Transfer Learning for Biomedical Relation Extraction

Introduction

Mario Sänger (WBI, HU Berlin)

saengema@informatik.hu-berlin.de

Outline

- Introduction
  - Natural Language Processing (NLP)
  - Relation Extraction (RE)
  - Transfer Learning (TL)

- Organization
  - General information
  - Requirements
  - Timeline

Introduction

- Natural Language Processing (NLP) -

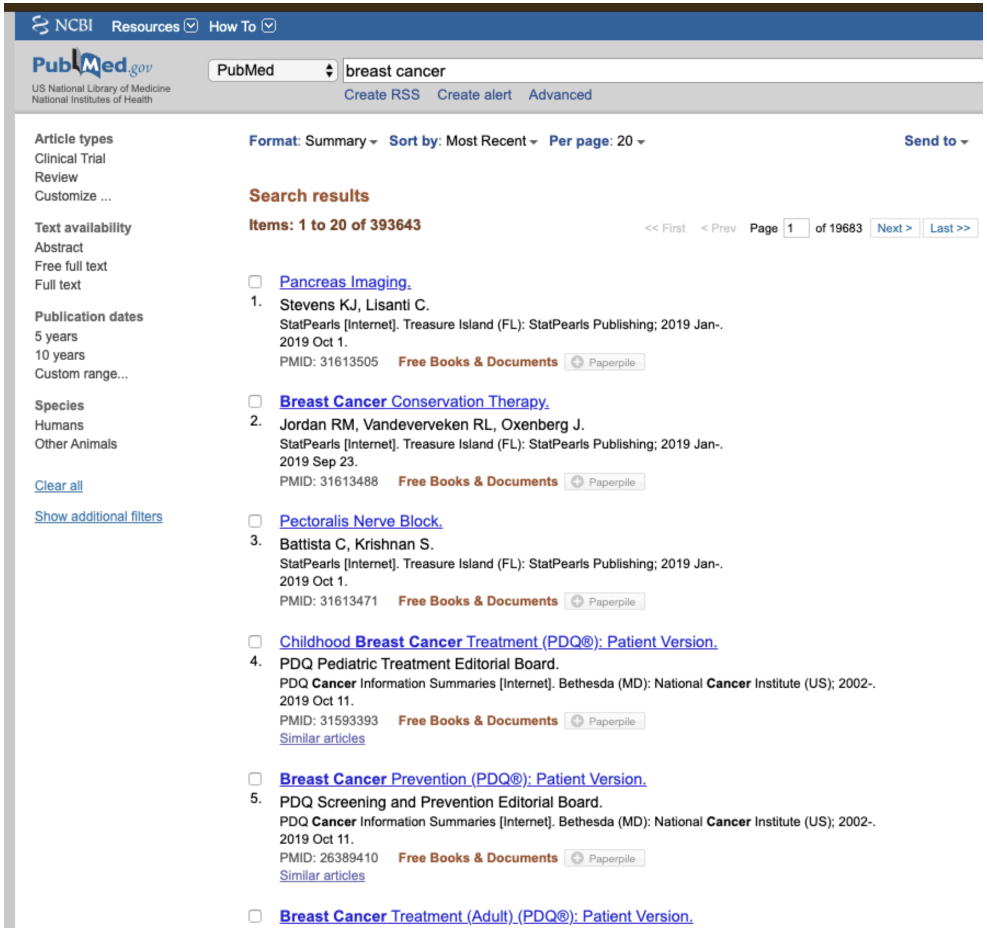
Natural Language Processing (NLP)

- Understanding human language is one of the main challenges for artificial (general) intelligence

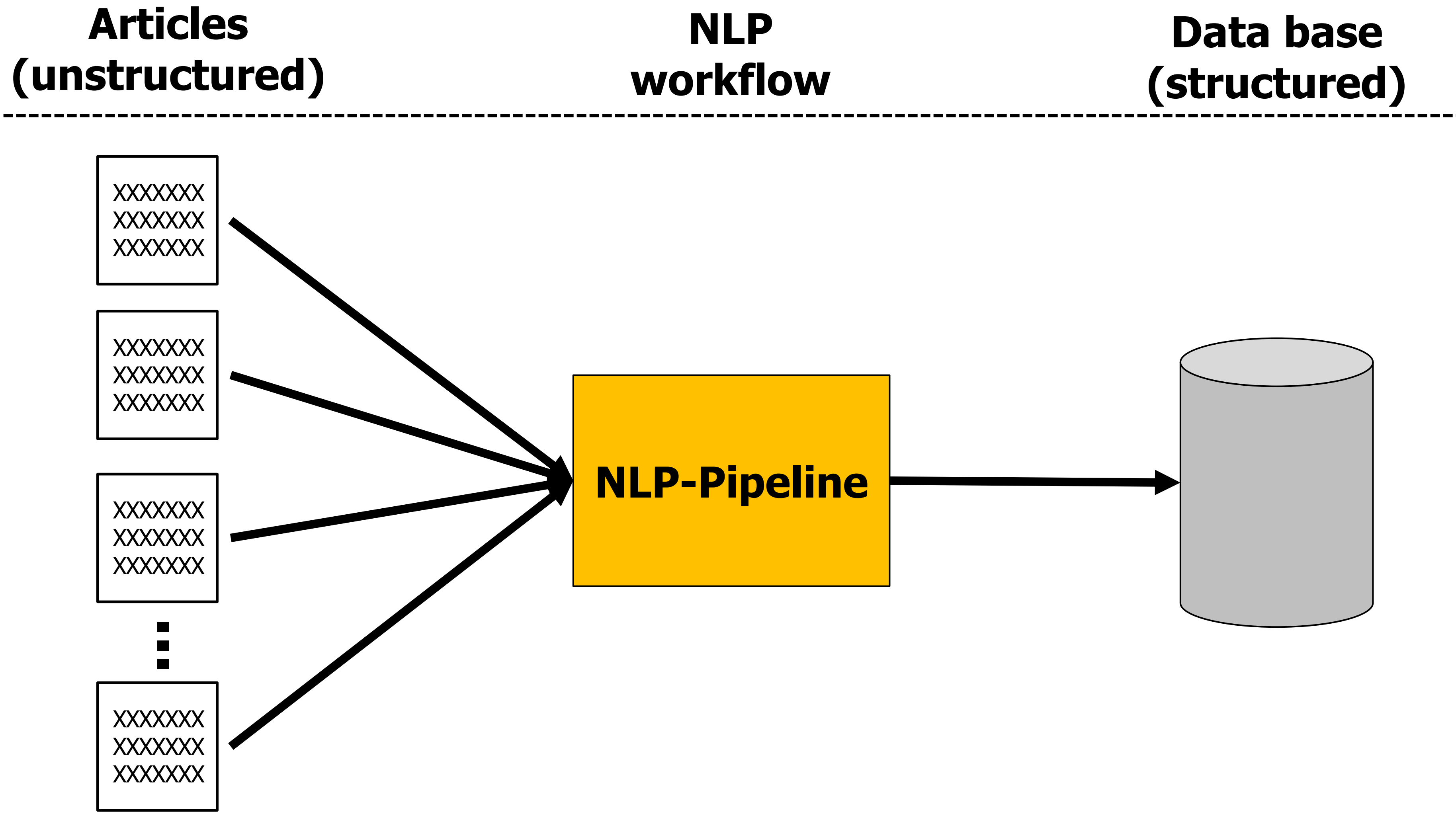
- Many of the world knowledge and information is given in textual form (e.g. scientific articles, newspaper articles, blogs, etc)
  - Rich source of information for a lot of tasks and applications
  - Often systematic access requires additional efforts

MEDLINE / PubMed

- MEDLINE: database of biomedical information
  - Houses more than 30 million abstract of scientific abstracts
  - Grows by about 1 million publications per year
- PubMed: search engine for MEDLINE
  - Used to retrieve relevant publications for an search interest



Information extraction (IE)



Natural Language Processing (NLP)

- Understanding human language is one of the main challenges for artificial (general) intelligence
- Many of the world knowledge and information is given in textual form (e.g. scientific articles, newspaper articles, blogs, etc)
  - Rich source of information for a lot of tasks and applications
  - Often systematic access requires additional efforts
- Plethora of different tasks and challenges (on varying level)
  - Syntactical: chunking, part-of-speech, syntactic parsing ...
  - Semantic: text classification, named entity recognition, relation extraction, ...

Part-of-speech tagging (POS)

- Task: assign the grammatical class to each word in a sentence
  - Simplest case: noun, verb, adjective, adverb, article, ...
  - Complex tag sets include morphological information (gender, case, tense, person, comparative, superlative, ...)

