Institute of Mathematics, Bhubaneswar, India
Natural Language Processing Webinar
1st May 2021

Deep Learning Approaches for Natural Language Processing

Shantipriya Parida
Idiap Research Institute
Martigny, Switzerland

idiap
RESEARCH INSTITUTE
www.idiap.ch

- **Areas of activity**: multilingual and multimodal interaction and multimedia information management, including human behavior modeling.
- **Staff**: 120+ (+50 across 16 start-ups)

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Agenda

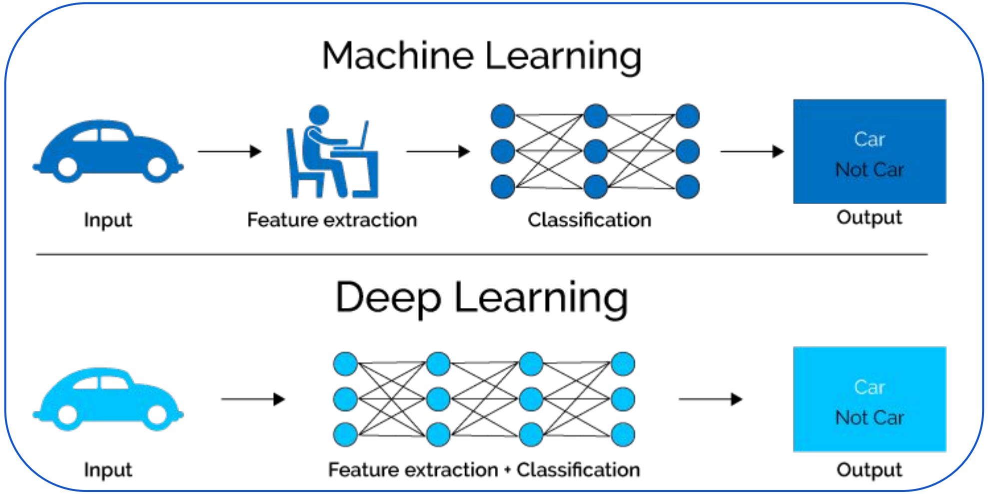
- Overview
- Deep Learning in NLP
- Case Studies
  - Case Study 1 - Machine Translation
  - Case Study 2 - Text Summarization
  - Case Study 3 - Language Detection
  - Case Study 4 - Fake News Detection
  - Case Study 5 - Operant Motive Classification
  - Case Study 6 - Crime Investigation
- Conclusion

Overview

- Natural language processing (NLP) helps computers communicate with humans in their own language and scales other language-related tasks.
- NLP makes it possible for computers to read text, hear speech, interpret it, measure sentiment and determine which parts are important.

What is Deep Learning ?

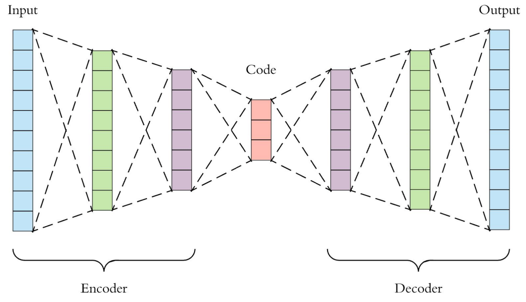
- A machine learning subfield of learning **representations** of data.
- Exceptionally effective at **learning patterns**.
- Deep learning algorithms attempt to learn (multiple levels of) representation by using a **hierarchy of multiple layers.**
- If you provide the system **tons of information**, it learns to respond in useful ways.



https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png

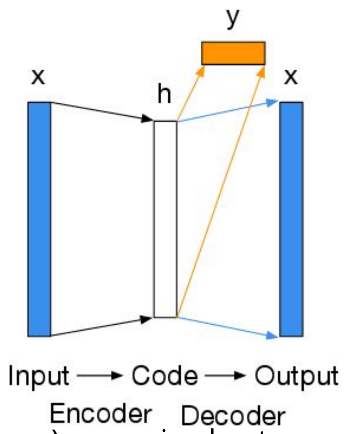
Deep Learning for NLP: Autoencoders

- Learning good representations lies at the core of Deep Learning (DL)
- Over the last few years, DL has made amazing advances in NLP
- Recently, autoencoders represent an alternative to contrastive unsupervised word learning
  - Are able to learn both linear and non-linear transformations
- Autoencoders can discover low-dimensional, less sparse, and robust features for classification

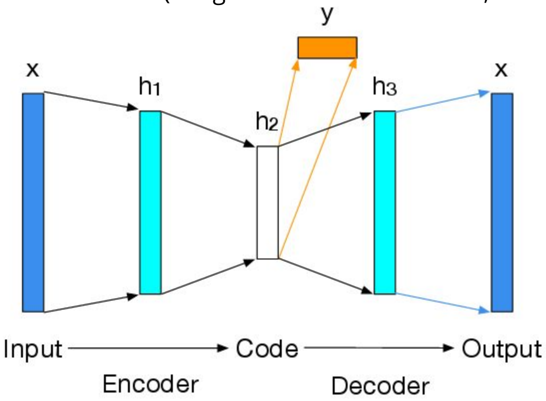


Supervised AutoEncoder

- A **supervised autoencoder** (**SAE**) is an autoencoder with the addition of a supervised loss on the representation layer
- The addition of supervised loss to the autoencoder acts as a regularizer and results better representation for the desired task
- Although **SAE** have been tested on many image classification tasks, they have **not** been extensively tested on **NLP** tasks



(Linear) supervised autoencoder
(Image source: Lei et al. NIPS, 2018)



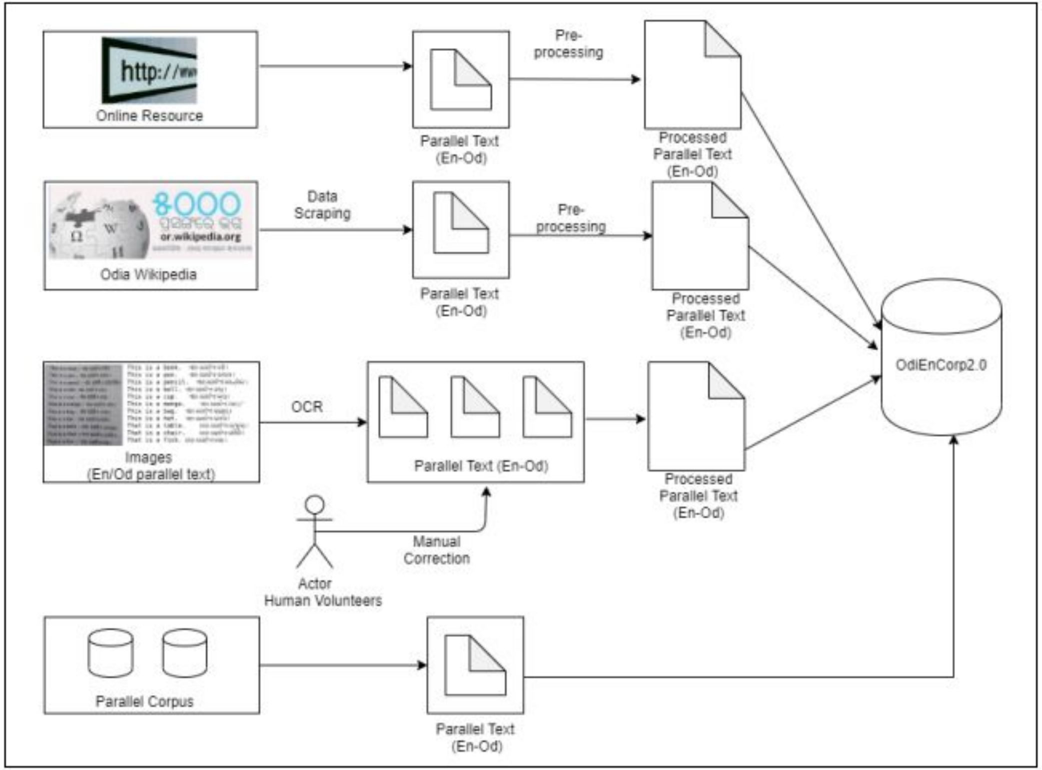
Deep supervised autoencoder
(Image source: Lei et al. NIPS, 2018)

