Supervised Learning for the ICD-10 Coding of French Clinical Narratives

Clément Dalloux[1], Vincent Claveau[1], Marc Cuggia[2], Guillaume Bouzillé[2] and Natalia Grabar[3]

[1] Univ Rennes, Inria, CNRS, IRISA, F-35000 Rennes, France
[2] Univ Rennes, CHU Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France
[3] UMR 8163 STL CNRS, Université de Lille, France

Context

- Programme de médicalisation des systèmes d’information (PMSI) :
  - Started in the 80’s as part of the reform of the French health system
  - 1982-2005 : epidemiology, monitoring health institutions’ activity
  - 2005-now : activity based financing to reduce inequalities of resources between health institutions
- Health facilities manually assign codes to reports for billing/funding purposes
- Aim: computer-aided coding of clinical narratives

ICD-10

- International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10)
    - Standardized, structured representation of clinical texts used for the billing of medical procedures, epidemiology, internal management, etc.
    - up to 16,000 codes representing numerous diagnostics and clinical and social situations

ICD-10

- Hierarchical structure, codes with up to 4 digits
- A lot of codes do not have children:
  - A38 Scarlet fever
  - B03 Smallpox
  - D70 Agranulocytosis
  - I10 Essential (primary) hypertension

- **ICD-10 Version:2016**
  - I Certain infectious and parasitic diseases
    - **A00-A09** Intestinal infectious diseases
      - **A00** Cholera
        - **A00.0** Cholera due to Vibrio cholerae 01, biovar cholerae
        - **A00.1** Cholera due to Vibrio cholerae 01, biovar eltor
        - **A00.9** Cholera, unspecified
      - **A01** Typhoid and paratyphoid fevers
        - **A01.0** Typhoid fever
        - **A01.1** Paratyphoid fever A
        - **A01.2** Paratyphoid fever B

Related work

- CLEF eHealth 2016-2018 shared task, classification of death certificates
  - French datasets : raw and aligned, stored in .csv files
  - 1 line per row and 3 lines, 10 tokens and 4 ICD codes per document on average
  - Best neural network system: (Atutxa et al., 2018), **0.838 F_1** (aligned) and **0.709 F_1** (raw)
  - Best rule based system: (Cossin et al., 2018), **0.670 F_1** (raw)
  - These systems are sentence based, **not multi-label**
- Are death certificates free or structured texts?

Related work

Line	Text	Normalized text	Codes
1	choc septique	choc septique	A41.9
2	peritonite stercorale sur perforation colique	peritonite stercorale	K65.9
2	peritonite stercorale sur perforation colique	perforation colique	K63.1
3	Syndrome de détresse respiratoire aiguë	Syndrome détresse respiratoire aiguë	J80.0

A sample document from the CépiDC French Death Certificates Corpus: aligned dataset (Névéol et al., 2018)

Material

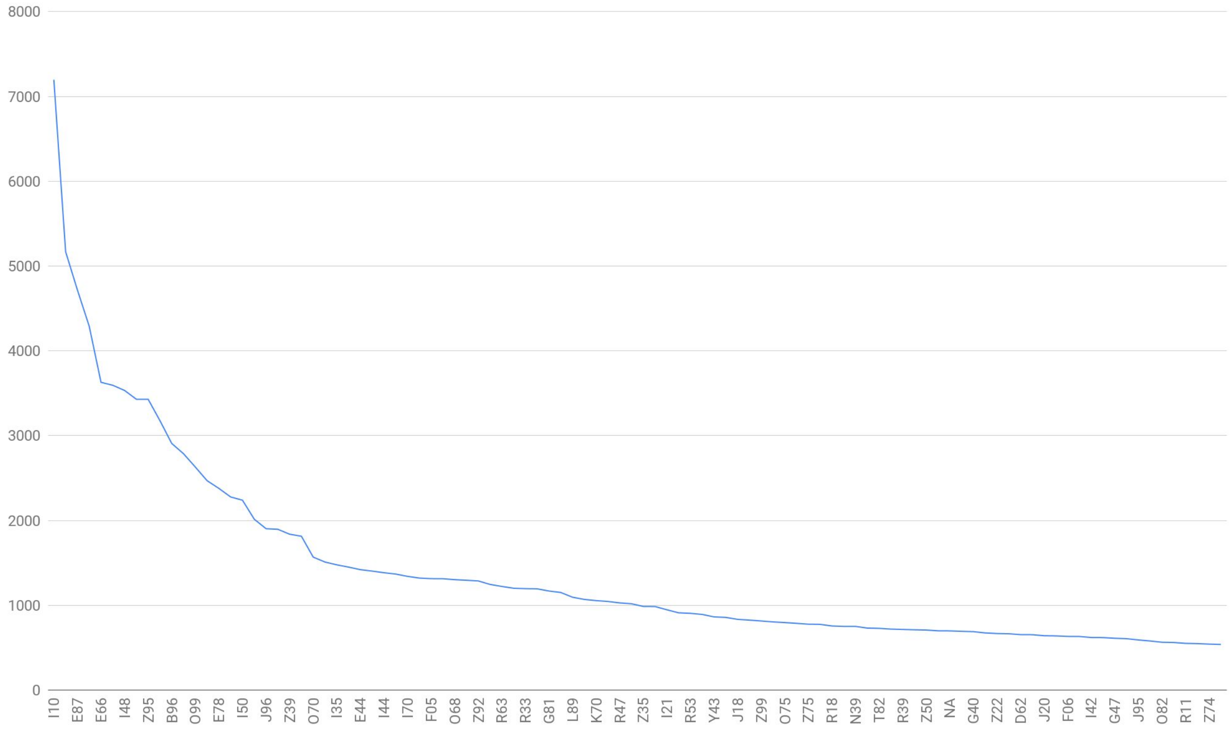
- CHU Server with 2 NVIDIA Tesla V100 GPUs
- Rennes University Hospital dataset :
    - 28,000 hospitalization records with the matching ICD-10 codes documents from all medical specialties (cardiology, urology, oncology, ophthalmology, etc.)
    - A typical document describes: reasons for hospitalization, patient's medical history, sometimes family history, laboratory measurements and findings, treatment, length of stay, release
    - Data limitations: changes in coding happen periodically (**ICD-11**)