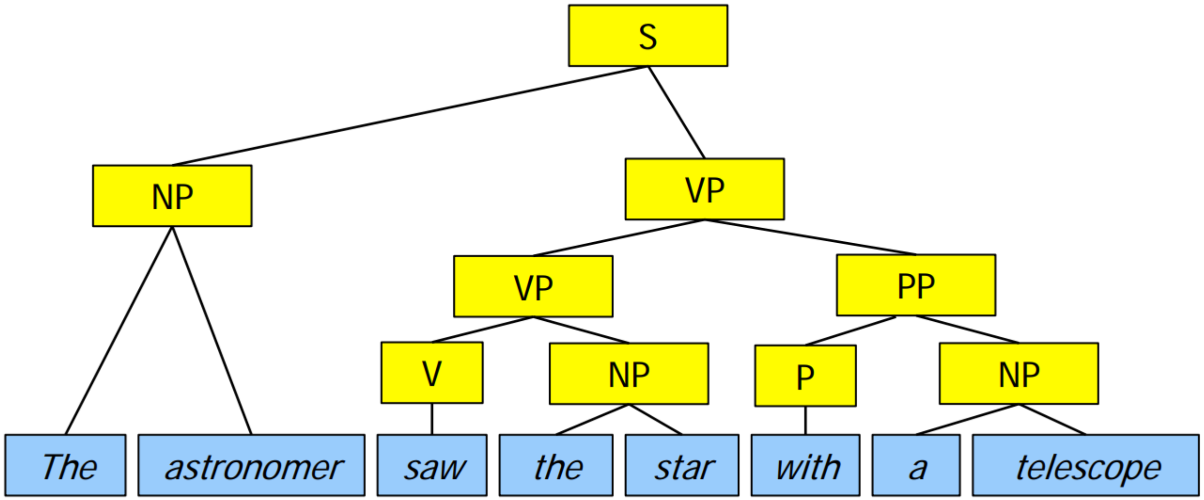
The	koala	put	the	keys	on	the	table
DT	NN	VBN	DT	NNS	P	DT	NN

- Grammatical class depends on the context / sentence!
  - Homonyms: “Win a grant” vs. “to grant access”
  - Intentional use: “We buy a house” vs. “Put the buy here”

Syntactic parsing

- Task: Infer the syntactical structure of a sentence
  - POS tagging studies the plain sequence of words in a sentence
  - But sentences have more and non-consecutive structures
  - Plenty of linguistic theories exist about the nature and representation of these structures / units / phrases / …

The astronomer saw the star with a telescope



Text classification

- A text is any sequence of tokens / words
  - Typical: books, scientific articles, news, emails, letters, ...
  - Atypical: tweets, reports with images and tables, spoken lang, ...

- Task: Assign each text to one of a given set of classes
  - Topic identification
  - Language identification
  - Spam detection
  - Sentiment analysis
  - Content-based messaging routing
  - Author identification
  - .....

Text classification

- Example: Topic identification
  - Classes: politics, economy, society, culture, sports, ...

Keimbelastete Wurst

Ikea nimmt Wilke-Wurstaufschnitt aus Sortiment

Bundesweit ergreifen Händler wegen keimbelasteter Produkte eines hessischen Herstellers Vorsichtsmaßnahmen. Verbraucherschützer werfen den Behörden Versäumnisse vor.

7. Oktober 2019, 16:59 Uhr / Quelle: ZEIT ONLINE, dpa, AFP, tst / 80 Kommentare

Nach zwei Todesfällen durch keimbelastete Fleischwaren des nordhessischen Wurstproduzenten Wilke ist auch der Möbelkonzern Ikea vom Rückruf betroffen. Über einen Großhändler habe Ikea Deutschland Wurstaufschnitt für Kunden- und Mitarbeiterrestaurants von diesem Hersteller erhalten, sagte eine Sprecherin und bestätigte damit Angaben der Verbraucherorganisation Foodwatch.

Ikea war nach eigenen Angaben am

https://www.zeit.de/wirtschaft/unternehmen/2019-10/keimbelastete-wurst-wilke-ikea-verkaufsstopp

Economy?

Sport?

Politics?

Text classification

- Example: Topic identification
  - Classes: politics, economy, society, culture, sports, ...

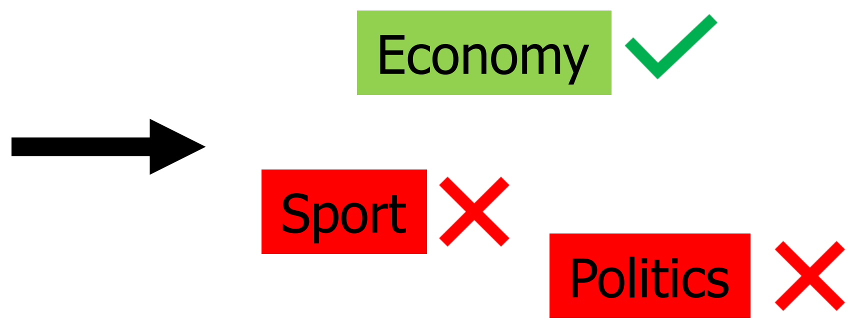
Keimbelastete Wurst

Ikea nimmt Wilke-Wurstaufschnitt aus Sortiment

Bundesweit ergreifen Händler wegen keimbelasteter Produkte eines hessischen Herstellers Vorsichtsmaßnahmen. Verbraucherschützer werfen den Behörden Versäumnisse vor.

7. Oktober 2019, 16:59 Uhr / Quelle: ZEIT ONLINE, dpa, AFP, tst / 80 Kommentare

Nach zwei Todesfällen durch keimbelastete Fleischwaren des nordhessischen Wurstproduzenten Wilke ist auch der Möbelkonzern Ikea vom Rückruf betroffen. Über einen Großhändler habe Ikea Deutschland Wurstaufschnitt für Kunden- und Mitarbeiterrestaurants von diesem Hersteller erhalten, sagte eine Sprecherin und bestätigte damit Angaben der Verbraucherorganisation Foodwatch.



https://www.zeit.de/wirtschaft/unternehmen/2019-10/keimbelastete-wurst-wilke-ikea-verkaufsstopp

Text classification

- Example: Sentiment analysis
  - Classes: Positive / negative or star rating (e.g. 5 or 10 star scale)

Sehr gute Maschine — Titel

7. Januar 2018

Farbe: Schwarz | Stil: Single | **Verifizierter Kauf**

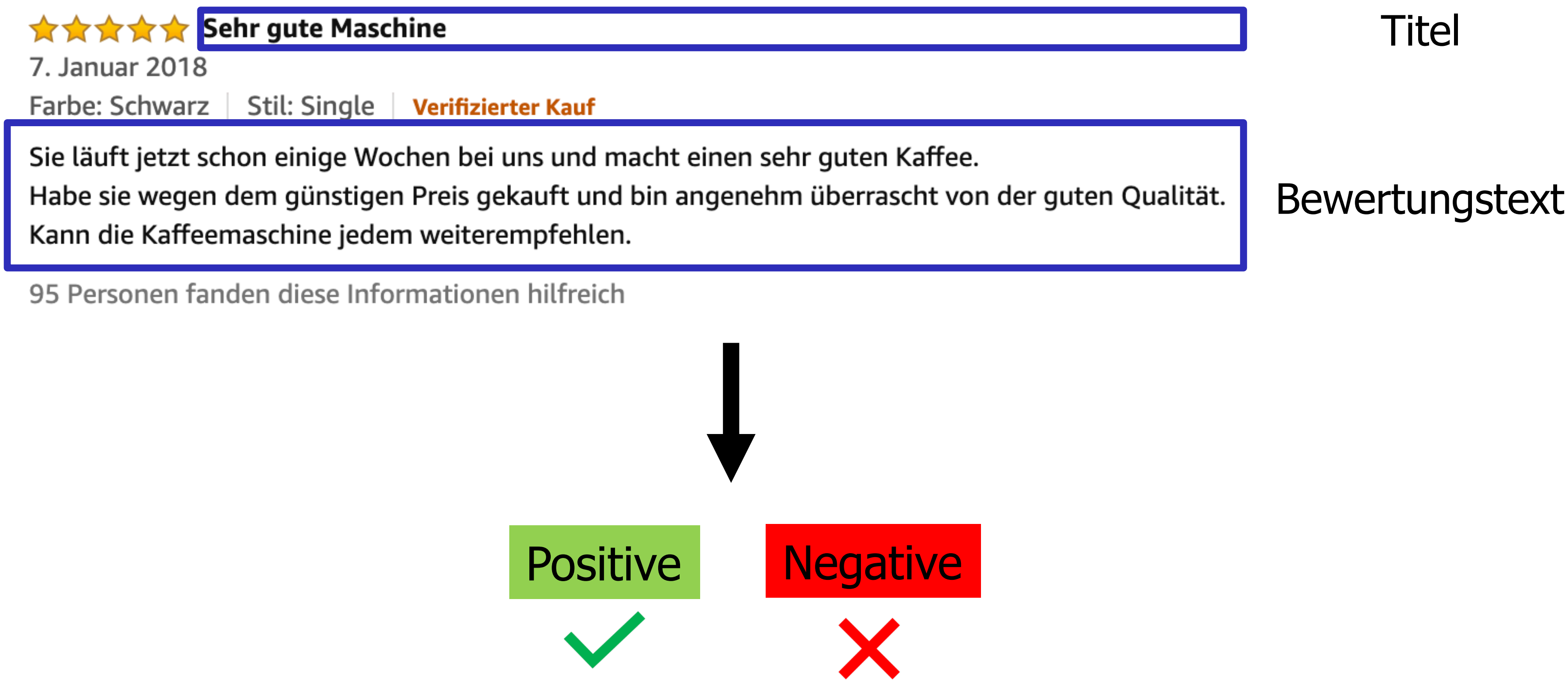
Sie läuft jetzt schon einige Wochen bei uns und macht einen sehr guten Kaffee.
Habe sie wegen dem günstigen Preis gekauft und bin angenehm überrascht von der guten Qualität.
Kann die Kaffeemaschine jedem weiterempfehlen. — Bewertungstext

95 Personen fanden diese Informationen hilfreich

Positive? Negative?

