Feature-augmented model for multilingual discourse relation classification

Eleni Metheniti, Chloé Braud, Philippe Muller

UT3 - IRIT - CNRS

firstname.lastname@irit.fr

Institut de Recherche en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

UNIVERSITÉ TOULOUSE III PAUL SABATIER Université de Toulouse

cnrs dépasser les frontières

ANITI ARTIFICIAL & NATURAL INTELLIGENCE TOULOUSE INSTITUTE

KEEP CALM AND ANDIAMO AVANTI

source image Andiamo

DISRPT

- DISRPT Shared Task: Discourse Relation Classification across Formalisms

IMPLICIT

ENG.PDTB.PDTB (wsj_2315)

UNIT 1: RJR moved 11 employees of the group back to New York in September
UNIT 2: because "there was supposed to be a future."
DIRECTION: 1<2
ORIGINAL LABEL: Contingency.Cause.Reason
DISRPT LABEL: Contingency.Cause

EXPLICIT

ENG.RST.RSTDT (wsj_1317)

UNIT 1: Known as a traditional executive,
UNIT 2: he is very much in the conservative American Express mold.
DIRECTION: 1>2
ORIGINAL LABEL: elaboration-additional-e
DISRPT LABEL: elaboration

DISRPT 2021

- DISRPT 2021 Shared Task: Discourse Relation Classification across Formalisms
- 16 datasets, 11 languages, 3 frameworks
- 🥇 DisCoDisCo (Gessler et al., 2021)

IMPLICIT

ENG.PDTB.PDTB (wsj_2315)

UNIT 1: RJR moved 11 employees of the group back to New York in September
UNIT 2: because "there was supposed to be a future."
DIRECTION: 1<2
ORIGINAL LABEL: Contingency.Cause.Reason
DISRPT LABEL: Contingency.Cause

EXPLICIT

ENG.RST.RSTDT (wsj_1317)

UNIT 1: Known as a traditional executive,
UNIT 2: he is very much in the conservative American Express mold.
DIRECTION: 1>2
ORIGINAL LABEL: elaboration-additional-e
DISRPT LABEL: elaboration

DisCoDisCo system

- Monolingual, one-per-corpus classifiers based on BERT

DisCoDisCo system

- Monolingual, one-per-corpus classifiers based on BERT

- **Features** in embedding layer (not the same for each classifier!):
  - Categorical features: *genre, head word's POS, list of common words…*
  - Numerical features: *length of unit, position of unit in document, distance…*
  - Boolean features: *same speaker, (unit) is sentence, (unit) is discontinuous*

DisCoDisCo system

- Monolingual, one-per-corpus classifiers based on BERT
- **Features** in embedding layer (not the same for each classifier!):
  - Categorical features: *genre, head word's POS, list of common words…*
  - Numerical features: *length of unit, position of unit in document, distance…*
  - Boolean features: *same speaker, (unit) is sentence, (unit) is discontinuous*
- **Direction annotation** with special tokens:

`[CLS] } unit 1 > unit 2`

`[CLS] unit 1 < unit 2 {`

DISRPT 2023

- DISRPT 2023: 26 datasets, 13 languages, 4 frameworks

DISRPT 2023

- DISRPT 2023: 26 datasets, 13 languages, 4 frameworks
- 🥇 HITS (Liu et al., 2023): monolingual or framework-based, large models
- 🥈 DiscReT (Metheniti et al., 2023): **multilingual only**, **label harmonization**, **switching units** for direction
- 🥉 DiscoFLAN (Anuranjana, 2023): generative models, **prediction filtering**

Our research questions

Which DisCoDisCo features are best?

- All features
- “Common” features

Our research questions

Which DisCoDisCo features are best?

- All features
- “Common” features
- New features:

  Language, Framework, Corpus (LCF)

```
[French, SDRT, fra.sdrt.annodis]
[English, RST, eng.rst.rstdt]
```

Our research questions

Which DisCoDisCo features are best?

- All features
- “Common” features
- New features:

  Language, Framework, Corpus (LCF)

How to handle relation direction best?