Case 1 : Machine Translation

- Odia is an Indian language belonging to the Indo-Aryan branch of the Indo-European language family.
- Odia is one of 22 official languages of India and sixth Indian language to be designated as a Classical language.
- There is a demand for English↔Odia machine translation system.
- There is lack of Odia resources, particularly parallel corpora.
- Existing few English-Odia corpora are small in size, cover few domains not very suitable for machine translation, which motivates us for OdiEnCorp 2.0.

Data Sources

- Data extracted from other online resources.
- Data extracted from Odia Wikipedia.
- Data extracted using Optical Character. Recognition (OCR).
- Data reused from existing corpora.



Block diagram of the Corpus building process

Data Processing

- Extraction of plain text.
    - Python script to scrape plain text from HTML page.
- Manual processing.
    - Correction of noisy text extracted using OCR-based approach.
- Sentence segmentation.
    - Paragraph segmented into sentences based on English full stop (.) and Odia Danda (|) or Purnaviram.
- Sentence alignment.
    - Manual sentence alignment for Odia Wikipedia articles where text in two language are independent of each other.

Final Datasize and Domain Coverage

- The composition of OdiEnCorp 2.0 with statistics for individual sources.

Source	Sentences	Tokens		Book Name and Author	
		English	Odia	(Parallel)	
Wikipedia Dump	5796	38249	37944	-	General Domain (Wiki data)
Glosbe Website	6222	40143	38248	-	Daily usage learning
Odisha District Website	761	15227	13132	-	General and Tourism Information
TamilCube Website	4434	7180	6776	-	Daily usage learning
OCR (Book 1)	356	4825	3909	A Tiger at Twilight by Manoj Dash	Literature
OCR (Book 2)	9499	117454	102279	Yajnaseni by Prativa Ray	
OCR (Book 3)	775	13936	12068	Wings of Fire by APJ Abdul Kalam with Arun Tiwari	
OCR (Book 4)	1211	1688	1652	Word Book by Shibashis Kar and Shreenath Chaterjee	
OCR (Book 5)	293	1492	1471	Spoken English by Partha Sarathi Panda and Prakhita Padhi	
Odia Virtual Academy (OVA)	1021	4297	3653	Sarala (Tribhasi) Bhasa Sikhana Petika	Daily usage learning
PMIndia	38588	690634	607611	-	Government Policies
OdiEnCorp 1.0	29346	756967	648025	-	Bible, Literature, Government Policies
Total	98302	1692092	1476768		

OdiEnCorp 2.0 parallel corpus details. Training, dev and test sets together

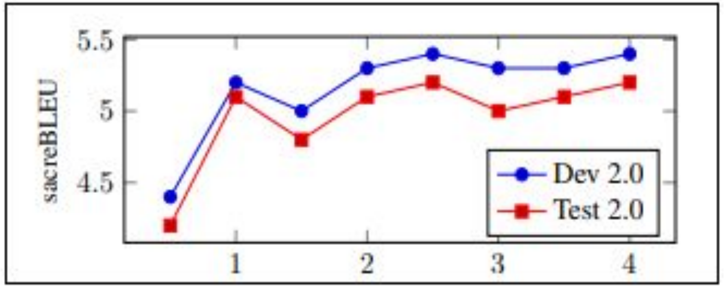
Baseline (Neural Machine Translation)

- Dataset
  - Removed duplicated sentence pairs and shuffled.

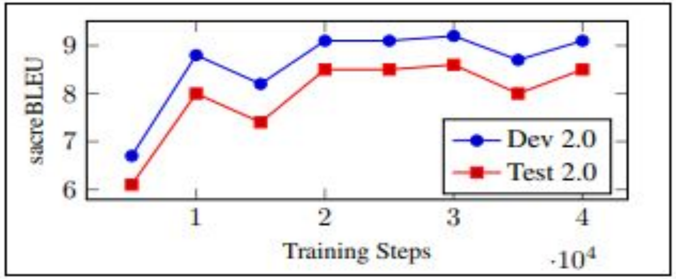
		#Tokens	
Dataset	#Sentences	EN	OD
Train 2.0	69260	1340371	1164636
Dev 2.0	13429	157951	140384
Test 2.0	14163	185957	164532

OdiEnCorp 2.0 processed for NMT experiments.

- NMT Setup
  - We used Transformer model as implemented in OpenNMT-py.
  - Generated vocabulary of 32K sub-word type jointly for source and target language.
  - Train using single GPU (learning rate: 0.2, 8000 warm-up steps).



Learning Curve (EN->OD)



Learning Curve (OD->EN)

Result

		sacreBLEU	
Training Corpus	Task	Dev 2.0	Test 2.0
OdiEnCorp 2.0	EN-OD	5.4	5.2
OdiEnCorp 2.0	OD-EN	9.2	8.6

Results for baseline NMT on Dev and Test sets for OdiEnCorp 2.0.

Availability

OdiEnCorp 2.0 is available for research and non-commercial use under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, CC-BY-NC-SA at :

http://hdl.handle.net/11234/1-3211

WAT 2020

Website: http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2020/index.html

NICT 情報通信研究機構 京都大学 KYOTO UNIVERSITY

WAT 2020
The 7th Workshop on Asian Translation

December, 2020
Suzhou, China
(Hosted by the AACL-IJCNLP 2020)

TRANSLATION TASK

Tasks:

- Scientific paper tasks: Asian Scientific Paper Excerpt Corpus (ASPEC)
  - English <--> Japanese
  - Chinese <--> Japanese
- Patent tasks: Japan Patent Office Patent Corpus 2.0 (JPC2)
  - Chinese <--> Japanese
  - Korean <--> Japanese
  - English <--> Japanese
  - Chinese -> Japanese expression pattern task
- Newswire tasks: JIJI Corpus
  - Japanese <--> English (description)
- News Commentary task:
  - Japanese <--> Russian
- IT and Wikinews tasks:
  - Hindi/Thai/Malay/Indonesian <--> English **(NEW!)** **(Multilingual Multi-domain evaluation task)** **(Collaboration with NICT-SAP)**
- Mixed-domain tasks:
  - UCSY and ALT corpora: Myanmar <--> English
  - ECCC and ALT corpora: Khmer <--> English
- Indic tasks:
  - UFAL (EnOdia) corpus: Odia <--> English **(NEW!)**
  - Bengali/Hindi/Malayalam/Tamil/Telugu/Marathi/Gujarati <--> English **(NEW!)** **(Modification of WAT 2018's Indic Multilingual evaluation task !)**
- Multimodal:
  - English --> Hindi
  - English <--> Japanese **(NEW!)**