Text classification

- Example: Sentiment analysis
  - Classes: Positive / negative or 5-star rating



Named Entity Recognition

- Task: Find all mentions of a given type of entities
  - General domain: person, organization, location, time specs, ...
  - Biomedical: gene, protein, mutations, drugs, cell line, species, ...

Z-100 is an arabinomannan extracted from Mycobacterium tuberculosis that has various immunomodulatory activities, such as the induction of interleukin 12, interferon gamma (IFN-gamma) and beta-chemokines. The effects of Z-100 on human immunodeficiency virus type 1 (HIV-1) replication in human monocyte-derived macrophages (MDMs) are investigated in this paper. In MDMs, Z-100 markedly suppressed the replication of not only macrophage-tropic (M-tropic) HIV-1 strain (HIV-1JR-CSF), but also HIV-1 pseudotypes that possessed amphotropic Moloney murine leukemia virus or vesicular stomatitis virus G envelopes. Z-100 was found to inhibit HIV-1 expression, even when added 24 h after infection. In addition, it substantially inhibited the expression of the pNL43lucDeltaenv vector (in which the env gene is defective and the nef gene is replaced with the firefly luciferase gene) when this vector was transfected directly into MDMs. These findings suggest that Z-100 inhibits virus replication, mainly at HIV-1 transcription. However, Z-100 also downregulated

Named Entity Recognition

- Task: Find all mentions of a given type of entities
  - General types: person, organization, location, time specs, ...
  - Biomedical: gene, protein, mutations, drugs, cell line, species, ...

Z-100 is an ***arabinomannan*** extracted from Mycobacterium tuberculosis that has various immunomodulatory activities, such as the induction of **interleukin 12**, **interferon gamma** (**IFN-gamma**) and beta-chemokines. The effects of ***Z-100*** on human immunodeficiency virus type 1 (HIV-1) replication in human **monocyte-derived macrophages** (**MDMs**) are investigated in this paper. In **MDMs**, ***Z-100*** markedly suppressed the replication of not only macrophage-tropic (M-tropic) HIV-1 strain (HIV-1JR-CSF), but also HIV-1 pseudotypes that possessed amphotropic Moloney murine leukemia virus or vesicular stomatitis virus G envelopes. ***Z-100*** was found to inhibit HIV-1 expression, even when added 24 h after infection. In addition, it substantially inhibited the expression of the pNL43lucDeltaenv vector (in which the ***env*** gene is defective and the ***nef*** gene is replaced with the ***firefly luciferase*** gene) when this vector was transfected directly into **MDMs**. These findings suggest that ***Z-100*** inhibits virus replication, mainly at HIV-1 transcription. However, ***Z-100*** also

Relation extraction (RE)

- Task: Find all mentioned relationships between the entities in a text

Z-100 is an *arabinomannan* extracted from Mycobacterium tuberculosis that has various immunomodulatory activities, such as the induction of **interleukin 12**, **interferon gamma** (**IFN-gamma**) and beta-chemokines. The effects of *Z-100* on human immunodeficiency virus type 1 (HIV-1) replication in human **monocyte-derived macrophages** (**MDMs**) are investigated in this paper. In **MDMs**, *Z-100* markedly suppressed the replication of not only macrophage-tropic (M-tropic) HIV-1 strain (HIV-1JR-CSF), but also HIV-1 pseudotypes that possessed amphotropic Moloney murine leukemia virus or vesicular stomatitis virus G envelopes. *Z-100* was found to inhibit HIV-1 expression, even when added 24 h after infection. In addition, it substantially inhibited the expression of the pNL43lucDeltaenv vector (in which the ***env*** gene is defective and the ***nef*** gene is replaced with the ***firefly luciferase*** gene) when this vector was transfected directly into **MDMs**. These findings suggest that *Z-100* inhibits virus replication, mainly at HIV-1 transcription. However, *Z-100* also downregulated expression of the cell surface receptors **CD4** and **CCR5** in **MDMs**, suggesting some inhibitory effect on HIV-1 entry. Further experiments revealed that *Z-100* induced **IFN-beta** production in these cells, resulting in induction of the 16-kDa **CCAAT/enhancer binding protein** (**C/EBP**) **beta transcription factor** that represses HIV-1 long terminal repeat transcription. These effects were alleviated by SB 203580, a

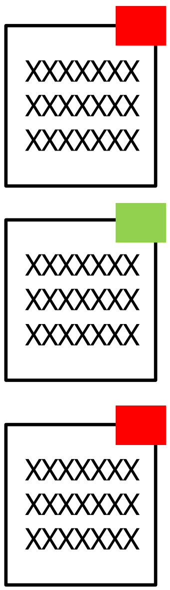
Overview NLP approaches

- 1. Rule-based approaches
  - Early methods are symbolic approaches using manually written rules to capture the meaning of text
  - Tied to the specific task / domain that they are designed for
  - Unable to deal with unseen or unexpected input

- 2. Statistical approaches:
  - Usage of mathematical models to “learn rules” from data
  - Human task is to create features that tell the model which characteristics is should take into account to make the prediction
  - Feature engineering is time-consuming, task-specific and needs domain expertise

Supervised Learning

- Given a set D of documents and a set of classes C. A classifier is a function $f: D \rightarrow C$

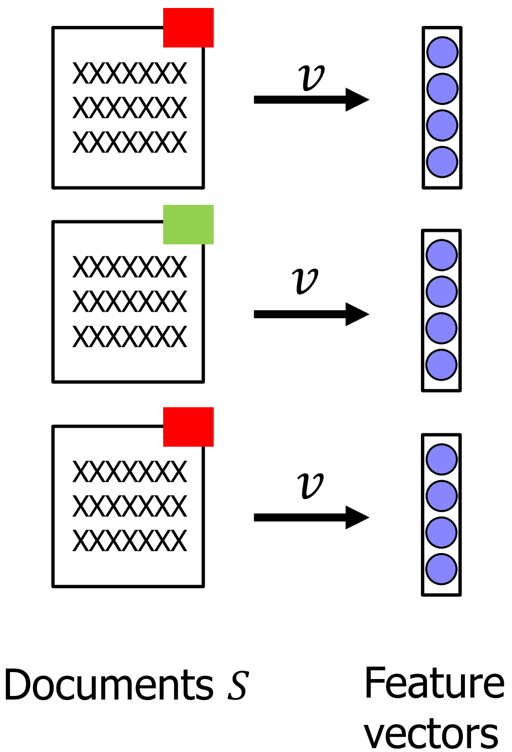


Documents S

- Obtain a set S of docs with their classes (training data)
- **Often, this is the most critical issue!**

Supervised Learning

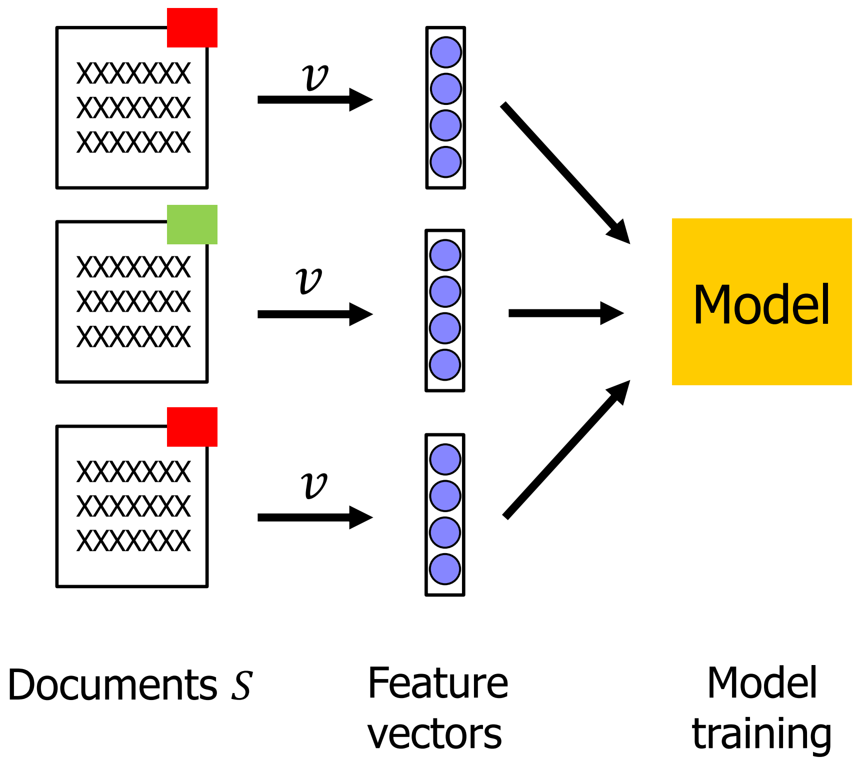
- Given a set D of documents and a set of classes C. A classifier is a function $f: D \rightarrow C$



- Design function v mapping a doc into feature vector (feature space)
  - Bag-of-words, TF-IDF
  - POS tags, language style, length
  - Syntactic properties
  - Other meta data (e.g. author, date)
  - ....