- No change
- DisCoDisCo direction special tokens
- DiscReT switching units

```
[French, SDRT, fra.sdrt.annodis]
[English, RST, eng.rst.rstdt]
```

Methodology

- **Multilingual**, **multi-framework** joint setting for DISRPT 2021
- Pretrained models: mBERT, DistilmBERT, XLM-RoBERTa (all in base size)

Methodology

- **Multilingual**, **multi-framework** joint setting for DISRPT 2021
- Pretrained models: mBERT, DistilmBERT, XLM-RoBERTa (all in base size)
- **Features**: added to vocabulary, before units. Input:

```
[CLS] language corpus framework feature_length feature_POS-Unit_1 … feature_percent
This is Unit 1 , [SEP] then follows Unit 2 .
```

Methodology

- **Multilingual**, **multi-framework** joint setting for DISRPT 2021
- Pretrained models: mBERT, DistilmBERT, XLM-RoBERTa (all in base size)
- **Features**: added to vocabulary, before units. Input:

```
[CLS] language corpus framework feature_length feature_POS-Unit_1 … feature_percent
This is Unit 1 , [SEP] then follows Unit 2 .
```

- **Direction**: see DisCoDisCo, DiscReT

Methodology

- **Multilingual**, **multi-framework** joint setting for DISRPT 2021
- Pretrained models: mBERT, DistilmBERT, XLM-RoBERTa (all in base size)
- **Features**: added to vocabulary, before units. Input:

```
[CLS] language corpus framework feature_length feature_POS-Unit_1 … feature_percent
This is Unit 1 , [SEP] then follows Unit 2 .
```

- **Direction**: see DisCoDisCo, DiscReT
- **Label filtering**: based on framework, not corpus

```
PDTB:{ expansion.conjunction: 0.2, ~~joint: 0.25~~, contingency.cause: 0.02 …}
```

Results

Model:	DisCoDisCo 2021	mBERT	DistilmBERT	XLM-R	mBERT	DistilmBERT	XLM-R
Direction:	Add. tokens	Add. tokens			Switching units		
No features	60.41	59.54	56.81	62.09	58.36	55.69	60.52
Common features	61.82	62.56	60.92	64.86	59.75	57.24	61.14
All features	-	63.09	60.28	64.50	62.33	59.08	63.95
LCF	-	61.76	59.17	64.13	58.34	55.69	60.52
LCF + Common	-	63.46	**62.01**	**65.91**	61.12	57.75	62.88
LCF + All	-	63.67	61.92	65.53	**63.89**	59.65	63.51

Results

Model:	DisCoDisCo 2021	mBERT	DistilmBERT	XLM-R	mBERT	DistilmBERT	XLM-R
Direction:	Add. tokens	Add. tokens			Switching units		
No features	60.41	59.54	56.81	62.09	58.36	55.69	60.52
Common features	61.82	62.56	60.92	64.86	59.75	57.24	61.14
All features	-	63.09	60.28	64.50	62.33	59.08	63.95
LCF	-	61.76	59.17	64.13	58.34	55.69	60.52
LCF + Common	-	63.46	**62.01**	**65.91**	61.12	57.75	62.88
LCF + All	-	63.67	61.92	65.53	**63.89**	59.65	63.51

Results: Features

- Best overall:
  - LCF + Common for XLM-R, DistilmBERT
  - LCF + All for mBERT
- +4% accuracy compared to DisCoDisCo (62% -> 66%)

- Big improvement for `spa.rst.sctb` (+16%) and `deu.rst.pcc` (+10%)
- 3-8% improvement on the rest…
- …but no improvement on `eus.rst.ert`

Results: Direction

- Overall: Using direction information > Not using it
- Without features: **Additional tokens** > Switching units > None (for all models)
- With features: **Additional tokens** for XLM-R and DistilmBERT
  Switching units for mBERT

Conclusion

1. Multilingual discourse relation classification can be better than monolingual!

2. Features work well for all models and languages!

3. Direction information is useful! Additional tokens over other manipulation

Thank you for your attention!

Find the project on GitLab:
gitlab.irit.fr/melodi/andiamo/relation-classification-features