



MaChAmp: Multi-task Learning to the Rescue in Resource Scarce Scenarios



Benchmarks in Natural Language Processing (NLP)

THE WALL STREET JOURNAL.

© 1981 Dow Jones & Company, Inc. All Rights Reserved.

VOL. CXXVII NO. 14 ★★

EASTERN EDITION

WEDNESDAY, JANUARY 21, 1981

PRINCETON, NEW JERSEY

35 CENTS

After the Crisis

Torn U.S.-Iranian Ties Won't Heal for a While Despite Hostage Pact

Mutual Recriminations Seem Likely, but in Long Run 'Normal' Relations Seen

'Stay Apart, Let Time Pass'

By KAREN ELIZABETH HOWE
Staff Reporter of The Wall Street Journal

WASHINGTON—The Iran hostage crisis has so polarized U.S.-Iranian relations that no one believes U.S.-Iranian relations is over, but its bitter residue will affect American foreign policy for the indefinite future.

After nearly 15 months of captivity, the 53 Americans held for 14 months remain for Algeria. In return, the U.S. released about \$1 billion if Iran's assets frozen in U.S. banks were returned to its owners by President Carter on Nov. 14, 1979.

That swap could eventually open the way for a more effective American policy to

For details of the hostages' release, see story on page E.

check Soviet influence in Iran, long a buffer between the two nations. The U.S. is the world's third largest oil producer, and President Reagan assumed office yesterday without the preoccupation with Iranian affairs which has shocked American foreign policy continually in the Middle East, for over a year.

Most of the world understood this was of

What's News—

Business and Finance

THREE MAJOR BANKS reported mixed results for the fourth quarter. Citicorp operating net plunged 20% to \$36 million despite a \$37 million after-tax gain from the real estate investors Hanover had a 12% increase to \$39.5 million, and Wells Fargo's profit rose 2% to \$84.4 million.

Fifth Pennsylvania Corp. had a fourth-quarter loss of \$4.6 million, due mainly to a \$3.6 million expansion of loan losses provisions and \$11.5 million to settle holder suits. The bank holding company's loss for the year totalled \$164.1 million.

(Starts on Page D)

American Telephone and the government may both be a victory in the "tug-of-war" reached to settle their antitrust case. In return for

its part in Westinghouse Electric

and some operating units, AT&T

may be allowed to enter the data processing field.

(Starts on Page E)

Cincinnati Bell canceled a \$6 million note owing it a day before the company was set to be delivered to investors. The surprise move came after reporters that a pending settlement

World-Wide

Agencies

THE HOSTAGES FLEW to Algiers as President Reagan huddled his Cabinet.

Two Algerians left Tehran at 11:30 p.m. Tuesday, ending a 14-month captivity of the former captives and Algiers digested in Algiers, the 52 Americans were welcomed by U.S. Ambassador John R. Bolton, the general manager of the airport, a 4-star general and the hostages were jubilant leaving Iran, where they had been detained by U.S.

The returnees said the Carter administration cleared up a little dispute over an appendix to the U.S.-Iranian framework agreement.

UNEMPLOYMENT among married men fell to 4.3% of the labor force in December, the lowest rate since November 1973, the Labor Department reports.

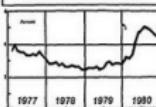
The returnees said the Carter administration cleared up a little dispute over an appendix to the U.S.-Iranian framework agreement.

Alors said Carter speculated that Iran delayed the hostages' departure until after the election to give the new president time to "cavort here our own entertainment."

Families cheered and wept at word of their captive relatives' return. Bells at the White House rang out as news of the release of the last 52 Americans reached the White House. In Algiers, N.Y., civil defense sirens blared for 45 seconds every hour for each day the Americans were held prisoner.

BY LAWRENCE DRAGGIA
Staff Reporter of The Wall Street Journal

Jobless Married Men



UNEMPLOYMENT among married men fell to 4.3% of the labor force in December, the lowest rate since November 1973, the Labor Department reports.