ODIANLP Team@WAT2020

System and WAT Task Label	WAT BLEU ODIANLP	WAT BLEU Best competitor
Indic Odia↔English translation task		
ODIAENen-od	**11.07***	9.85
ODIAENod-en	**18.31***	17.89

WAT2020 Automatic Evaluation Results

English to Odia			
Translation Type	English (Source)	Odia (Translation)	Gloss/Remark
Correct Translation	It is located on the bank of the River Sone which merges with River Ganges at Digha a few kilometers from Danapur.	ଏହା ଦାନପୁର ଠାରୁ କିଛି କିଲୋମିଟର ଦୂରରେ ଦିଘାଠାରେ ଗଙ୍ଗା ନଦୀ ସହ ମିଳିତ ହେଉଥିବା ସୋନ ନଦୀର କୂଳରେ ଅବସ୍ଥିତ ।	It is located on the bank of the river Sone which merges with river Ganges a few kilometer away from Danapur
Partial Correct Translation	The temple is maintained by the Bengal, Bihar and Odisha Digambara Jaina Tirthankara Committee Bimala Devi Jain is the local caretaker.	ଏହି ମନ୍ଦିରର ରକ୍ଷଣାବେକ୍ଷଣ ବଙ୍ଗଳା, ବିହାର ଓ ଓଡ଼ିଶା ଦିଗମ୍ବର ଜୈନ ତୀର୍ଥଙ୍କରୀ କମିଟି ବିମଳା ଦେବୀ ଜୈନ ରଙ୍କର ସ୍ଥାନୀୟ କ୍ୟାରିଅର ।	This temple is maintained by the Bengal, Bihar, and Odisha Digambara Jain Tirthankara committee Bimala Devi is the local career. (the word "caretaker" mistranslated into Odia)
Incorrect Translation	donator	ଦାନମାଦାତ	Mistranslated the English word "donator" into Odia

Sample Translation

ODIANLP Team@WAT2020

Odia to English			
Translation Type	Odia (Source)	English (Translation)	Gloss/Remark
Correct Translation	ତେଣୁ ତମେ ତାଙ୍କର ଶତ୍ରୁ ।	So you are his enemy.	Therefore you are his enemy
Partial Correct Translation	'ଭାରତୀୟ ସିନେମା ର ଜନକ' ବୋଲି ଅଭିହିତ କରାଯାଉଥିବା ଦାଦାସାହେବ ଫାଲକେ ଭାରତ ର ପ୍ରଥମ ପୂର୍ଣ ଦୀର୍ଘ ଚଳଚିତ୍ର ' ରାଜା ହରିଶ୍ଚନ୍ଦ୍ର' ନିର୍ମାଣ କରିଥିଲେ ୧୯୧୩ ମସିହା ରେ	Dadasaheb Phalke, who is described as the "father of Indian Cinema", built the first long film of India" in 1913.	Dadasaheb Phalke who called as "The father of Indian Cinema" build the first full length cinema "Raja Harishchandra" in 1913 (the movie name “Raja Harishchandra” missing)
Incorrect Translation	ଉଦ୍ଧୃତି ଚିହ୍ନ ପ୍ରାରମ୍ଭ	Open number	Begin of quotation mark (mistranslated the Odia word “ଉଦ୍ଧୃତି ଚିହ୍ନ”)

Case 2 - Text Summarization

(Usage of Synthetic Data for Text Summarization)

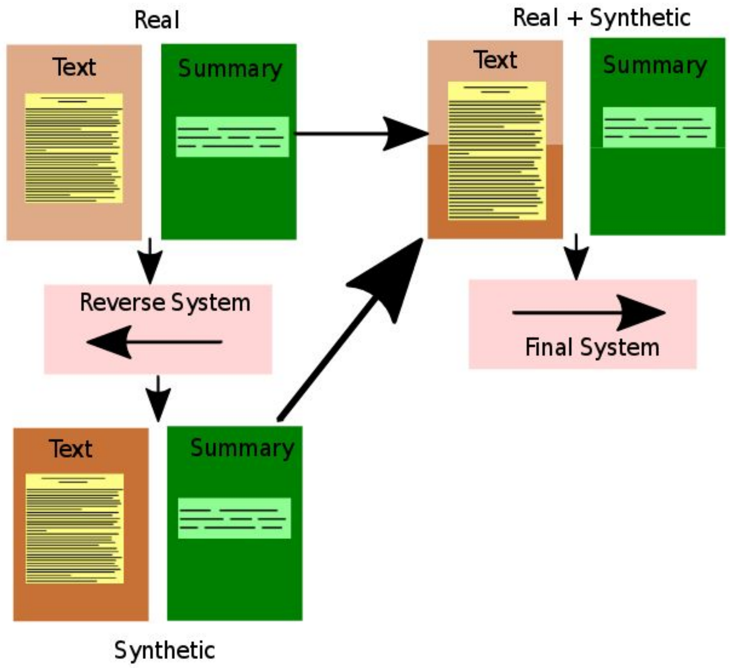
- Based on Idiap participation in the SwissText 2019 challenge (100’000/2’000) paragraphs and summaries for training/evaluation.

- *Use of synthetic data*: a popular approach in machine translation for the low resource conditions to improve the quality.

- Can such approaches work for the text summarization task ?.

Method

- Use a state-of-the-art "Transformer Model" as implemented in OpenNMT-py.
- Different experiments performed based on real and synthetic data.
- Synthetic data used to increase the size of the training data.
- To generate synthetic data :
    1. A system is trained in reverse direction i.e. source as summary and target as text.
    2. The reverse system is used to generate text for the given summary. Now, synthetic data is ready.
    3. Mix the real and synthetic data and train the final system.



Generation of synthetic data using reverse system.

Dataset

- Real data (SwissText dataset)

Dataset	#Text	#Summaries
Train	90K	90K
Dev	5K	5K
Test	5K	5K
Test Evaluation	2K	-

Statistics of experimental data (real) including the number of text and summaries.

- Synthetic data (Common Crawl)
    1. Build Vocabulary (using SwissText dataset, most frequent German words).
    2. Select sentences based on the prepared Vocabulary. From the selected sentences, randomly choose 100K.
    3. Generate synthetic data by using 100K sentences to input to the reverse trained model.

Dataset	#Text	#Summaries
Train	190K	190K

Statistics of experimental data (real + synthetic) including the number of text and summaries.