Supervised Learning

- Given a set D of documents and a set of classes C. A classifier is a function $f: D \to C$



- Find the characteristics of the docs in each class (model training)
  - Which feature values / ranges are characteristic?
  - What combinations or features are characteristic?
- Encode the model in a classifier function f operating on the feature vector: $v: D \to V$ and $f: V \to C$
- Classification: compute $f(v(d))$

Overview NLP approaches

- 1. Rule-based approaches
  - Early methods are symbolic approaches using manually written rules to capture the meaning of text
  - Tied to the specific task / domain that they are designed for
  - Unable to deal with unseen or unexpected input

- 2. Statistical approaches:
  - Usage of mathematical models to “learn rules” from data
  - Human task is to create features that tell the model which characteristics is should take into account to make the prediction
  - Feature engineering is time-consuming, task-specific and needs domain expertise

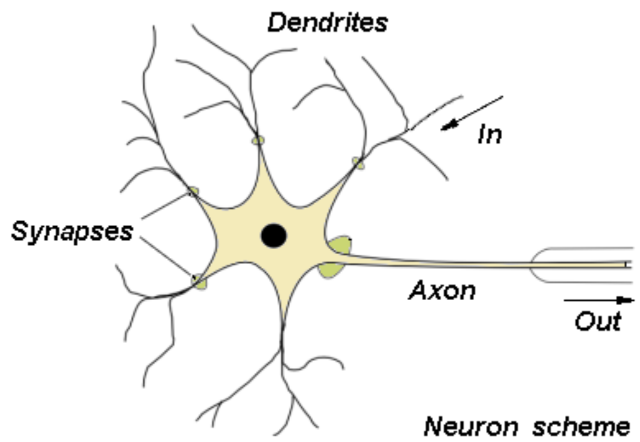
Overview NLP approaches

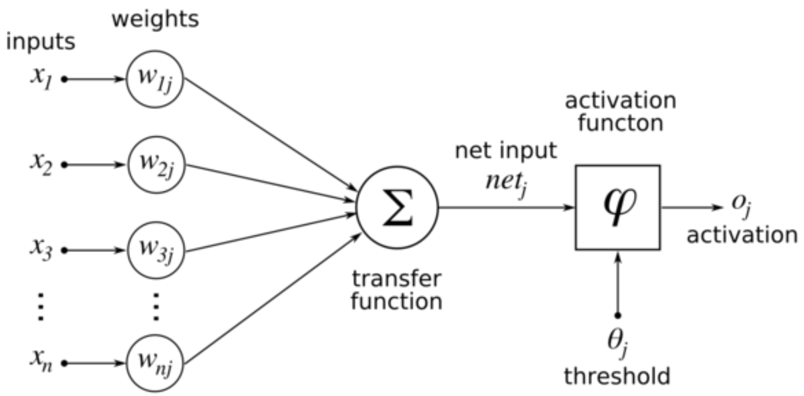
- 3. Deep Learning
  - Since about 2012 deep neural networks dominate NLP community
  - Automatically learn a multi-layered hierarchy of features based on “raw data”
    - Reduces need for feature engineering
  - Human task is to determine the most suitable architecture and training setting

Deep learning / neural networks

- A method for non-linear classification
  - Quite old, always present, extremely hyped since ~2010
  - Breakthrough based on (extreme) growth in computational resources and data availability

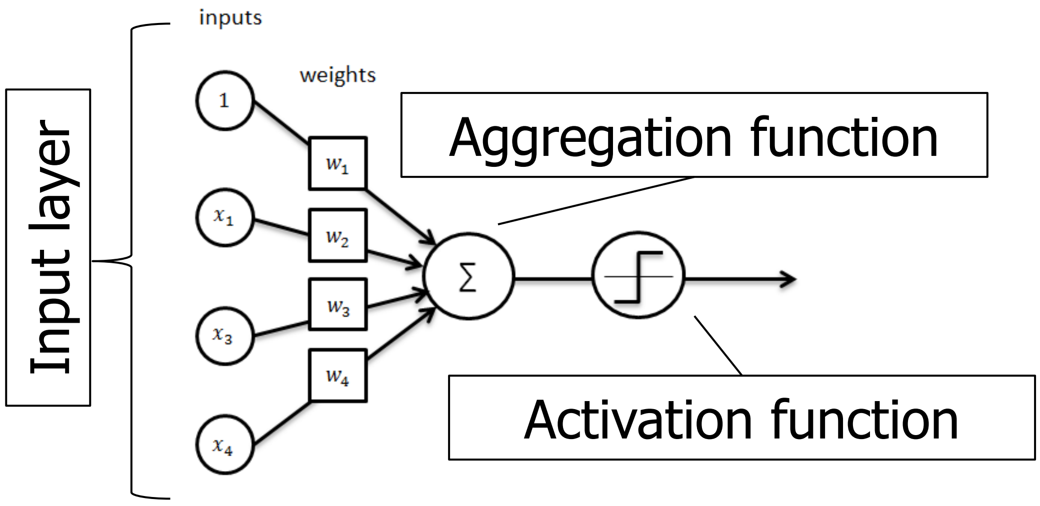
- Inspired by biological neural networks





Concepts of neural networks

- Nodes (~ artificial neurons): basic module of an neural network
  - Receives input signal from raw data or other nodes
  - Aggregates values from the input layer (e.g. weighted sum)
  - Outputs a real value based on an activation function

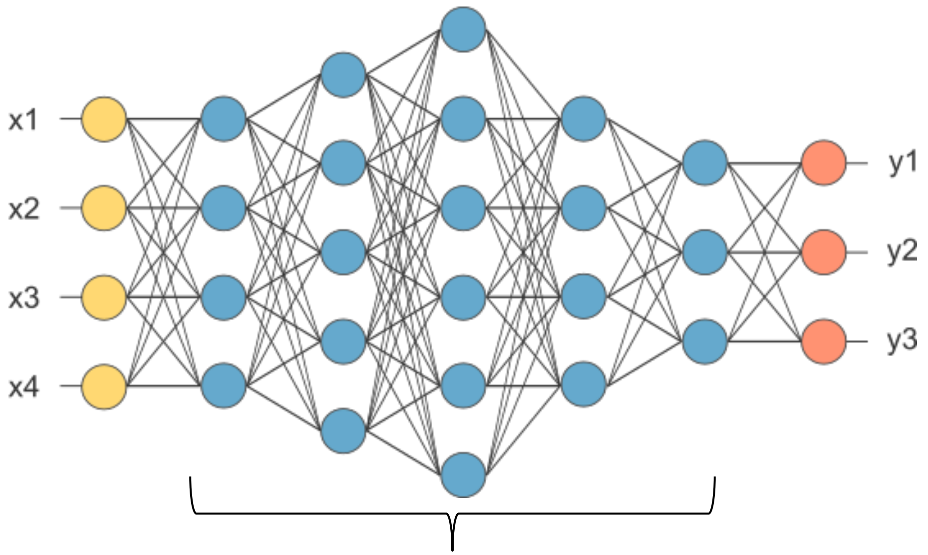


http://www.opennn.net/

Concepts of neural networks

- Nodes are organized and stacked in layers
  - Each layer represents / learns more complex abstractions of the data based on the abstractions of the layer before
  - This results in a hierarchy of features

Input layer



Output layer

Hidden layers

http://www.opennn.net/

Introduction

- Relation Extraction (RE) -

Relation extraction (RE)

- Task: Find all mentioned relationships between the entities in a text
  - Who is the CEO of a company?
  - What product aspect does a user like?
  - Which proteins interact with which other proteins?
  - Which genetic mutations cause which diseases?
  - ....
- Often, RE depends on pre-recognized entities
  - Can be modelled as joint inference problem (later)

Relation extraction (RE)

- Task: Find all mentioned relationships between the entities in a text

Z-100 is an ***arabinomannan*** extracted from Mycobacterium tuberculosis that has various immunomodulatory activities, such as the induction of **interleukin 12**, **interferon gamma** (**IFN-gamma**) and beta-chemokines. The effects of *Z-100* on human immunodeficiency virus type 1 (HIV-1) replication in human **monocyte-derived macrophages** (**MDMs**) are investigated in this paper. In **MDMs**, *Z-100* markedly suppressed the replication of not only macrophage-tropic (M-tropic) HIV-1 strain (HIV-1JR-CSF), but also HIV-1 pseudotypes that possessed amphotropic Moloney murine leukemia virus or vesicular stomatitis virus G envelopes. *Z-100* was found to inhibit HIV-1 expression, even when added 24 h after infection. In addition, it substantially inhibited the expression of the pNL43lucDeltaenv vector (in which the ***env*** gene is defective and the ***nef*** gene is replaced with the ***firefly luciferase*** gene) when this vector was transfected directly into **MDMs**. These findings suggest that *Z-100* inhibits virus replication, mainly at HIV-1 transcription. However, *Z-100* also downregulated expression of the cell surface receptors **CD4** and **CCR5** in **MDMs**, suggesting some inhibitory effect on HIV-1 entry. Further experiments revealed that *Z-100* induced **IFN-beta** production in these cells, resulting in induction of the 16-kDa **CCAAT/enhancer binding protein** (**C/EBP**) **beta transcription factor** that represses HIV-1 long terminal repeat transcription. These effects were alleviated by SB 203580, a