Live, From St. Paul,
Here's 'A Prairie Home Companion'

A Lazy Way of Music
And Small-Town Humor
Scores Big on Public Radio

BY LAWRENCE DRAGGIA
Staff Reporter of The Wall Street Journal

LAKE WOBEGON, Minn.—If you've ever heard Garrison Keillor's radio show, the home of Ralph's Pretty Good Grocery, Our Lady of Perpetual Responsibility Church, the 45th State, and the 45th State, you know he's a good man.

And that it can't be found on any map but only on your radio for two hours every Sunday evening.

The mythical town of Lake Wobegon (pronounced wobegon) comes to life each week on "A Prairie Home Companion," a down-home radio show with a hay-Midwest-

Tax Report

A Special Summary and Forecast
Of Federal and State Tax Developments

FOREIGN CORPORATION rules are varied—with some exceptions.

Congress attacked vacation abuses in 1976 by limiting the amount of business travel abroad to two a year and lengthening spending and reporting requirements. A new law repeats the two-meeting test, which requires that a corporation to be liable for tax if it meets in a foreign country during a single meeting that "as reasonable" to have outside North America at will. However, a provision in the new Price Waterhouse & Co. bill means a meeting deductible under the old two-year rule may not be under the new rules.

This test involves the purposes and activities of the organization and the meetings, the reasonableness of the travel, and the nature of the other gatherings. Meetings on cruise ships aren't deductible at all. The law broadens the definition of foreign countries to include Canada and Mexico, besides the U.S. and its possessions, that should stimulate more international travel.

The pending U.S.-Canada tax treaty would give Canada the same benefit. However, unless Canada has made drastic revisions it won't take effect.

THEY'LL KEEP THE HOUSE AND ALL THE PETS.

The gain from selling your home normally is taxable, unless you put it into a new home by certain rules. But up to \$100,000 of the gain from your residence may be excluded once from your gross income. This applies to the first five years you live there for three of the first four years.

Xena, who's 56, and his wife, Vickie, sold their house at 95% value in 1980.

Fire Hazard

Safety Officials Fear
Skyscraper Holocaust
Could Kill Thousands

They Cite Buildings' Design
And Location, Lax Codes,
Poison Gas From Plastic

Owners Note Record Is Good

By PHILIP E. REEDER
Staff Writer of The Wall Street Journal

Fire fighters say it's over and over, to anyone they can get to listen: A major fire disaster in a skyscraper will be terrible, but will rival an earthquake or flood in death and destruction.

The difference is in how many will die through a skyscraper during working hours, trapping thousands of people and upper floors, than in a community where a fire is concentrated or spread in death in a few minutes.

But don't think it's more than likely. "I'm not talking about probability, I'm talking about certainty," says Michael F. Lafley, a former captain and 20-year veteran of the New York City Fire Department. He is currently safety officer at Comcast's 10-story Manhattan office tower.

With the current record, Mr. Lafley can tick off their concerns:

—More people than ever work in offices today, and most of them are in buildings in many in suburbs likely to be served by vol-