Material

- **6,116 unique ICD-10 codes**, 400+ not included in the train set
- **1,345 tokens** per document on average (10 for Clef eHealth)
- 8 unique codes per document on average (4 for Clef eHealth)

	Training	Development	Test
Documents	22,400	2,800	2,800
Tokens	30,307,091	3,718,715	3,654,599
Total ICD codes	182,112	23,010	22,561
Unique ICD codes	5,735	2,824	2,887
Unique unseen ICD codes		191	212
Codes < 10 examples	3,885	2,320	2,391
Codes > 100 examples	382	32	31

Distribution of codes



Multi-label classification of French clinical narratives

- Task 1: with all available ICD-10 codes, **6,116 classes**
  - Subtasks: with the top 1,000 and 100 classes
- Task 2: with codes from A00 to Z99, **1,549 classes**
  - higher hierarchical level (A00.0 ➔ A00)
  - still relevant for health institutions
  - Subtask: with the top 100 classes

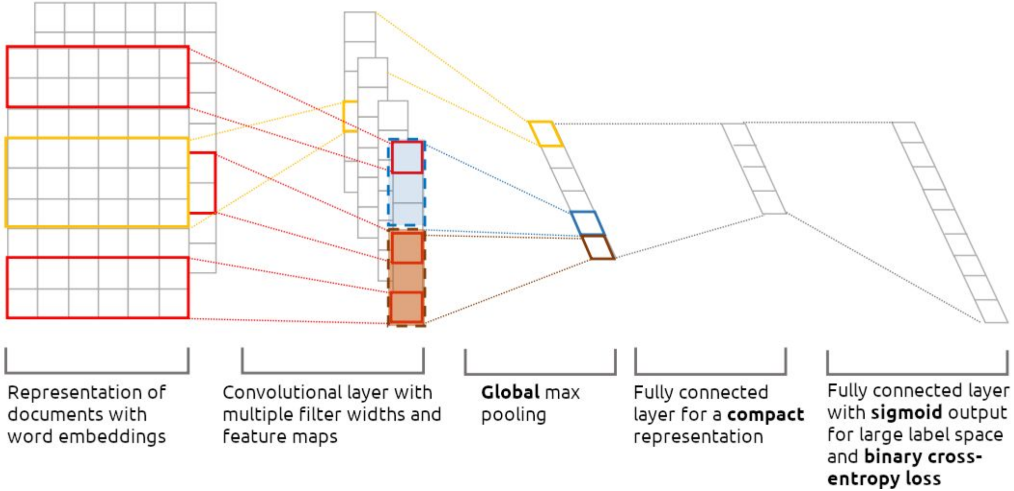
Methods

- Baselines:
  - Dictionary based system
- Deep learning systems:
  - Bidirectional Long Short-Term Memory neural network
  - Convolutional neural network
  - Word vectors: fastText vectors trained on our dataset with 300 dimensions

