \begin{pmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nm} \end{pmatrix} \tag{2}$$

Indexing involves assignment of weights with respect to importance of term inside a document and in context of the whole collection. Salton's $tf \cdot idf_t$ is a popular automated indexing method

$$ntf_{ij} = \frac{tf_{ij}}{\max_k(tf_{ik})}, \quad idf_i = \log\left(\frac{N}{N_i}\right), \quad t_{ij} = ntf_{ij} \cdot idf_i \tag{3}$$

A query $\boldsymbol{q} = (t_{q1}, t_{q2}, \ldots, t_{qm})$ is interpreted as a vector of searched terms. The document – query matching is based on the best match principle and evaluated using some similarity measure.

Feature–based Information Retrieval

Feature–based IR models represent documents as vectors of feature functions' values.

Feature–based Information Retrieval

Feature–based IR models represent documents as vectors of feature functions' values.

Feature functions are arbitrary functions of document and query.

Feature–based Information Retrieval

Feature–based IR models represent documents as vectors of feature functions' values.

Feature functions are arbitrary functions of document and query.

Ranking functions that combine the values of feature functions into a single relevance score are essential for good overall IR performance.

Deep Learning and Information Retrieval

Deep learning is in information retrieval most often used for extracting semantically meaningful features (of different kinds) for subsequent document ranking stages.

Selected applications of deep learning in IR

- Semantic hashing with deep autoencoders for document indexing and retrieval
- Deep-structured semantic modeling (DSSM) for document retrieval
- Deep stacking networks for information retrieval

IR application: Semantic hashing

A deep belief network is used to transform a word–count–based document representation into a hash (signature) that represents it well.

Hashing

- an IR strategy for providing efficient access to information based on a key
- uses a hash function to map documents from document space into a hash (signature) space
- both similar and different requirements than in security domain (e.g. order–preserving hashing)

IR application: Semantic hashing (cont.)

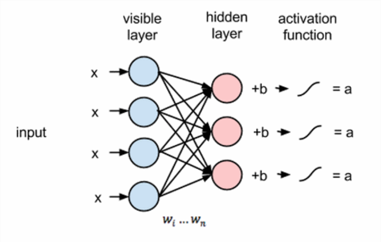
Restricted Boltzmann machines

- 2–layer neural net with visible (input) layer and hidden layer
- layers composed of nodes connected across layers (but not inside)
- each node makes a stochastic decision whether to transmit an input or not

IR application: Semantic hashing (cont.)

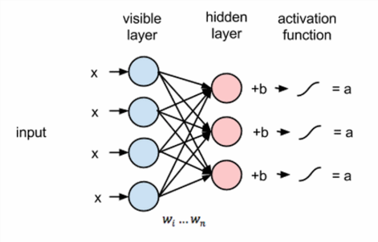
Restricted Boltzmann machines

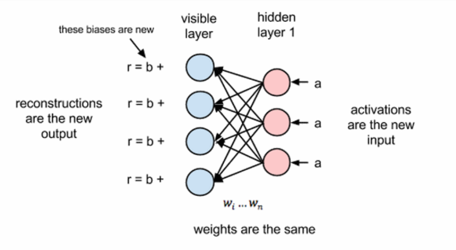
- 2–layer neural net with visible (input) layer and hidden layer
- layers composed of nodes connected across layers (but not inside)
- each node makes a stochastic decision whether to transmit an input or not



Restricted Boltzmann machines

- 2–layer neural net with visible (input) layer and hidden layer
- layers composed of nodes connected across layers (but not inside)
- each node makes a stochastic decision whether to transmit an input or not





IR application: Semantic hashing (cont.)