Types of relation extraction problems

- Binary RE: Only the entities that are in a certain relation
  - Output: tuples / pairs of entities with fixed semantics

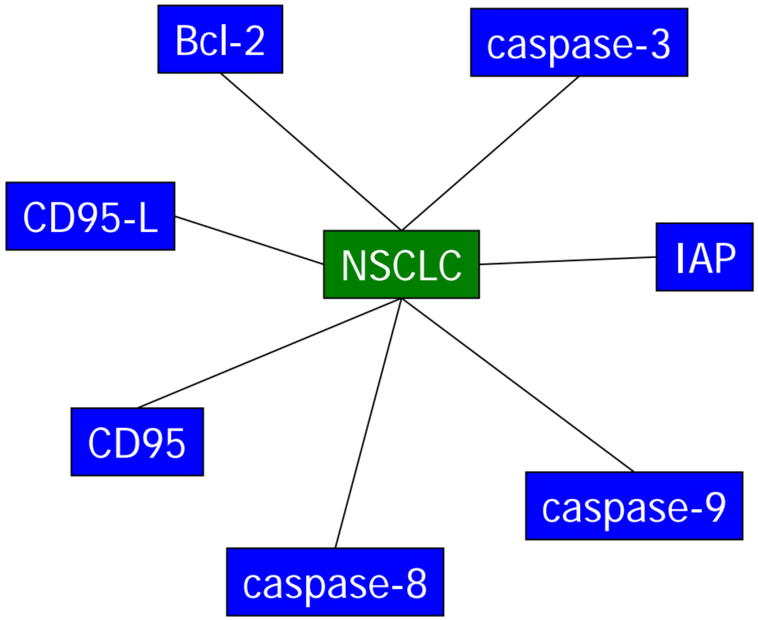
- Multi-class RE: Infer the type (of a given set of types) of the relation between the entities
  - Detect entities and deduce semantics of their relation
  - Example: Brad Pitt and Angelina Jolie - *married*?, *brother_of*?, *parent_of*?, *co_worker_of*?,....
  - Output: entity tuples and specific relation type

- Event extraction: extraction of complex relations with more than two entities and multiple relations

Approaches for RE

- (Very) Simple: Co-occurrence-based extraction
  - Very often, entities mentioned in a sentence are in a certain relationship to each other

NSCLC often becomes resistant to chemotherapy due to multiple defects found in expression of ***CD95-L***, ***CD95*** and members of the ***Bcl-2*** and ***IAP*** family, as well as ***caspase-8***, ***-9*** and ***-3*** as examined by ..



Co-occurrence: 28 relationships, 21 false positives

Co-occurrence-based RE

- All pairs of entities appearing together in a context
  - A sentence, a paragraph, a window of n words
    - Larger context: higher recall (e.g. across sentences), lower precision
  - Best context size for a given relationship can be learned

- Yields high recall yet poor precision in general
  - Problems with enumerations, nested structures, long sentences, …
  - Completely agnostic to relationship type

- Improvement: Pre-filtering sentences for "type'ness"
  - For instance, filter by a set of verbs or trigger words

Pattern-based approaches for RE

- Language pattern (aka Hearst Pattern)
  - Look at words occurring in sentences expressing a relationship
    - … GENE regulates expression of GENE …
    - … GENE is strongly suppressed by GENE …
  - Adding part-of-speech
    - … GENE VRB NOM PRP GENE …
    - … GENE is ADV VRB PRP GENE …

Balanced precision/recall requires many rules!

- Different levels of generality
  - … GENE .* VRB .* GENE
    - Simple rule, high recall, low precision
  - ... GENE [is] ADV? {regulat|suppres} NOM? PRP GENE
    - Complex rules, lower recall, higher precision

Pattern-based approaches for RE

- Most pattern-based systems work on hand-crafted sets of pattern
  - Recall: Users love pattern/rule-based approaches
  - Good recall quickly requires hundreds of pattern - large effort!
  - Need to be created for any type of relationship separately
    - Protein-protein, gene-disease, disease-drug, …
- One idea: Learn patterns from weakly labelled data
  - Semi-supervised learning
  - More specific term: distant supervision
  - User-friendly: patterns can be inspected, removed, modified, …

(Rough) Idea

- Assume we seek protein-protein-interactions (PPI)
  - Fortunately, there exist databases of PPIs, e.g. IntAct
- Hypothesis: If a pair of proteins known to interact (from a database) co-occur in a sentence, then this sentences expresses a PPI
  - Can be used to quickly find thousands of relevant sentences
- Sentences are then turned into patterns
  - These patterns can be matched against new text to find novel PPIs

Classification-based RE

- Idea: Classify each pair of entities
  - Consider each entity pair (in a sentence) as an object
  - Compute a feature vector for this object
    - POS tags, distance, words, words in between, path in the dependency tree connecting the two, neighbourhood, trigger words, …
  - Learn a model from training data
  - Classify each object (~ pair) as having the relationship or not

- Approaches can be categorised into three groups:
  - Feature-based: utilise hand-crafted lexical and syntactical features
  - Kernel-based: model similarity between syntactical trees
  - Deep learning: usage of neural networks

Introduction

- Transfer Learning (TL) -

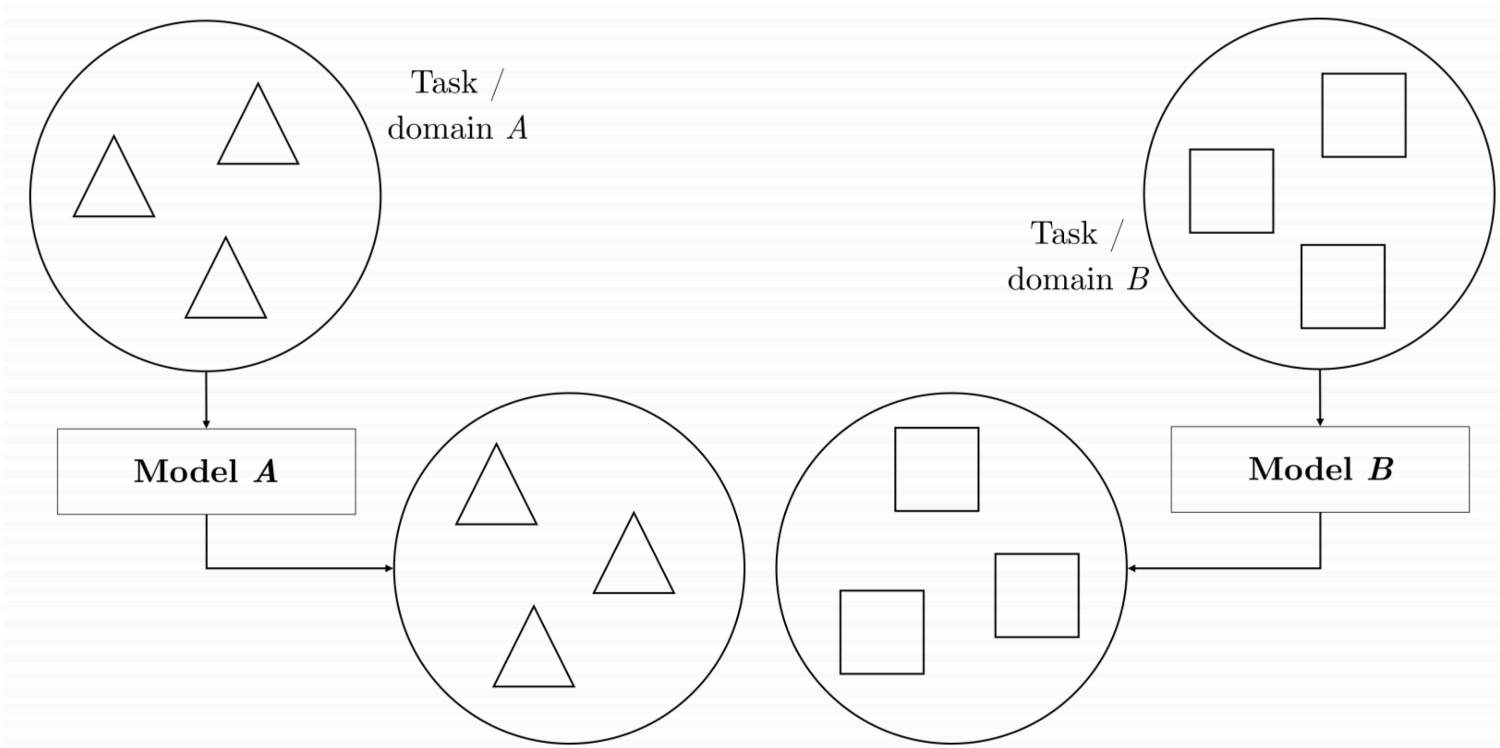
Transfer Learning (TL)