Evaluation

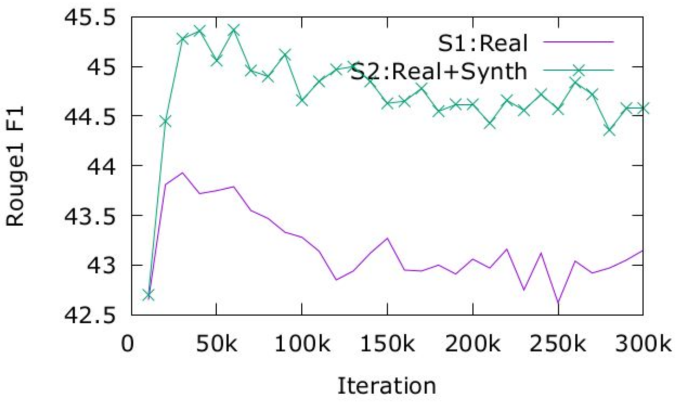
Setting	Dataset	Rouge_1_F1	Rouge_2_F1
S1	Dev	43.9	28.5
	Test	39.7	22.9
S2	Dev	45.4	29.8
	Test	55.7	41.8

Evaluation results of our models

Team	Rouge_1	Rouge_2
Shantipriya Parida, and Petr Motlicek (s2)	40.2	22.2
Dmitrii Aksenov, Georg Rehm, Julian Moreno Schneider	40.4	21.9
Nikola Nikolov	34.7	19.3
Valentin Venzin, Jan Deriu, Didier Orel, Mark Cieliebak	39.8	23.4
Pascal Fecht	40.9	23.5

SwissText 2019 Text Summarization Challenge Result

Source: http://ceur-ws.org/Vol-2458/summarization_challenge.pdf



Learning curves in terms of Rouge 1 F1 Score on dev set

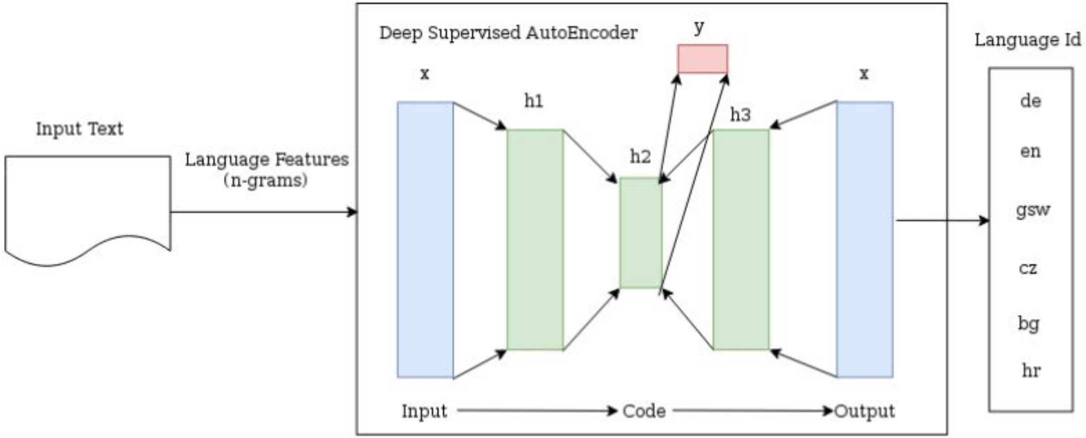
- Evaluations made using Rouge (Recall-Oriented Understudy for Gisting Evaluation) score, a popular metric for text summarization.

Case3: Language and Dialect Detection

- Its challenging to detect languages that have similar origins or dialects (e.g. German dialect identification, Indo-Aryan language identification)
- It may not be possible to distinguish related dialects with very similar phoneme and grapheme inventories for some languages.

Method Description

- We used character n-gram for extracting features from the input text.
- Extracted features are input to the deep supervised autoencoder (SAE).
- Bayesian optimizer used for selecting the optimal hyperparameters.



Proposed model architecture

Dataset

DSL Dataset: Discriminating between Similar Language (DSL) contains 13 different languages based on 6 different language group. We used DSLCCv2.0 in our experiment.

Ling10 Dataset: It contains 190,000 sentences categorized into 10 languages (*English, French, Portuguese, Chinese Mandarin, Russian, Hebrew, Polish, Japanese, Italian, Dutch*).

ILI Dataset: The Indo-Aryan Language Identification (ILI) dataset contains 5 closely-related languages of the Indo-Aryan language family – Hindi (also known as Khari Boli), Braj Bhasha, Awadhi, Bhojpuri, and Magahi.

Group Name	Language	Id
South Eastern Slavic	Bulgarian Macedonian	bg mk
South Western Slavic	Bosnian Croatian Serbian	bs hr sr
West-Slavic	Czech Slovak	cz sk
Ibero-Romance(Spanish)	Peninsular Spain Argentinian Spanish	es-ES es-AR
Ibero-Romance(Portuguese)	Brazilian Portuguese European Portuguese	pt-BR pt-PT
Astronesian	Indonesian Malay	id my

DSL Language Group. Similar languages with their language code.

Result

Dataset	Training	Development	Test
DSL	252,000	28,000	14,000
Ling10	140,000	-	50,000
ILI	70,351	10,329	9,692

Dataset Statistics

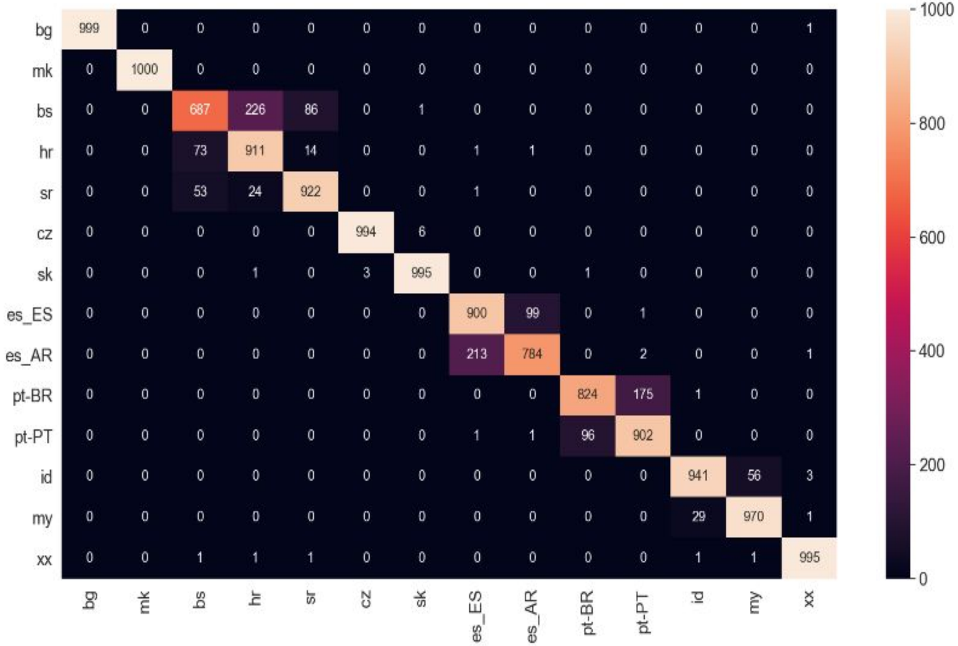
Model	Dataset	Accuracy (Test Set)
SAE (char-3gram)	Ling10	100%
SAE (char-3gram)	DSL	92%
SAE (char-3gram)	ILI	85%