Deep belief network

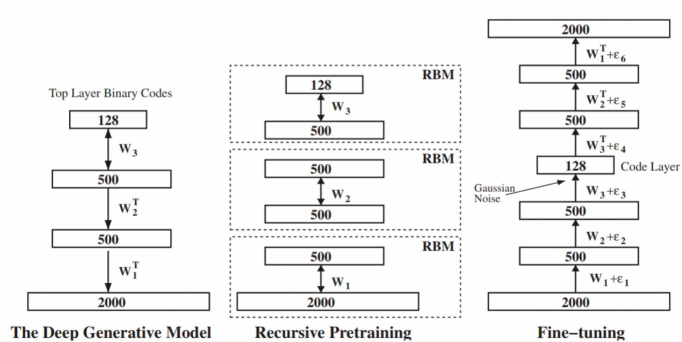
- probabilistic generative model composed of multiple layers of stochastic, hidden (latent) variables
- the top two layers have undirected, symmetric connections between them
- the lower layers receive top–down, directed connections from the layer above
- nodes are Restricted Boltzmann Machines

IR application: Semantic hashing (cont.)

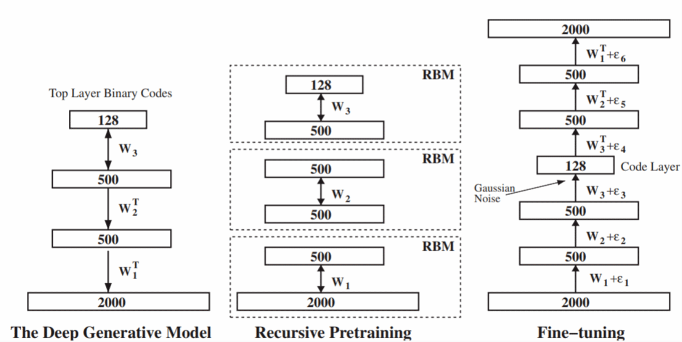
DBN for hashing

- composed of an undirected associative memory (top 2 layers) + deep belief network with directed top–down connections composed of stacked RBMs
- lowest layer takes a word–count representation of the document and the top layer shows a binary signature (hash, code) associated with the input = ENCODER
- RBMs in the opposite direction form a network that maps the hash to word–count representation = DECODER
- ENCODER + DECODER = deep AUTOENCODER for document coding and retrieval

IR application: Semantic hashing (cont.)



IR application: Semantic hashing (cont.)



Retrieval process

- create a hash (128 bit) of query (forward pass with thresholding)
- compute hamming distance between query hash and document hashes
- explore the semantic neighborhood of the query

IR application: Deep-structured semantic modeling

A special DNN, DSSM, is used to capture the semantic properties of query and documents in a way similar to latent semantic indexing and to rank the documents.

IR application: Deep-structured semantic modeling

A special DNN, DSSM, is used to capture the semantic properties of query and documents in a way similar to latent semantic indexing and to rank the documents.

It maps high–dimensional sparse text features into low–dimensional dense semantic features (in a semantic space).

IR application: Deep-structured semantic modeling

A special DNN, DSSM, is used to capture the semantic properties of query and documents in a way similar to latent semantic indexing and to rank the documents.

It maps high–dimensional sparse text features into low–dimensional dense semantic features (in a semantic space).

The model is trained in a supervised manner using clickthrough data recorded on the Web.

IR application: Deep-structured semantic modeling

A special DNN, DSSM, is used to capture the semantic properties of query and documents in a way similar to latent semantic indexing and to rank the documents.

It maps high–dimensional sparse text features into low–dimensional dense semantic features (in a semantic space).

The model is trained in a supervised manner using clickthrough data recorded on the Web.

The relevance is computed as cosine similarity between the query and docs in the semantic space.

$$cos(\boldsymbol{q}, \boldsymbol{d}) = \frac{\boldsymbol{q} \cdot \boldsymbol{d}}{||\boldsymbol{q}|| \cdot ||\boldsymbol{d}||}$$

Convolutional neural networks

A kind of NN designed for data with known grid–like topology (time series, images).

Convolutional neural networks

A kind of NN designed for data with known grid–like topology (time series, images).

Uses convolution (convolution–like operation) instead of general matrix multiplication in at least one of their layers.

Convolutional neural networks

A kind of NN designed for data with known grid–like topology (time series, images).

Uses convolution (convolution–like operation) instead of general matrix multiplication in at least one of their layers.

Most popular type of Deep NN.

Convolutional neural networks (cont.)