What to do about non-standard (or *non-canonical*) language in NLP

Barbara Plank
University of Groningen
b.plank@rug.nl

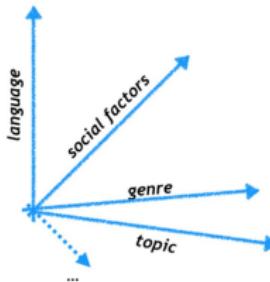
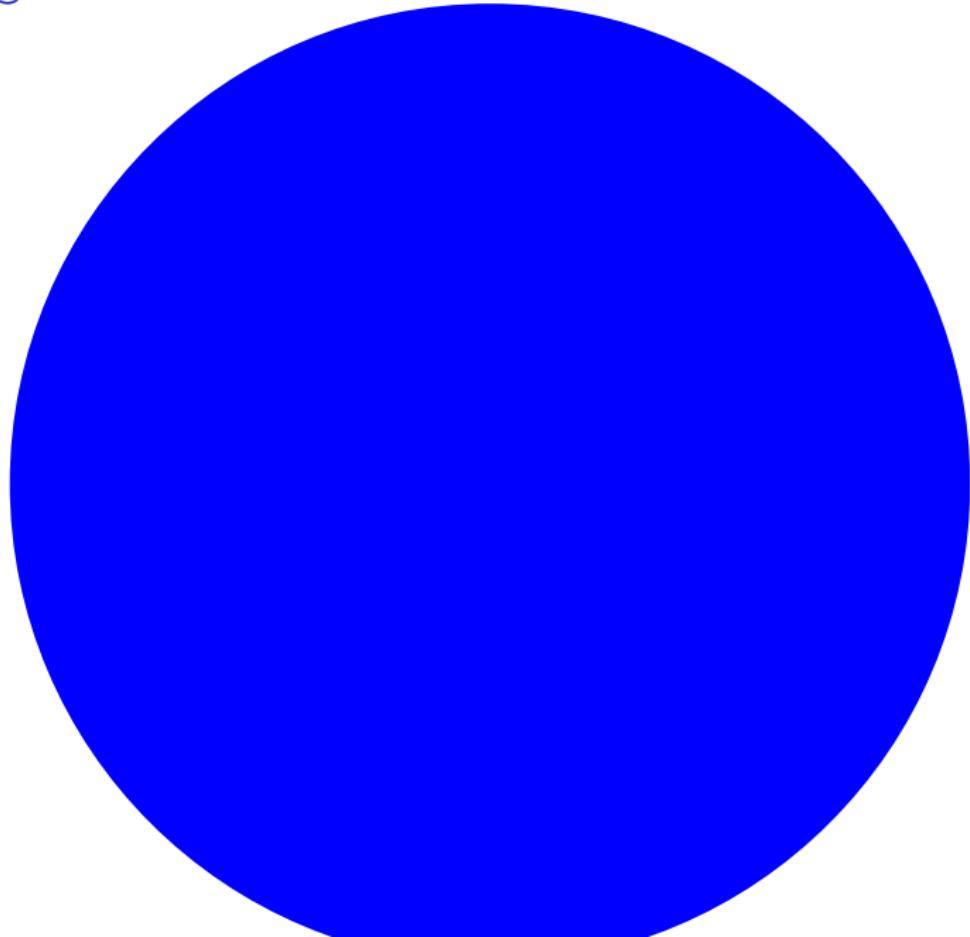


Figure 2: What's in a *domain*? Domain is an overloaded term. I propose to use the term *variety*. A dataset is a sample from the *variety space*, a unknown high-dimensional space, whose dimensions contain (fuzzy) aspects such as language (or dialect), topic or genre, and social factors (age, gender, personality, etc.), amongst others. A domain forms a region in this space, with some members more prototypical than others.

Language varieties that are annotated for NLP tasks(in red)



What can we do?

- ▶ Annotate more?
 - ▶ Language is not static!
- ▶ Cross-domain, cross-lingual learning

Multi-task learning to the rescue!

Standard in NLP:

- ▶ Pre-train a **language model** on raw data (billions of words)
- ▶ Fine-tune the language model on NLP-annotated data (thousands of words)

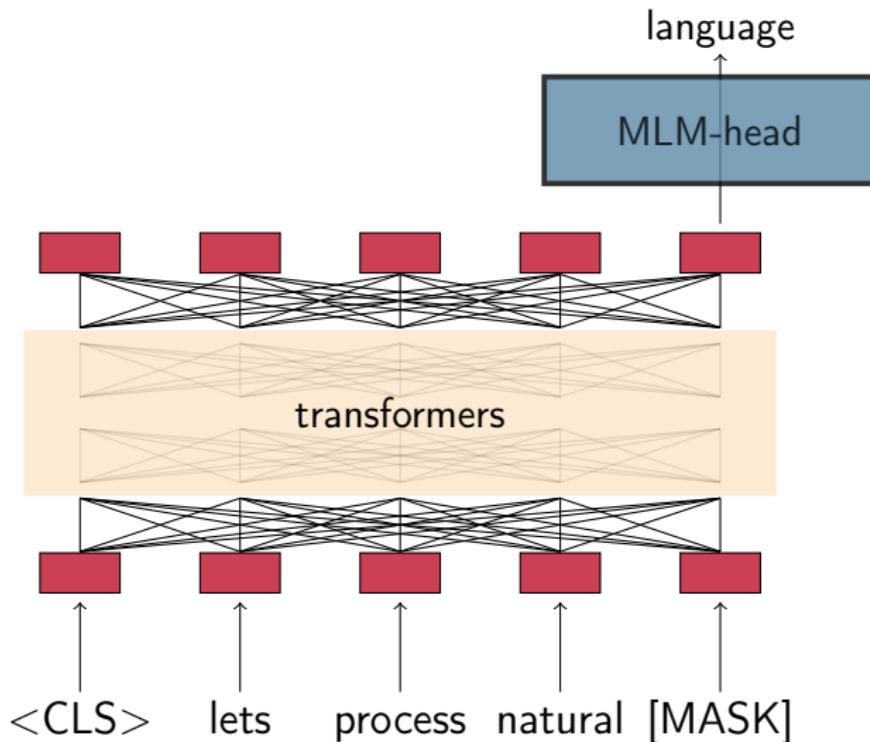
What is a language model?