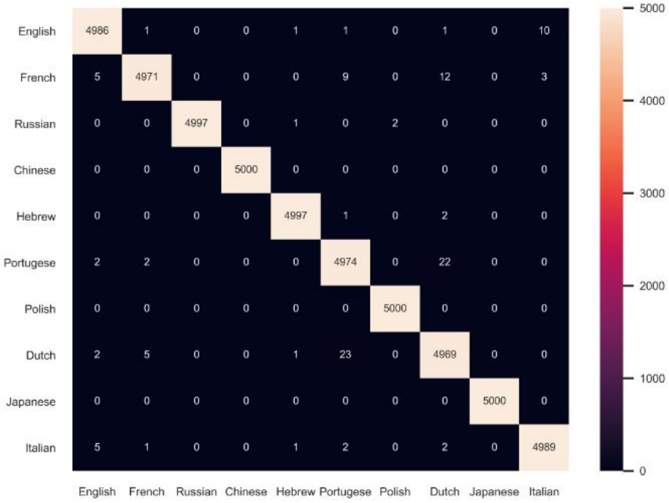
Performance on test dataset.

Parameter	DSL	Ling10	ILI
ngram-range	1-3	1-3	1-3
number of target	14	10	5
embedding dimension	300	300	300
supervision	'clf'	'clf'	'clf'
converge threshold	0.00001	0.00001	0.00001
number of epochs	300	500	500

SAE model configurations for the dataset.

Result (Confusion Matrix)





Case 4: Fake News Detection @MEX-A3T

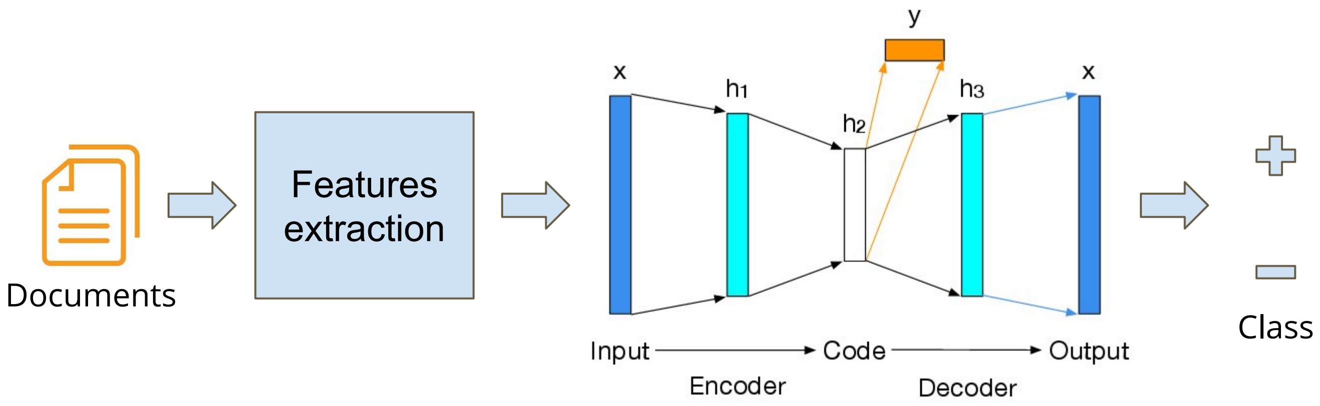
- The goal of IberLEF is to encourage the research community to organize competitive text processing, understanding and generation tasks in order to define new research challenges and setting new state-of-the-art results for the Natural Language Processing community, involving at least one of the following Iberian languages: Spanish, Portuguese, Catalan, Basque or Galician
- MEX-A3T 2020 had the following tracks:
  - Fake News detection
  - Aggressiveness detection
  - Both tracks contain documents in Mexican Spanish

Fake News Detection

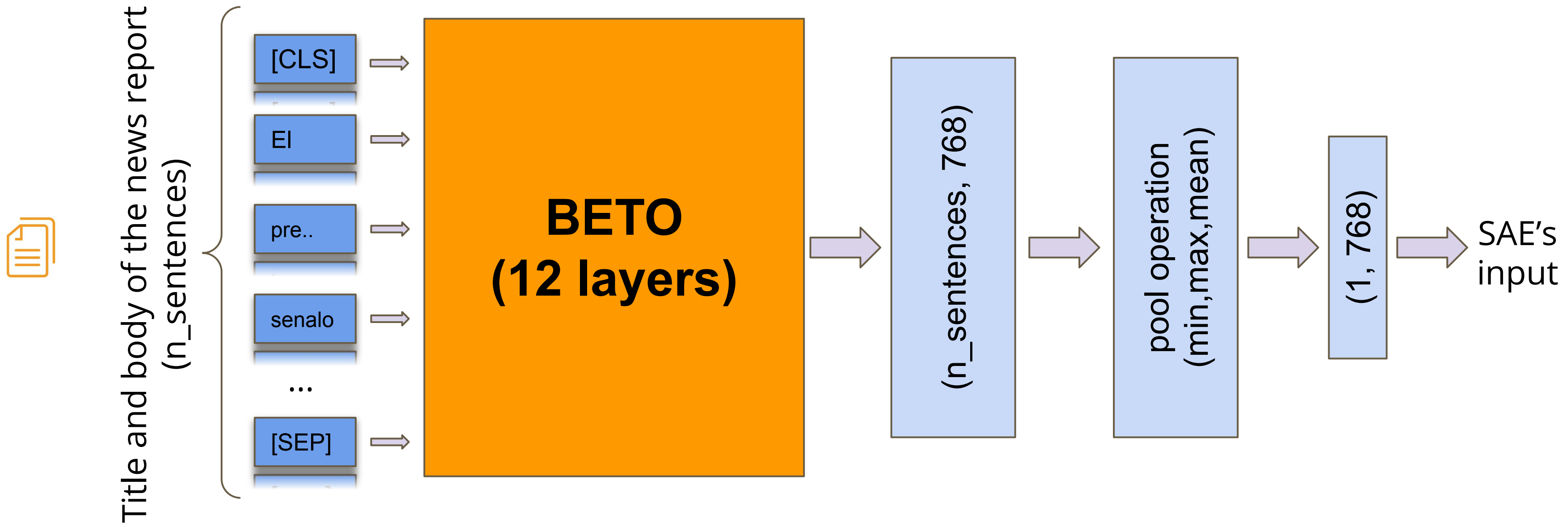
- **Fake news** provides information that aims to manipulate people for different purposes: terrorism, political elections, advertisement, satire, among others
- In social networks, misinformation extends in seconds among thousands of people
- A fake news detection system aims to help users detect and filter out potentially deceptive news
- The dataset consist of 971 documents, 676 for training and 295 for test
  - Documents are real news extracted from differents news media in Mexico