Convolution

An operation on two real–valued functions that results in a weighted average of measurements. Prefers (gives more weight to) recent measurements.

$$s(t) = \int x(a)w(t-a)da$$

$$s(t) = (x * w)(t)$$

where w is a pdf, x is an input, w is a kernel, and the output is called a feature map.

Convolutional neural networks (cont.)

Convolution

An operation on two real–valued functions that results in a weighted average of measurements. Prefers (gives more weight to) recent measurements.

$$s(t) = \int x(a)w(t-a)da$$

$$s(t) = (x * w)(t)$$

where w is a pdf, x is an input, w is a kernel, and the output is called a feature map.

In practice, a discrete convolution is used

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a)$$

Convolutional neural networks (cont.)

Convolution causes

- sparse interactions
- parameter sharing
- equivariant representations
- allows use of inputs of variable size

Convolutional neural networks (cont.)

Convolution causes

- sparse interactions
- parameter sharing
- equivariant representations
- allows use of inputs of variable size

Sparse interactions

Also known as sparse connectivity, means that each input affects only a small number of outputs.

The network still can detect important features (e.g. lines in an image) and is suitable for learning hierarchies of (relatively simple) concepts.

Sparse interactions reduce memory complexity (fewer parameters, orders of magnitude lower number of weights).

Convolutional neural networks (cont.)

Convolution causes

- sparse interactions
- parameter sharing
- equivariant representations
- allows the use of inputs of variable size

Parameter sharing

Is achieved by using the same parameter for more than one function in the model.

Each member of the kernel is used at every position of the input so only a single set of parameters is learned.

Convolutional neural networks (cont.)

Convolution causes

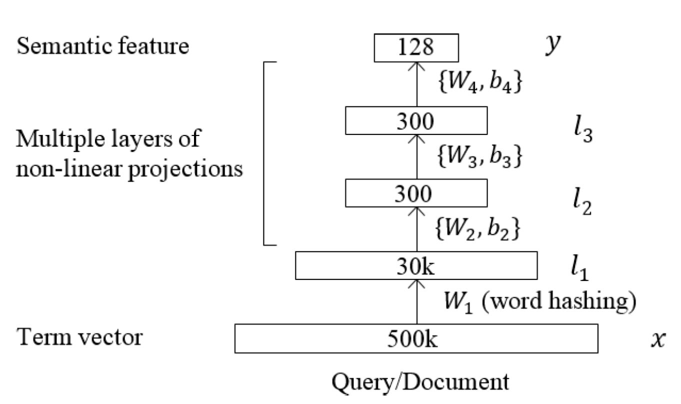
- sparse interactions
- parameter sharing
- equivariant representations
- allows use of inputs of variable size

Equivariance

If input changes, output changes in the same (similar) way.

Example: if a transformation (e.g. translation) is applied to the input, it is equivalent to applying the same transformation to output.

IR application: Deep-structured semantic modeling (cont.)



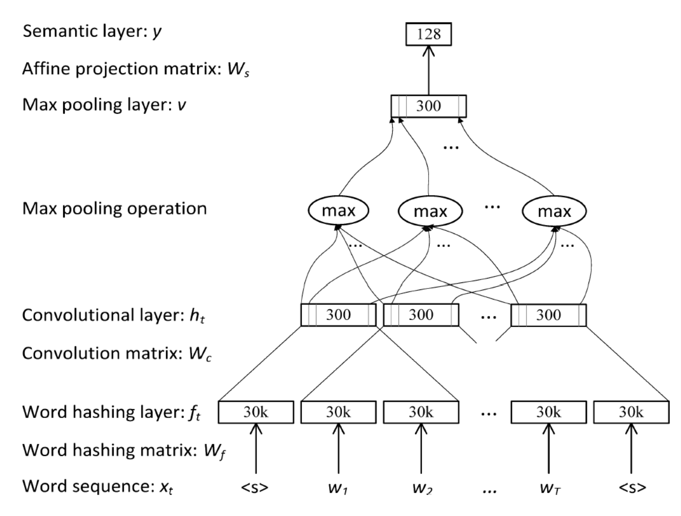
IR application: Deep-structured semantic modeling (cont.)

DSSM was later extended by convolutional layer to enable dealing with word sequences and contextual structures.