i  ved

Øvelse 34. Kig på stuen side 76 og indsæt passende præpositioner.

- Der står nogle bøger _____ regolen.
- Der står et fjernsyn _____ sofaen.
- Der ligger et gulvtæppe _____ sofabordet.
- Der står en kop _____ telefonen.
- Der står en vase _____ sofabordet.
- Der hænger nogle billeder _____ vasen.

What is a language model?



Framework: MaChaMp

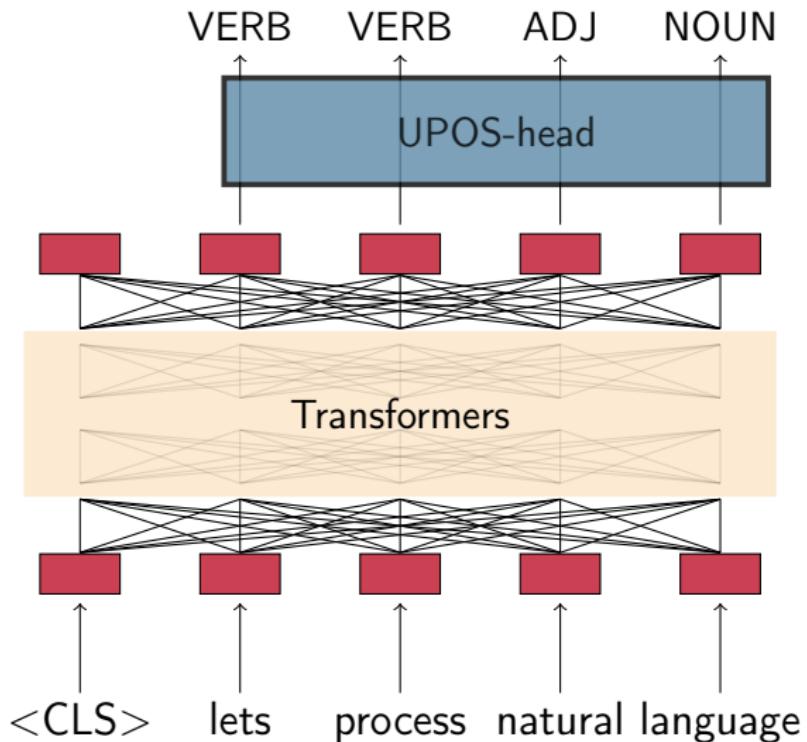
Massive Choice, Ample Tasks (MACHAMP):

 A Toolkit for Multi-task Learning in NLP 

Rob van der Goot  **Ahmet Üstün**  **Alan Ramponi**   **Ibrahim Sharaf** 
Barbara Plank 

IT University of Copenhagen  University of Groningen  University of Trento 
Fondazione the Microsoft Research - University of Trento COSBI  Factmata 
`robv@itu.dk, a.ustun@rug.nl, alan.ramponi@unitn.it
ibrahim.sharaf@factmata.com, bapl@itu.dk`

MaChAmp



- ▶ This is the default setup for all NLP tasks these days; sharing happens over time: MLM \Rightarrow TGT task
- ▶ MaChAmp can do much more!, we mainly focus on multi-task learning after the first pre-training



MachAmp #001

Abilities:

Joint training	loss weighing
sequential training	layer attention
dataset embeddings	dataset smoothing

MachAmp is a multi-task NLP toolkit, it can seemingly effortlessly handle multiple NLP tasks simultaneously. It supports a wide variety of NLP tasks, and can easily handle multiple datasets at once.