Methodology

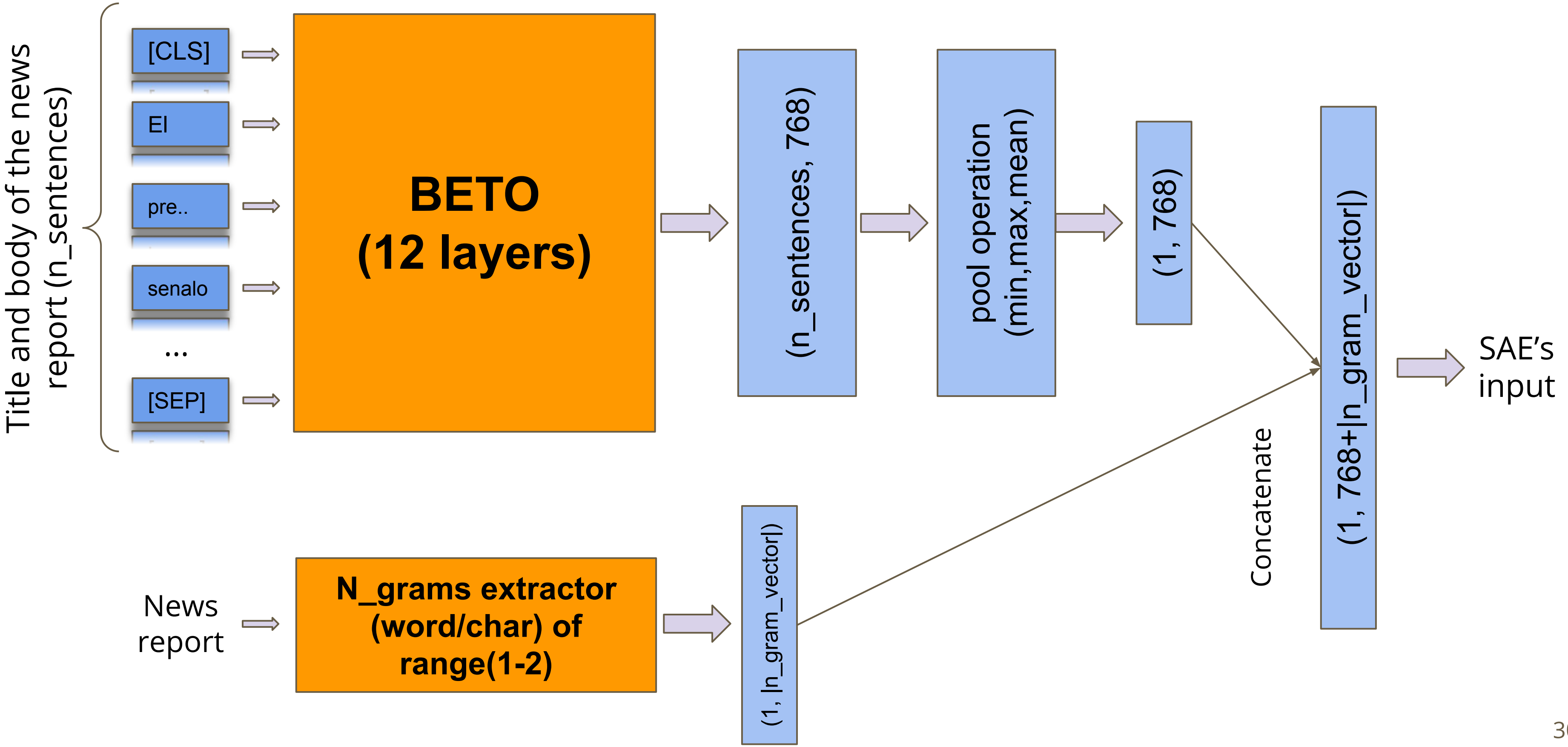
- Our goal was to evaluate the pertinence of deep SAE in these tasks
- As input features we used:
  - Spanish pre-trained **BERT** encodings (BETO)
  - Traditional text representation techniques such as **word** and **char n-grams** (ranges 1-2 and 1-3)
  - Combinations of BETO encodings plus traditional words/char n-grams vectors



Features extraction



Features extraction



Results (fake news task)

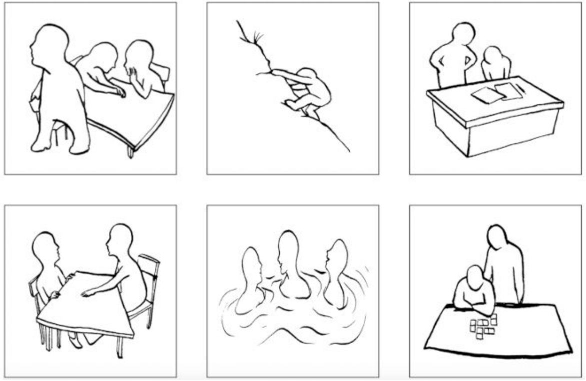
The best performance among 6 participating institutions

Table 3. Results in validation and test phases reported in F-score for
real-news (F-), and macro average of F-score (Fm).

		Validation phase			*Test phase*			
Input features	min-df,max-df	Fm	F+	F-	ID	Fm	F+	F-
W(1,2)	0.01, 0.5	0.775	0.793	0.758	-	-	-	-
W(1,3)	0.01, 0.5	0.778	0.798	0.758	-	-	-	-
C(1,2)	0.01, 0.5	0.697	0.719	0.674	-	-	-	-
C(1, 3)	0.01, 0.5	0.757	0.768	0.745	-	-	-	-
B(min-pooling)		**0.843**	0.842	0.845	2	**0.856**	**0.844**	**0.868**
B(max-pooling)		0.830	0.830	0.830	-	-	-	-
B(mean-pooling)		0.833	0.831	0.835	-	-	-	-
C(1, 3)+W(1,2)	0.01, 0.5	0.805	0.807	0.802	-	-	-	-
B+W(1,2)	0.01, 0.3	**0.845**	0.846	0.844	1	0.850	0.840	0.859
B+C(1,3)	0.01, 0.3	0.834	0.834	0.835	-	-	-	-
B+W(1,2)+C(1,3)	0.01, 0.3	0.833	0.831	0.835	-	-	-	-
B+W(1,2)+C(1,3)	0.01, 0.5	**0.848**	0.846	0.850	-	-	-	-
Third best system (in the track)						0.817	0.819	0.817
BOW-RF (baseline-given by track organizers)						0.786	0.785	0.787

Case 5: Operant Motive Detection

- According to Psycholinguistics theory, how we use language reveals information about our personality traits, educational level, age, etc.
- Operant methods are psychometrics, which are captured by having participants write free texts associated with faint images
  - Clinical research indicates that operant motives provide the possibility to assess behavioral long-term developments
- M - power, A - affiliation, L - achievement, F - freedom, 0 - zero, and corresponding levels (0 to 5).