The convolutional layer allows mapping variable–length word sequences to a low–dimensional vector in the latent semantic space.

Documents and queries are treated as sequences instead of the traditional bag–of–words approach.

IR application: Deep-structured semantic modeling (cont.)



IR appl.: Information retrieval by deep stacking networks

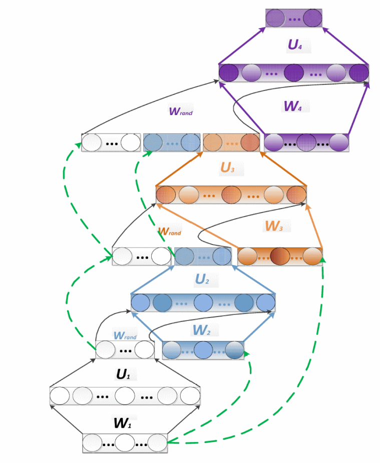
Deep stacking networks

- deep model equipped with parallel and scalable learning
- simple modules (functions, classifiers) are composed first and then stacked on top of each other (to learn complex classification strategies).
- simple training procedure (optimizing MSE)
- connection weights between hidden units and linear output units learned by convex optimization

Deep stacking networks for IR

- stacked network used to provide a binary decision about document relevance
- results correlate very well with IR quality measures

IR appl.: Information retrieval by deep stacking networks (cont.)



Deep Learning and Image Colorization

Grayscale Image Colorization

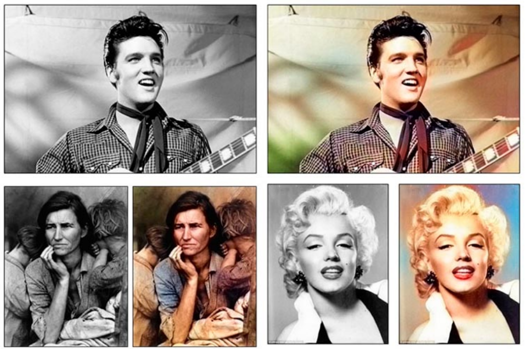


Image colorization is a traditional and appealing image processing task with dozens of applications (incl. restoration), however, it is ill–posed.

Traditional Image Colorization

Manual colorization is rather straightforward and depends on the ability of the user to identify all parts of the image, their context, and color.

Interactive colorization requires user input whereas non–interactive does not.

Interactive methods require the user to define the color of some areas and they are later on extended into the full image (segmentation vs. color–flow function).

Automated colorization is also often cast as a search problem: similar reference images are sought and their colors are transfered to target image.

Traditional Image Colorization (cont.)

Scribble–based colorization requires manually specifying desired colors of certain regions. The scribble colors are propagated under the assumption that adjacent pixels with similar luminance should have similar color, with the optimization relying on Normalized Cuts.

Transfer–based methods rely on the availability of related reference image(s), from which color is transferred to the target grayscale image. Mapping between source and target is established automatically, using correspondences between local descriptors, or in combination with manual intervention.

Deep Learning–based Image Colorization

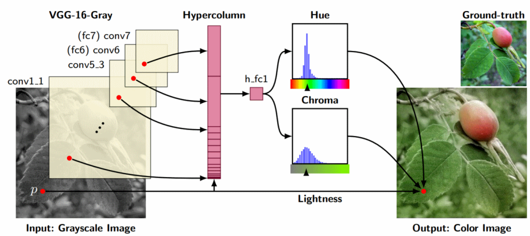
Deep convolutional architecture for image colorization

- Grayscale image processed by a deep convolutional network
- identifies spatially localized hypercolumns (multilayer slices) as per–pixed descriptors
- trained to predict hue and chroma distributions for each pixel from its hypercolumn

Deep Learning–based Image Colorization

Deep convolutional architecture for image colorization

- Grayscale image processed by a deep convolutional network
- identifies spatially localized hypercolumns (multilayer slices) as per–pixed descriptors
- trained to predict hue and chroma distributions for each pixel from its hypercolumn



Questions?

Questions?

Thank you for the 2017 Czech-Austrian Summer School on ”Deep Learning and Visual Data Analysis”