More information on machamp-nlp.github.io

IT UNIVERSITY OF CPH

Examples of tasks

Input	Output
classification Smell ya later!	negative
mlm Gotta [MASK] em all	catch
multclas That will be 5\$	inform request
multiseq I never caught Snorlax	per:1 n:sin _ tens:past n:sin
regression You're playing cats	1.2
seq I want to be the best	PRN VB PART AUX DT ADJ
seq_bio Ash from Pallet Town	Ash:PERS Pallet Town:LOC
tok Gary, Gary, he's the man.	Gary , Gary , he 's the man .
dependency Brock wants to fight	root [subj] [comp] [mark]
Brock wants to fight	Brock wants to fight

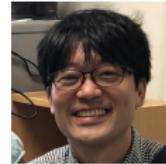
Rob van der Goot

Examples of tasks

Input	Output
classification	
Smell ya later!	negative
mlm	
Gotta [MASK] em all	catch
multclas	
That will be 5\$	inform request
multiseq	
I never caught Snorlax	per:1 n:sin _ tens:past n:sin
regression	
You're playing cats	1.2
seq	
I want to be the best	PRN VB PART AUX DT ADJ
seq_bio	
Ash from Pallet Town	Ash:PERS Pallet Town:LOC
tok	
Gary, Gary, he's the man.	Gary , Gary , he 's the man .
dependency	
Brock wants to fight	<p>Brock wants to fight</p>

xSID: Cross-lingual Slot and Intent Detection

Rob van der Goot, Ibrahim Sharaf, Aizhan Imankulova, Ahmet Üstün, Marija Stepanović, Alan Ramponi, Siti Oryza Khairunnisa, Mamoru Komachi and Barbara Plank



Example: Languages in EU covered by voice assistants



*as of March, 2020



87

* Slide by Barbara Plank

Slot and Intent Detection

I'd like to see the showtimes for Silly Movie 2.0 at the movie house
Intent: SearchScreeningEvent

ar	أود أن أرى مواعيد عرض فيلم Silly Movie 2.0 في دار السينما
da	Jeg vil gerne se spilletiderne for Silly Movie 2.0 i biografen
de	Ich würde gerne den Vorstellungsbeginn für Silly Movie 2.0 im Kino sehen
de-st	I mecht es Programm fir Silly Movie 2.0 in Film Haus sechn
en	I'd like to see the showtimes for Silly Movie 2.0 at the movie house
id	Saya ingin melihat jam tayang untuk Silly Movie 2.0 di gedung bioskop
it	Mi piacerebbe vedere gli orari degli spettacoli per Silly Movie 2.0 al cinema
ja	映画館の Silly Movie 2.0 の上映時間を見て。
kk	Мен Silly Movie 2.0 бағдарламасының кинотеатрда көрсетілім уақытын көргім келеді
nl	Ik wil graag de speeltijden van Silly Movie 2.0 in het filmhuis zien
sr	Želetabih da vidim raspored prikazivanja za Silly Movie 2.0 u bioskopu
tr	Silly Movie 2.0'ininema salonundaki seanslarını görmek istiyorum
zh	我想看 Silly Movie 2.0 在影院 的放映

ar	أود أن أرى مواعيد عرض فيلم Silly Movie 2.0 في دار السينما
da	Jeg vil gerne se spilletiderne for Silly Movie 2.0 i biografen
de	Ich würde gerne den Vorstellungsbeginn für Silly Movie 2.0 im Kino sehen
de-st	I mecht es Programm fir Silly Movie 2.0 in Film Haus sechn
en	I'd like to see the showtimes for Silly Movie 2.0 at the movie house
id	Saya ingin melihat jam tayang untuk Silly Movie 2.0 di gedung bioskop
it	Mi piacerebbe vedere gli orari degli spettacoli per Silly Movie 2.0 al cinema
ja	映画館の Silly Movie 2.0 の上映時間を見せて。
kk	Мен Silly Movie 2.0 бағдарламасының кинотеатрда көрсетілім уақытын көргім келеді
nl	Ik wil graag de speeltijden van Silly Movie 2.0 in het filmhuis zien
sr	Želelabih da vidim raspored prikazivanja za Silly Movie 2.0 u bioskopu
tr	Silly Movie 2.0'ın sinema salonundaki seanslarını görmek istiyorum
zh	我想看 Silly Movie 2.0 在 影院 的放映