Sample images that are shown to subjects during the OMT test

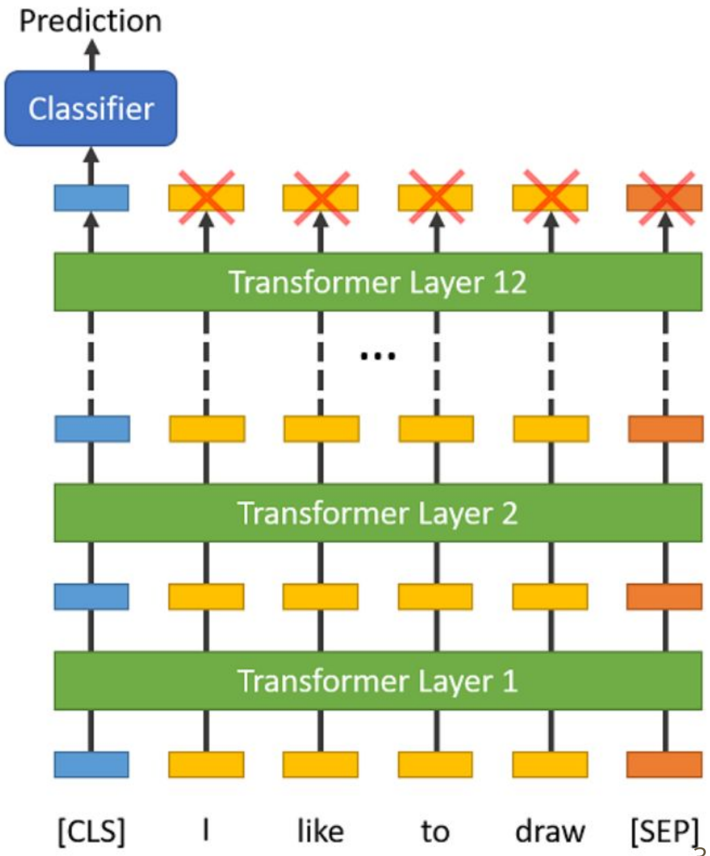
Task details

- The task is to predict motivational style solemnly based on tex
- The dataset:
  - Language: German
  - Training: 167,200*
  - Development: 20,900
  - Test: 20,900
  - Highly unbalanced

	Training	
	Average (σ)	*Total*
Tokens	20.27 ($\pm$12.08)	3,389,945
Vocabulary	18.07 ($\pm$9.82)	267,620
LR	0.92 ($\pm$ 0.08)	0.08
	Development	
	Average (σ)	*Total*
Tokens	20.38 ($\pm$12.17)	425,880
Vocabulary	18.17 ($\pm$9.94)	55,606
LR	0.92 ($\pm$ 0.08)	0.13
	Test	
	Average (σ)	*Total*
Tokens	20.24 ($\pm$12.01)	423,018
Vocabulary	18.05 ($\pm$9.76)	55,592
LR	0.92 ($\pm$0.08)	0.13

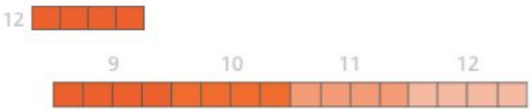
Methodology(1/3)

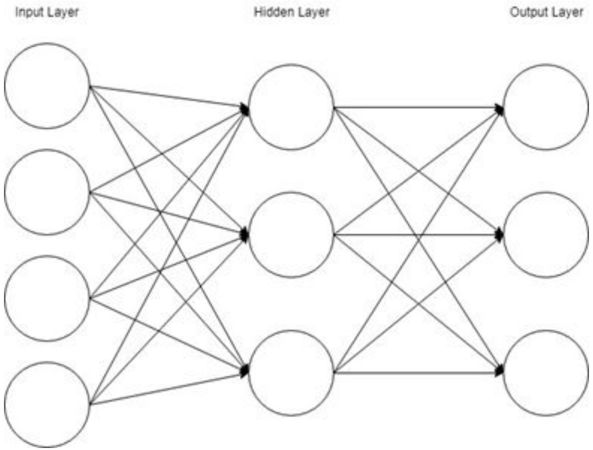
- **Simple transformers**: we add an untrained layer of neurons on the end, and re-train the model with the OMT classification task at the output
- **max_length** parameter is set to **90**, and models are re-trained up to **2 epochs**
- Three different configurations:
    - BERT
    - XLM
    - DistilBERT



Methodology(2/3)

- **Fully connected neural network (FC):** the FC is feed with the representation of the textual descriptions using:
    - **Pre-train** BERT
    - **Fine-tuned** BERT
- We reported results using two distinct ways for building the sentences representation
    - Last Hidden Layer
    - Concat Last Four Hidden

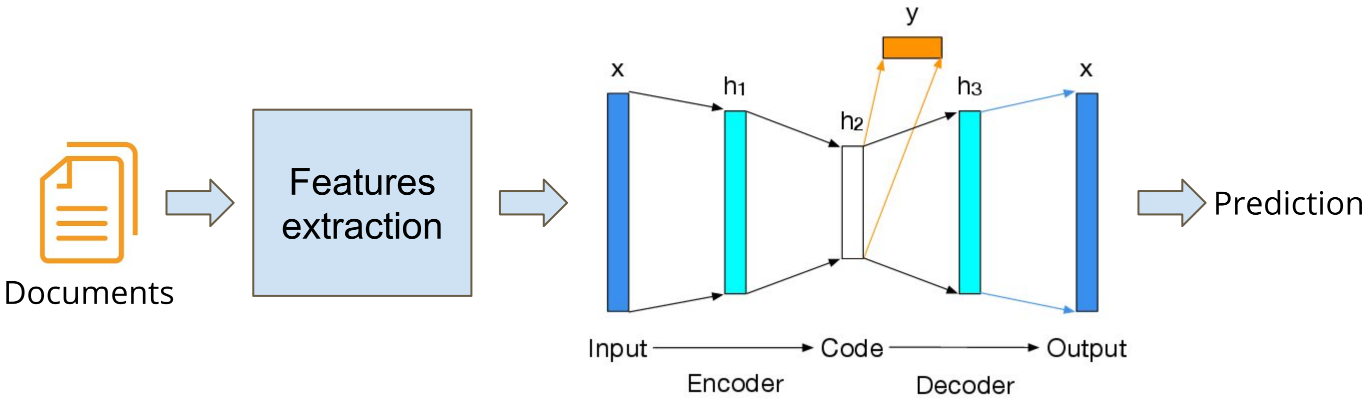




Hyper Parameter	Range
number of layers	3
number of hidden layers	1
nodes in hidden layer	16
activation function	ReLU

Methodology(3/3)

- We evaluate the performance of deep supervised autoencoders in the OMT task
- As input features we used:
    - German pre-trained and fine-tuned BERT encodings
    - Traditional text representation techniques such as word and char n-grams (ranges 1-2 and 1-3)
    - Combinations of BERT encodings plus traditional words/char n-grams vectors



IDIAP RESEARCH INSTITUTE

Results (*test phase*)

The 2nd best performance among 3 (official) participating institutions

Method	*Configuration type*	*Configuration sub-type*	**F1-macro** (*dev*)	**F1-macro** (*test*)
ST	Bert	bert-base-german-cased	**0.694**	**0.698**
ST	XLM	xlm-mlm-ende-1024	0.688	0.686
ST	DistilBert	distilbert-base-german-cased	0.692	0.688
FC	Bert (pre-trained)	LHL	0.589	0.589
FC	Bert (pre-trained)	Concat4LHL	0.616	0.579
FC	Bert (fine-tuned)	LHL	0.673	0.671
FC	Bert (fine-tuned)	Concat4LHL	0.675	0.230
Baseline	SVM	*tf-idf*	0.639	0.644
1st place	–	–	–	0.704