- Needs of humans are diverse and complex
  - Constantly require NLP techniques to solve new tasks

- Supervised learning requires a sufficient number of examples for every new tasks
  - Usually, for every task a new model is trained from scratch
  - Plethora of tasks, domains and languages makes it infeasible to manually annotate examples for each setting

Traditional setup

- Learn a new model for every task



http://ruder.io/thesis/neural_transfer_learning_for_nlp.pdf

Transfer Learning (TL)

- Needs of humans are diverse and complex
  - Constantly require NLP techniques to solve new tasks
- Supervised learning requires a sufficient number of examples for every new tasks
  - Usually, for every task a new model is trained from scratch
  - Plethora of tasks, domains and languages makes it infeasible to manually annotate examples for each setting
- Knowledge from different tasks isn't combined and model training starts from a random initialization
  - Antithetic to human language / knowledge acquisition

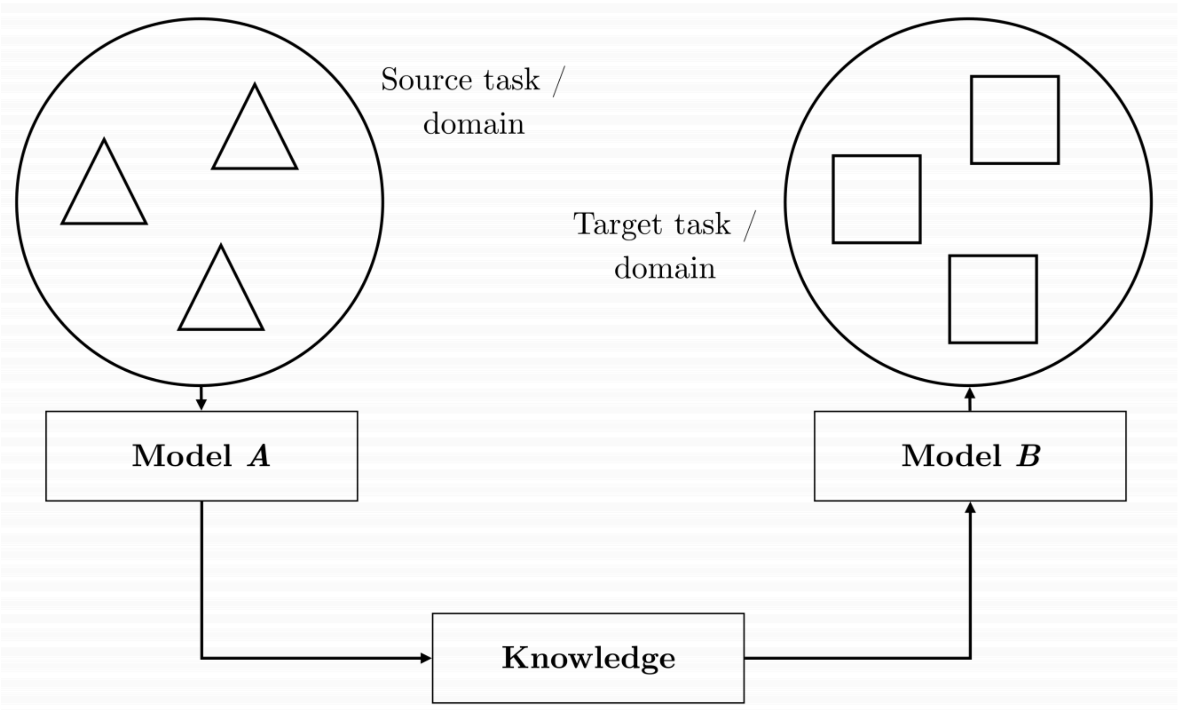
Transfer Learning (TL)

- Recent studies show that ML algorithms are often brittle in similar way like rule-based approaches
  - Heavily conform the training data characteristics
  - Can't adapt well if conditions change

- TL addresses this problem and tries to transfer knowledge between different tasks
  - Utilize learned knowledge from other tasks, domains and/or languages to solve the target task

Transfer learning setup

- Utilize knowledge learned from a task A to solve the target task B



http://ruder.io/thesis/neural_transfer_learning_for_nlp.pdf

Transfer Learning (TL)

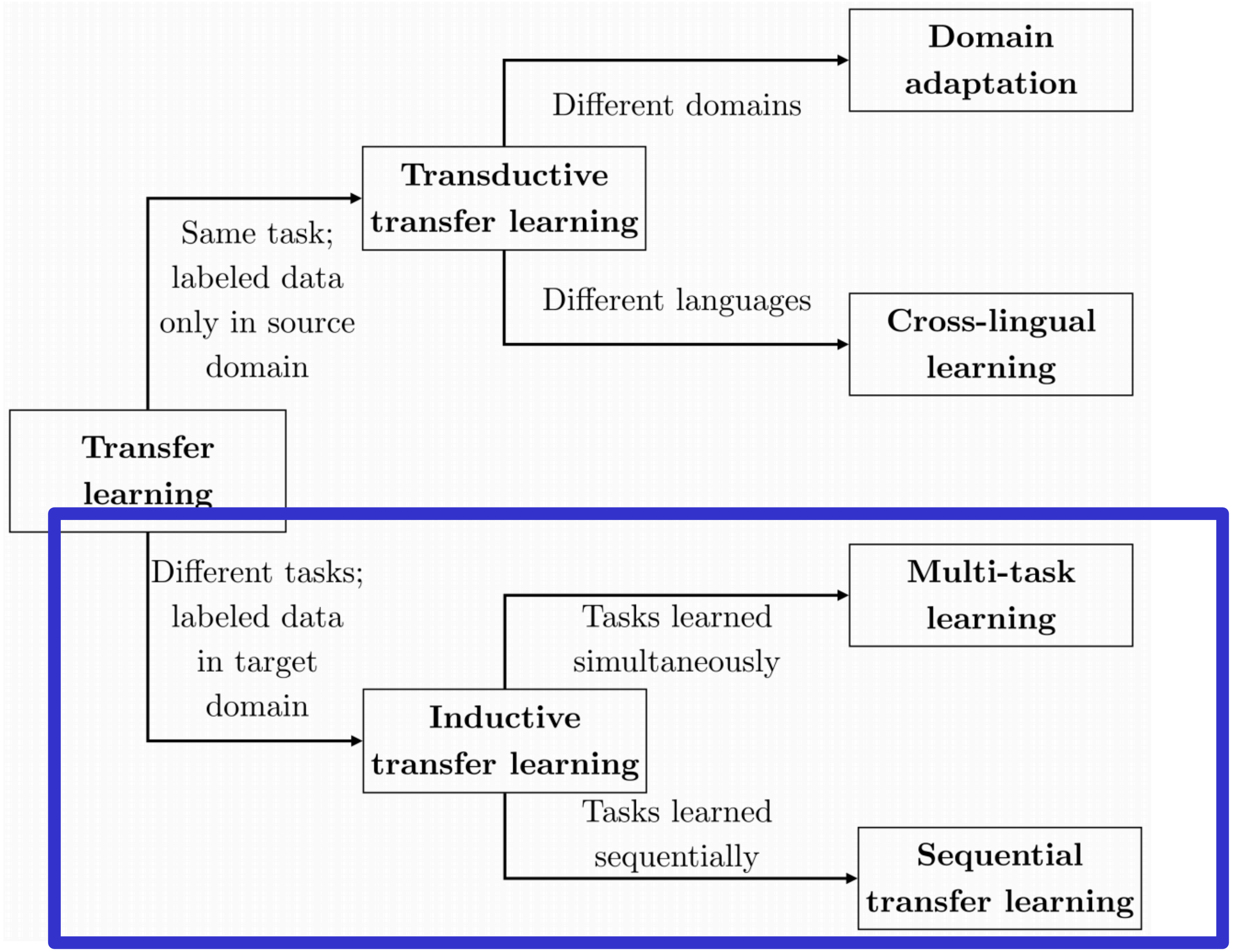
- Recent studies show that ML algorithms are often brittle in similar way like rule-based approaches
  - Heavily conform the training data characteristics
  - Can't adapt well if conditions change
- TL addresses this problem and tries to transfer knowledge between different tasks
  - Utilize learned knowledge from other tasks, domains and/or languages to solve the target task
- Goal: develop models that generalizes better

Transfer learning scenarios

- NLP problem settings can differ in several aspects:
  - Different feature distributions
    - E.g. movie vs. coffee machine reviews (~ different topics)
  - Different feature spaces
    - E.g. reviews in different languages
  - Different label distributions
    - E.g. different sentiment distribution between product categories
  - Different label spaces
    - E.g. positive-negative vs. 10-star rating

- In NLP, two categories of TL: transductive and inductive transfer learning

Transfer learning taxonomy for NLP



http://ruder.io/thesis/neural_transfer_learning_for_nlp.pdf

Multi-task learning (MTL)