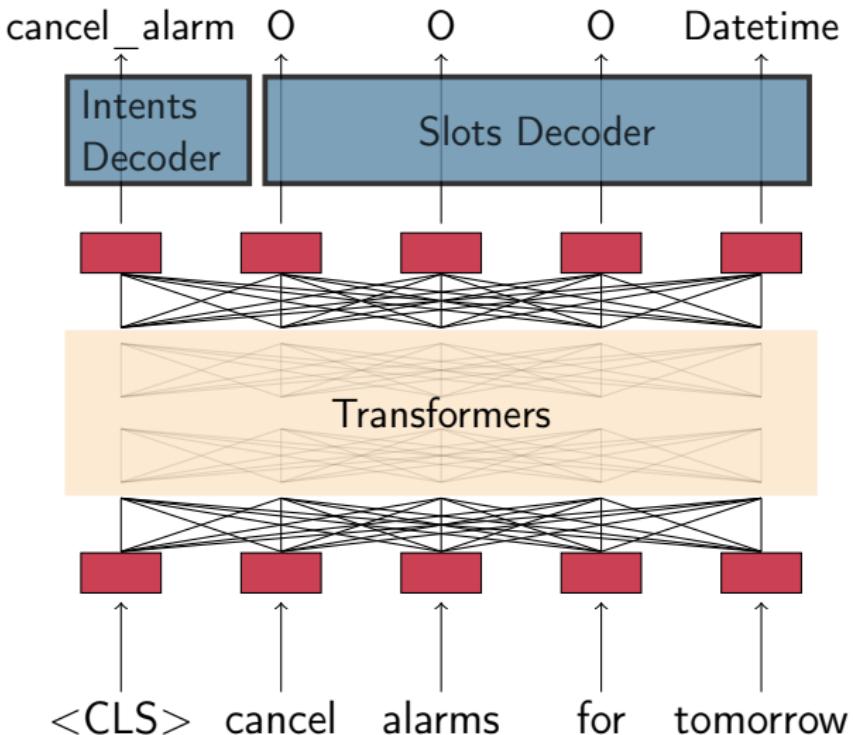
- ▶ Lithuanian
- ▶ Bavarian
- ▶ Swiss German
- ▶ Neapolitan

Experiments

Baselines

- ▶ Baseline: language model with joint intent+slots

Baseline



Experiments

Baselines

- ▶ Baseline: language model with joint intent+slots
- ▶ Stronger baseline: translate training data to target language (NMT-TRANSFER)

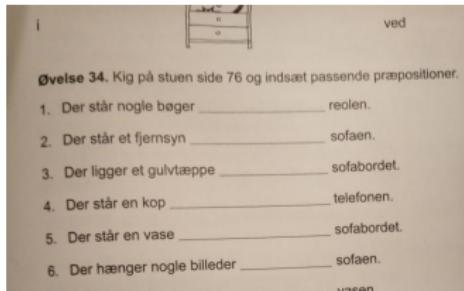
Experiments

Baselines

- ▶ Baseline: language model with joint intent+slots
- ▶ Stronger baseline: translate training data to target language (NMT-TRANSFER)

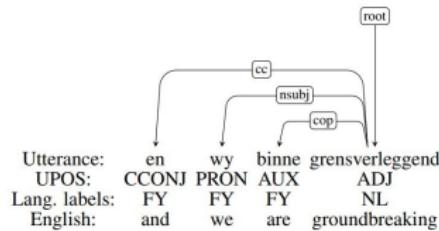
New models:

- ▶ Train on auxiliary task in target language:
 - ▶ Masked language modeling (AUX-MLM)
 - ▶ Neural machine translation (AUX-NMT)
 - ▶ UD-parsing (AUX-UD)



► MLM:

► NMT:



► UD-parsing:

Experiments

Evaluate 2 language models

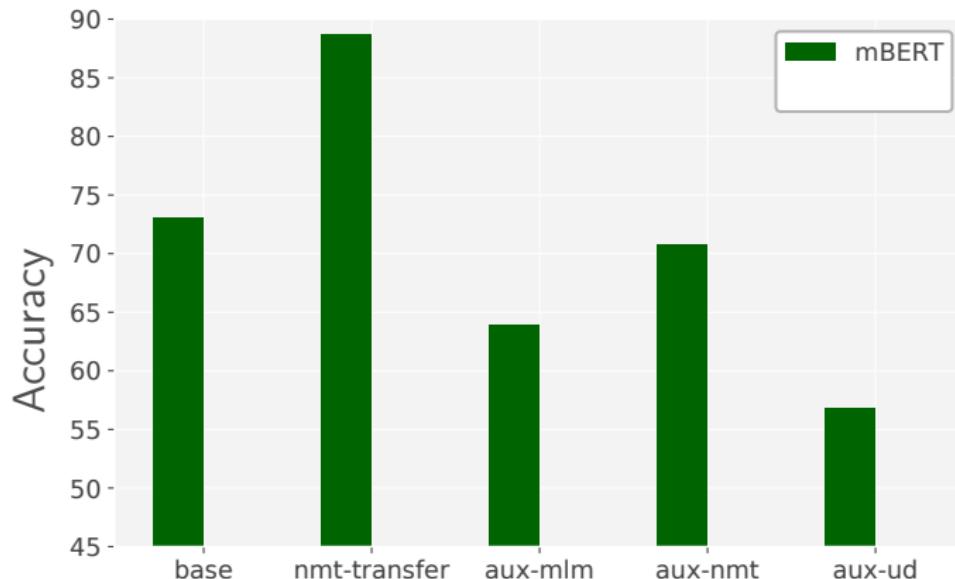
- ▶ mBERT: trained on 104 languages (12/13)
- ▶ XLM15: trained on 15 languages (5/13)

Results

model	Time (minutes)
base	46
nmt-transfer	5,213
aux-mlm	193
aux-nmt	373
aux-ud	79

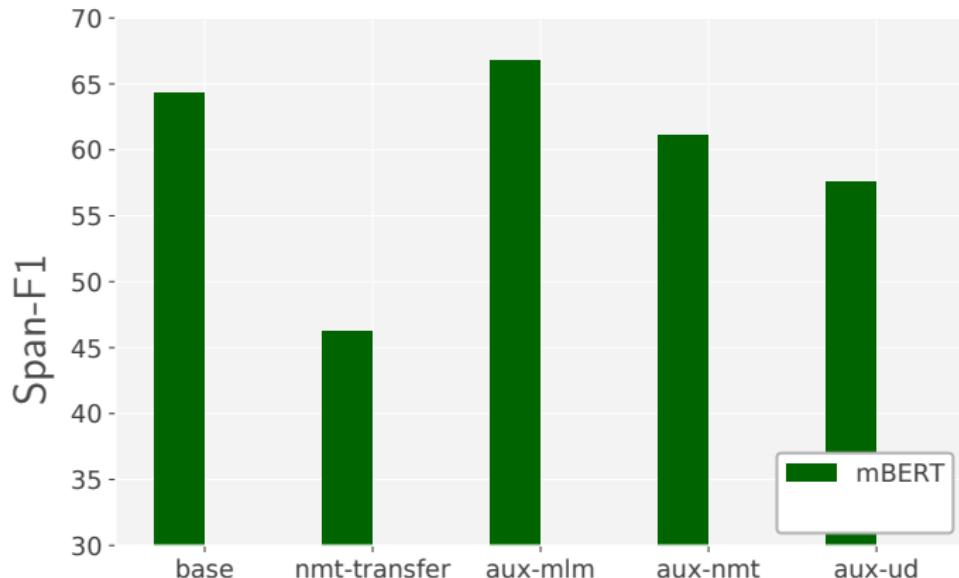
Table: Average minutes to train a model, averaged over all languages and both embeddings. For nmt-transfer we include the training of the NMT model.