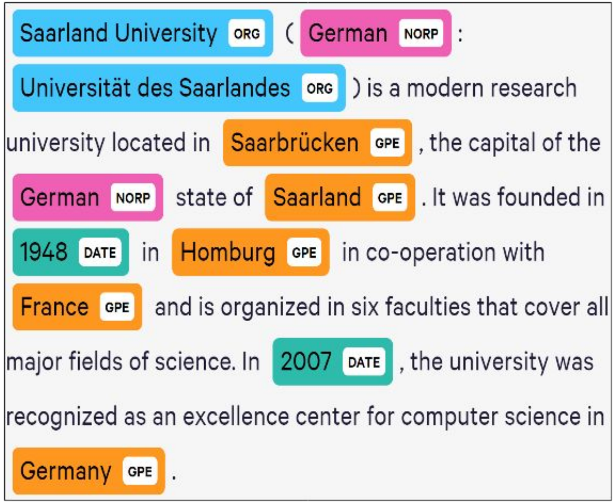
Case 6 - Crime Investigation

ROXANNE

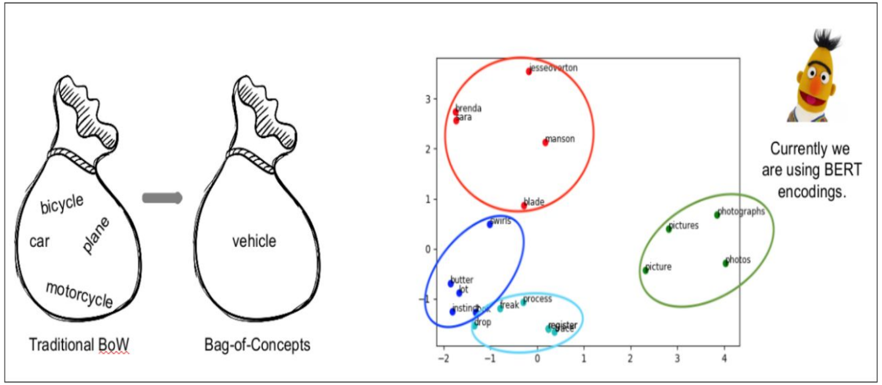
(Real time netwOrk, teXt, and speaker ANalytics for combating orgaNized crimE)

- FCT H2020 project (http://roxanne-euproject.org).
- Various NLP technologies applied (potentially including summarization, entity detection, and topic detection).

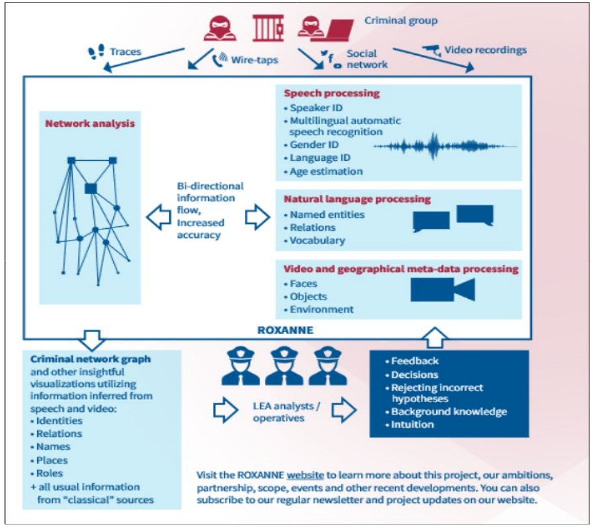
Entity Detection



Topic Detection



Project Overview



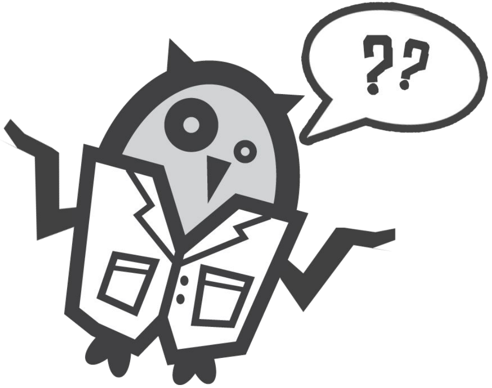
Conclusions and future work

- SAE with Bayesian Optimization for the language detection task found effectively for discriminating between very close languages or dialects
- SAE are able to generalize well, however, they seem to perform better on texts extracted from formal written
    - Fake news detection, best performance, documents extracted from real news media
- SAE are less computationally expensive as compared to attention based DL models (e.g., transformers)
    - They do not require high volume of data

Conclusions and future work

- If there are plenty of data, fine-tuning of previous attention based DL models is an immediate and promising solution
  - This was the case for the OMT task
- We would like to evaluate the impact of hyper-parameter tuning
- Test **SAE** on other NLP tasks, e.g., topic detection and topic tracking
- Extending OdiEnCorp 2.0 with more parallel data, again by finding various new sources.

Q&A



Contact information:

- Email: shantipriya.parida@idiap.ch
- Twitter: @Shantipriyapar3
- Web : https://www.idiap.ch/~sparida/

References

[1] Parida, S., Dash, S. R., Bojar, O., Motlıcek, P., Pattnaik, P., & Mallick, D. K. **OdiEnCorp 2.0: Odia-English Parallel Corpus for Machine Translation.** In *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020* (p. 14).

[2] Parida, S., Villatoro-Tello, E., Kumar, S., Motlicek, P., & Zhan, Q. (2020). **Idiap Submission to Swiss-German Language Detection Shared Task.** In Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS).

[3] Villatoro-Tello, E., Parida, S., Kumar, S., Motlicek, P., & Zhan, Q. (2020). **Idiap & UAM participation at GermEval 2020: Classification and Regression of Cognitive and Motivational Style from Text**. In Proceedings of the GermEval 2020 Task 1 Workshop in conjunction with the 5th SwissText & 16th KONVENS Joint Conference (pp. 11-16).

[4] Parida, S., Villatoro-Tello, E., Kumar, S, Fabien, M. & Motlicek, P. (2020). "**Detection of Similar Languages and Dialects Using Deep Supervised Autoencoders**". In Proceedings of the 17th International Conference on Natural Language Processing, ICON2020.

[5] Parida, S., Motlicek, P. "**Abstract Text Summarization: A Low Resource Challenge**". In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 5996-6000, 2019.

[6] Ramírez-de-la-Rosa, G., Kumar, S., Parida, S., Motlicek, P. **Idiap and UAM Participation at MEX-A3T Evaluation Campaign**. Esaú Villatoro-Tello, Gabriela Ramírez-de-la-Rosa, Sajit Kumar, Shantipriya Parida, Petr Motlicek. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020), CEUR WS Proceedings vol. 2664 (2020)

Prof. Esá u Villatoro-Tello
Idiap Research Institute, Switzerland
And
UAM Mexico

Assoc. Prof. Ondřej Bojar
Charles University, Czech Republic

Dr. Shantipriya Parida
Idiap Research Institute, Switzerland

Dr. Petr Motlicek
Idiap Research Institute, Switzerland

Thanks to all of our collaborators

Assoc. Prof. Satya Ranjan Dash
KIIT University, India

Debasish Kumar Mallick
KIIT University, India

Sajit Kumar
GreatLearning, India

Priyanka P. Pattnaik
Tech Mahindra, India

Thank You