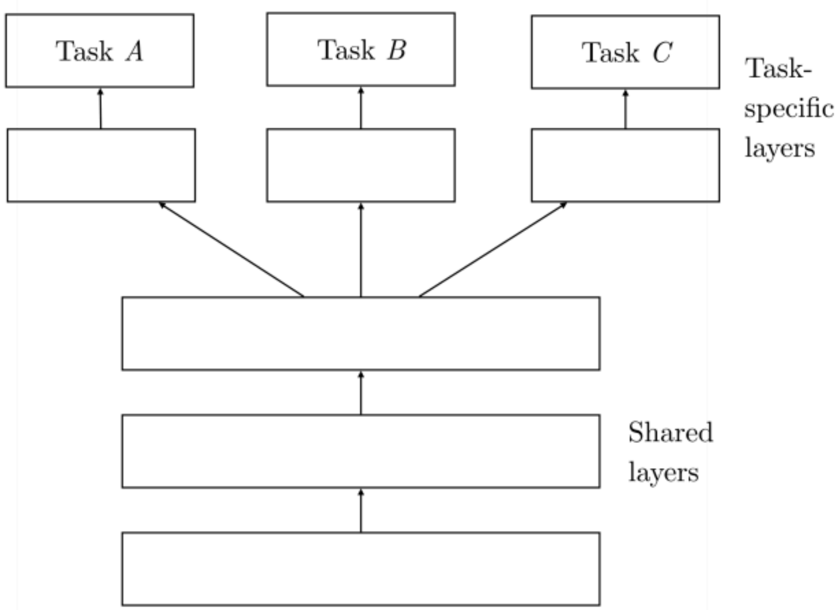
- Motivation: For learning new tasks, we often apply the knowledge we have acquired by learning related tasks
  - First provide necessary skills to master more complex techniques

- Learn multiple tasks simultaneously
  - Introduces an inductive biases provided by auxiliary tasks
  - Prefer hypotheses that can explain more than one task
  - Implicitly reduces chances of overfitting

- Share representations / features between related tasks
  - Two methods: Hard and soft parameter sharing

Multi-task learning (MTL)

- Hard parameter sharing
  - Most commonly used form of MTL in neural networks
  - Share hidden layers between several tasks
  - Use task-specific output layers for each task

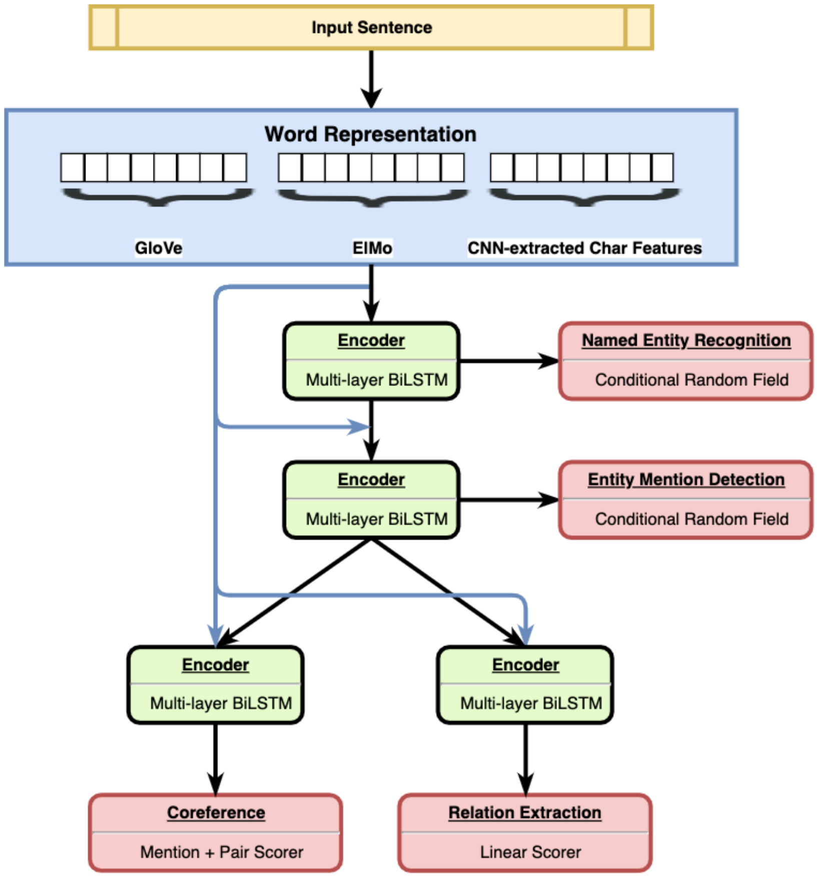


http://ruder.io/thesis/neural_transfer_learning_for_nlp.pdf

Example: Hard parameter sharing

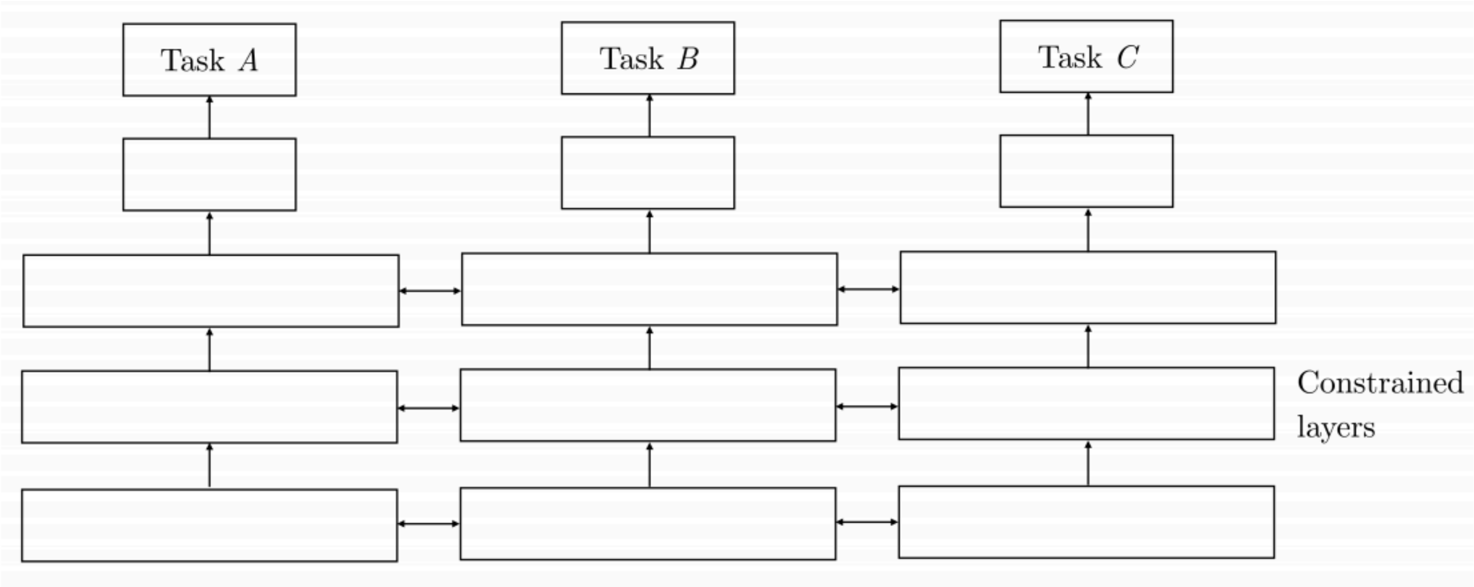
- Sanh et al. [1] use a hierarchical multi-task model solving four semantic tasks
  - Named entity recognition
  - Entity mention detection
  - Coreference resolution
  - Relation extraction

- Achieves SOTA performance on three of the four tasks



Multi-task learning (MTL)

- Soft parameter sharing
  - Each task has it‘s own model with it‘s own parameters
  - Regularize the distance between the parameters
    - ~ Encourage the parameters to be similar



http://ruder.io/thesis/neural_transfer_learning_for_nlp.pdf

Why does MTL work?