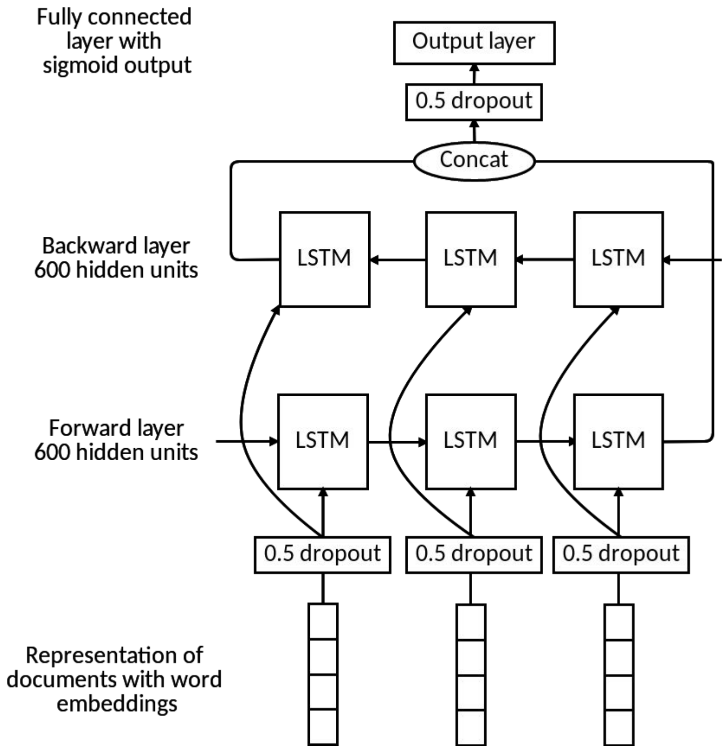
Dictionary based system

- Clef ehealth 2018 (Cossin et al., 2018):
  - Uses the ICD-10 terminologies to assign codes to each text line
  - runs Levenshtein distance and a module of synonyms expansion
- Our system uses:
  - Clef eHealth’s dictionnaire2015.csv to assign codes (147,342 entries)
  - Levenshtein distance between dictionary entries and “sentences”

Neural networks



Best parameters: Filter width = [2,3,4,5]; Feature maps: 2000; FCL: 2000 hidden units



Results

- The rule based system yields mediocre results
- The CNN yields the best results by far
- Much lower results than the systems presented at CLEF eHealth 2018, same task?

	System	Precision	Recall	F_1
Task 1	Dict	0.0722	0.0356	0.0476
	BiLSTM	0.6513	0.1387	0.2287
	CNN	0.5029	0.3301	**0.3986**
Task 1 top 1000c	CNN	0.5604	0.3839	0.4557
Task 1 top 100c	CNN	0.6317	0.5667	0.5974
Task 2	Dict	0.0522	0.0725	0.0607
	BiLSTM	0.6309	0.2837	0.3914
	CNN	0.6018	0.4625	**0.5230**
Task 2 top 100c	CNN	0.6666	0.5873	0.6245

Conclusions

- First experiments on ICD-10 coding of French clinical **narratives**
- Multi-label classification experiments with several deep learning algorithms
  - CNN yields the best results by far
  - Using a still relevant higher hierarchical level works better
- Can a dictionary based expert system on narratives work?
  - Implement a better sentence detection method
  - Implement a better matching method than Levenshtein distance

Perspectives

- Building a bigger dataset from the CHU database
  - Adding all available structured data as additional input to our system
  - Training BERT, Flair, ELMo embeddings on all texts from the CHU database
- Adding more preprocessing to clinical narratives
  - medical history detection, negation detection, etc.
- Using Perform a thorough error analysis:
  - Does the prediction fail or are the documents lacking codes?

References

Atutxa, A., Casillas, A., Ezeiza, N., Fresno, V., Goenaga, I., Gojenola, K., and Perez-de Viñaspre, O. (2018). Ixamed at clef ehealth 2018 task 1 : Icd10 coding with a sequence-to-sequence approach.

Cossin, S., Jouhet, V., Mougin, F., Diallo, G., and Thiessard, F. (2018). Iam at clef ehealth 2018 : Concept annotation and coding in french death certificates.

Névéol, A., Robert, A., Grippo, F., Morgand, C., Orsi, C., Pelikan, L., ... \& Zweigenbaum, P. (2018, September). CLEF eHealth 2018 Multilingual Information Extraction Task Overview: ICD10 Coding of Death Certificates in French, Hungarian and Italian. In CLEF (Working Notes).