Results (intents)



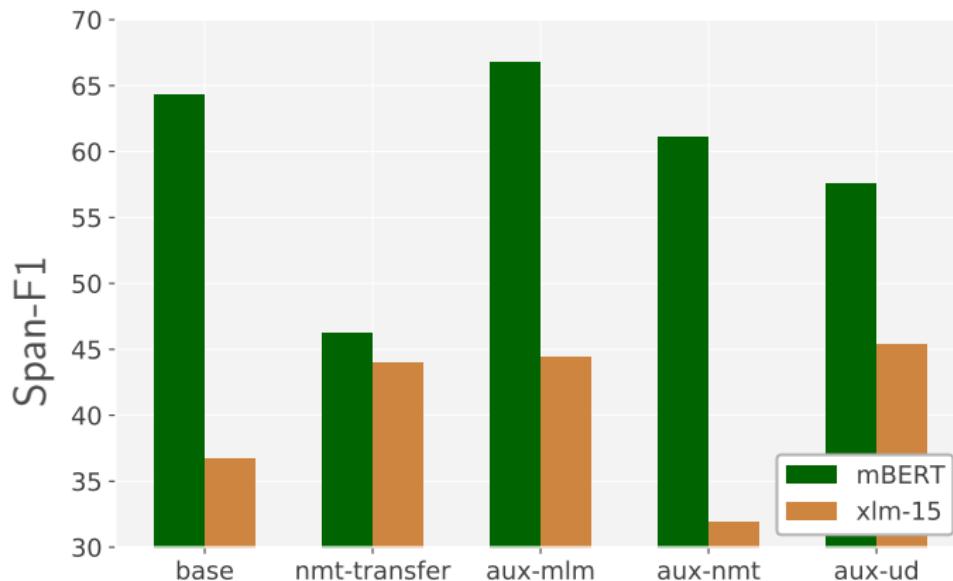
- ▶ NMT-transfer is hard to outperform, but costly
- ▶ Even baseline hard to beat

Results (slots)



- ▶ NMT-transfer performs bad (due to alignment)
- ▶ Seen languages: only MLM helps

Results (slots)



- ▶ NMT-transfer performs bad (due to alignment)
- ▶ Seen languages: only MLM helps
- ▶ Unseen languages: More explicit tasks (UD) are faster and lead to better performance

How to implement this

Requirements:

- ▶ Training data
- ▶ In tab-separated format
- ▶ Couple hours of GPU time

How to implement this

```
xsid.json:  
{ "xSID": {  
    "train_data_path": "data/xsid-0.5/en.train.conll",  
    "validation_data_path": "data/xsid-0.5/en.dev.conll",  
    "word_idx": 1,  
    "tasks": {  
        "slots": {  
            "task_type": "seq_bio",  
            "column_idx": 3  
        }  
        "intents": {  
            "task_type": "classification",  
            "column_idx": -1  
        }  
    }  
}
```

```
python3 train.py --dataset_configs xsid.json
```

How to implement this

```
mlm-da.json:  
{ "MLM": {  
    "train_data_path": "data/spoken/da.train.txt",  
    "validation_data_path": "data/spoken/da.dev.txt",  
    "tasks": {  
        "language_modeling": {  
            "task_type": "mlm",  
        }  
    }  
}  
}  
python3 train.py --dataset_configs xsid.json mlm-da.json
```

How to implement this

```
mlm-da.json:  
{ "MLM": {  
    "train_data_path": "data/spoken/da.train.txt",  
    "validation_data_path": "data/spoken/da.dev.txt",  
    "tasks": {  
        "language_modeling": {  
            "task_type": "mlm",  
        }  
    }  
}  
}  
  
python3 train.py --dataset_configs xsid.json mlm-da.json
```

Need better performance?: try other language models.

Interested in SOTA NLP?:

- ▶ Try MaChAmp:
<https://github.com/machamp-nlp/machamp.git>
- ▶ Get in touch (slack/e-mail)