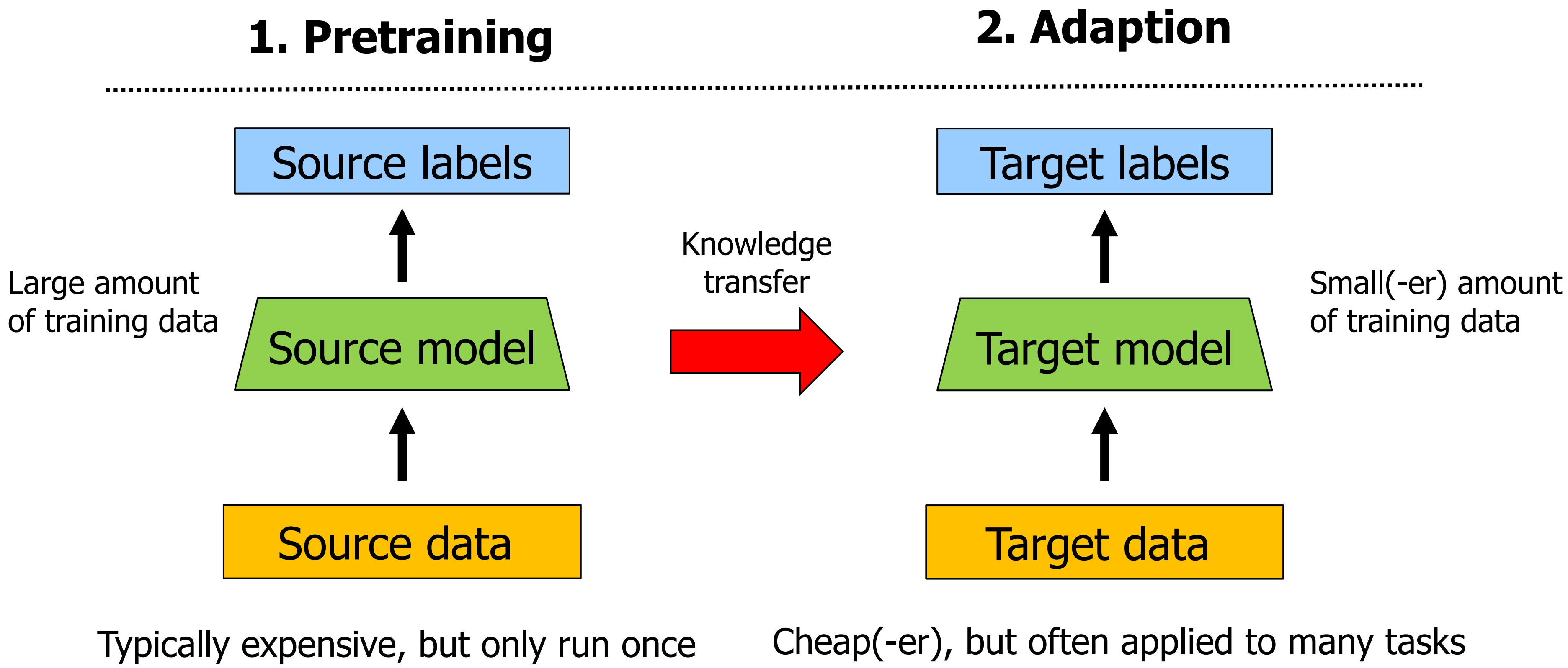
- Implicit data augmentation
  - MTL effectively increases the sample size for model training
- Representation bias
  - Learning of more general representations / features through averaging noise patterns of different tasks
  - MTL biases the model to prefer representations /features that other tasks also prefer
- Additional regularization
  - Prevent model to overfit to a single task

Sequential Transfer Learning (STL)

- Most frequently used transfer learning scenario in NLP
- Source and target task are different and training is performed in sequence
  - Each task is learned separately – no joint learning!
  - Goal: Improve target task performance by learned information from source task
- Useful in three scenarios
  - Data for the tasks are not available at the same time
  - Source task has much more data than the target task
  - Adaption to many target tasks is necessary

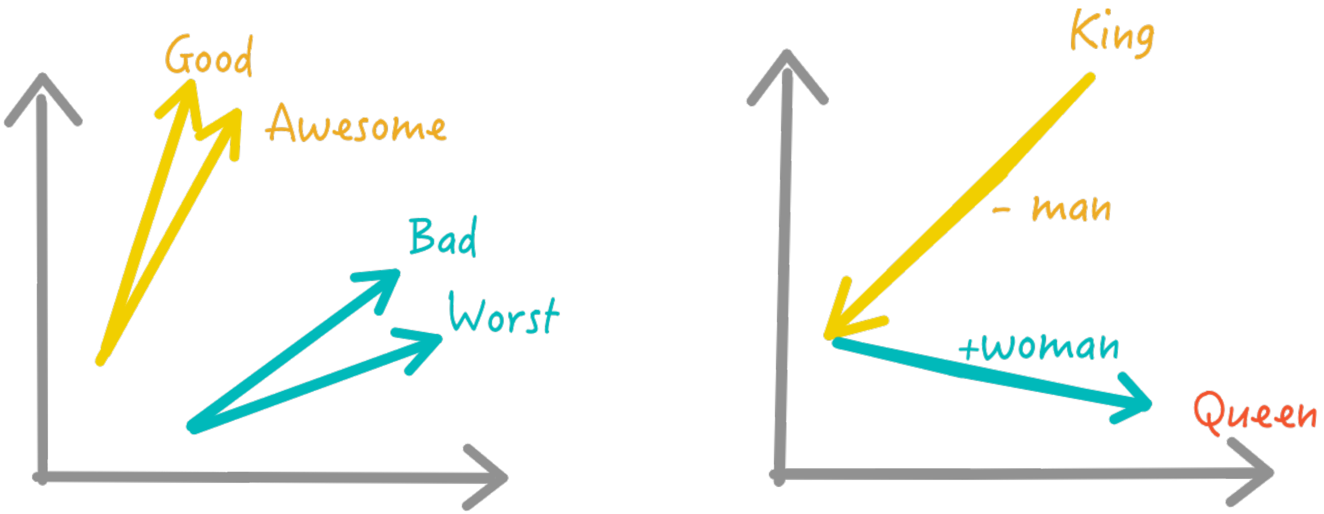
Sequential Transfer Learning (STL)

- STL typically consists of two stages / phases: pretraining and adaption phase



Example: Word2Vec

- Mikolov et al. [2,3] learn word representations based on word co-occurrences in large text corpora



https://mlwhiz.com/images/word2vec.png

- STL: Use word representations as input features for the target tasks

Questions?

Organization

Who should be here?

- Master Informatik
  - Also: Wirtschaftsinformatik, Ms. Education, Diplominformatik (?)

- Ability to read English papers
- Advanced programming skills (preferably Python)

- Ideally
  - Knowledge on machine learning / natural language processing
  - Knowledge in statistics, probability theory, math
  - Or willingness to learn this

How it will work

- Every group (2-3 students) has to implement a transfer learning approach
  - In general, starting point will be a concrete paper / implementation
  - Approach has to be adapted to biomedical domain (e.g. use other databases or resources, different tasks)
  - Free choice to change the approach or combine it with other solutions - own ideas are very welcome!!!

- We will release the training sets in November
  - A set of biomedical relation extraction corpora (e.g. PPI)
  - Understand, program, test and optimize your approach
    - We can provide access to two GPU servers of our research group
  - Include approach description in seminar talk and presentation

How it will work

- We will evaluate your method on held-back test data
  - You will be given an unlabelled test set for each corpora
  - Apply your approach to the test data and submit results
  - Comparison of the results in a competition
  - Discussion of the results during the presentations at the end of the semester
- Small price for best average scores among all groups and the group with the fanciest idea! ☺

Timeline

- Today: Introduction and group formation
  - Look to your left / right and find a friendly looking person ☺
  - ... or search via: https://cutt.ly/mesBK9K
- Introductory lectures:
  - Thursday 24.10.: Crash course on machine learning for NLP
  - Thursday 31.10.: Deep learning for natural language processing
  - Time? Additionally hands-on session?
- Topic selection
  - Topics will be released next week
  - Choose a topic until 31.10.

Timeline

- Flash presentation (12.12.)
  - Present approach and ideas in 5 minutes
  - Before 05.12.: meet me to discuss the topic / approach

- Blockseminar (first week in February)
  - Present your approach (~ 30 minutes) at the Blockseminar
  - Discussion of the competition results
  - Before 31.01.: meet me to discuss the slides

- Seminar thesis before 31.03.!
  - Write seminar thesis (~20 pages)

General information

- Literature research is necessary
  - The proposed papers are only anchors - further work must be considered!
  - You will likely have to get familiar with the basics of the approaches first!
- Read critically!
  - Don't be afraid of not entirely accurate statements - as long as there are good reasons to do so
  - Justify and argue - uncritical copying is out of place
- Don’t be afraid to ask questions!
- Be creative! ☺

General literature

- Text books:
  - Manning et al: *Foundations of statistical natural language processing*
  - Bishop: *Pattern recognition and machine learning*
  - Goodfellow et al.: *Deep learning*

- Overview papers:
  - Fabrizio: "*Machine learning in automated text categorization.*" ACM computing surveys (CSUR), 2002
  - Deyu et al.: "*Biomedical relation extraction: from binary to complex.*" Computational and mathematical methods in medicine, 2014
  - Zhang, Yijia, et al. "*Neural network-based approaches for biomedical relation classification: A review.*" Journal of biomedical informatics, 2019

Online courses

- Fast.ai – Introduction to machine learning:
  https://www.fast.ai/2018/09/26/ml-launch/

- Coursera – machine learning:
  https://de.coursera.org/learn/machine-learning

- Stanford - Natural Language processing with deep learning:
  https://www.youtube.com/playlist?list=PLoROMvodv4rOhcuXMZkNm7j3fVwBBY42z

Transfer Learning

- Ruder, Sebastian. *Neural Transfer Learning for Natural Language Processing*. Diss. NATIONAL UNIVERSITY OF IRELAND, GALWAY, 2019. (http://ruder.io/thesis/)

- Ruder, Sebastian. "*An overview of multi-task learning in deep neural networks.*" arXiv preprint arXiv:1706.05098, 2017

- Ruder, Sebastian, et al. "*Transfer Learning in Natural Language Processing.*" Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials. 2019.

Thank you for your interest!

See you next week!

Literature

[1] Sanh, Victor, Thomas Wolf, and Sebastian Ruder. "A hierarchical multi-task approach for learning embeddings from semantic tasks.“ *Proceedings of the AAAI Conference on Artificial Intelligence*.

[2] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* (2013).

[3] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*. 2013.