THE UNIVERSITY OF ARIZONA.

Sushma Akoju <sushmaakoju@arizona.edu>

Regarding NL to FOL course project

Sushma Akoju <sushmaakoju@arizona.edu> Thu, Nov 24, 2022 at 1:50 PM
To: Mihai Surdeanu <surdeanu@gmail.com>

Dear Sir,

I updated the report and added a section for measuring Representational Similarity analysis using Tree kernel for comparing predictions and ground truth FOL.

But I am working on a toy example, I hope to share it by end of this week for how this metric could work for T5 encoder-decoder architecture when comparing ground truth FOL statements.

Best Regards,
Sushma Akoju

On Thu, Nov 24, 2022 at 11:38 AM Mihai Surdeanu <surdeanu@gmail.com> wrote:

> **External Email**
>
> The tree-kernel based one. In addition of the standard MT measures you already have.
>
> On November 24, 2022 at 11:37:39 AM, Sushma Akoju (sushmaakoju@arizona.edu) wrote:
>
> > Dear Sir,
> >
> > Thank you. Sorry I did not understand which measure do you suggest is relevant?
> >
> > Best Regards,
> > Sushma Akoju
> >
> > On Thu, Nov 24, 2022 at 10:52 AM Mihai Surdeanu <surdeanu@gmail.com> wrote:
> >
> > > **External Email**
> > >
> > > Yes, that measure makes sense to me.
> > >
> > > On Thu, Nov 24, 2022 at 10:07 Sushma Akoju <sushmaakoju@arizona.edu> wrote:
> > >
> > > > Dear Sir,
> > > >
> > > > Thank you for the feedback.
> > > > Sure, Eduardo also mentioned I include a correlation metric but later suggested since I have only two systems to compare with, so he just suggested including one table with just metrics and errors. Do you suggest including these or any other approaches for correlation metrics? Can we use alignment attachment relevant metrics? One approach that crossed my mind was: https://arxiv.org/pdf/1905.06401.pdf - Representational Similarity Analysis (RSA) using Tree Kernels (this was the paper I have asked questions about RSA and tree kernels) - can we use this for comparing the substructure similarity between FOL syntax trees?
> > > >
> > > > Example for prediction:
> > > > prediction FOL - Shows12133(z) AND (Friends(x,y) AND Performer(Shows, y)) AND Attend(x,y)
> > > > Ground truth FOL - Shows(z) AND (Friends(x,y)) AND (Performer(z,y) AND Attend(x,y))

He suggested following two papers where correlation was measured:
https://arxiv.org/pdf/1603.08023.pdf - pearson's & Spearman coefficient

https://aclanthology.org/2020.findings-emnlp.345/ - Kendall's Tau τ coefficient (Kendall, 1938),

	Twitter				Ubuntu			
Metric	**Spearman**	p-value	**Pearson**	p-value	**Spearman**	p-value	**Pearson**	p-value
Greedy	0.2119	0.034	0.1994	0.047	0.05276	0.6	0.02049	0.84
Average	0.2259	0.024	0.1971	0.049	-0.1387	0.17	-0.1631	0.10
Extrema	0.2103	0.036	0.1842	0.067	0.09243	0.36	-0.002903	0.98
METEOR	0.1887	0.06	0.1927	0.055	0.06314	0.53	0.1419	0.16
BLEU-1	0.1665	0.098	0.1288	0.2	-0.02552	0.8	0.01929	0.85
BLEU-2	0.3576	< 0.01	0.3874	< 0.01	0.03819	0.71	0.0586	0.56
BLEU-3	0.3423	< 0.01	0.1443	0.15	0.0878	0.38	0.1116	0.27
BLEU-4	0.3417	< 0.01	0.1392	0.17	0.1218	0.23	0.1132	0.26
ROUGE	0.1235	0.22	0.09714	0.34	0.05405	0.5933	0.06401	0.53
Human	0.9476	< 0.01	1.0	0.0	0.9550	< 0.01	1.0	0.0

Table 3: Correlation between each metric and human judgements for each response. Correlations shown in the human row result from randomly dividing human judges into two groups.

Best Regards,
Sushma Akoju

On Thu, Nov 24, 2022 at 9:29 AM Mihai Surdeanu <surdeanu@gmail.com> wrote:

External Email

It looks good!
The only I would add is measuring the correlation between human judgments and automated measures. I wonder if these automated measures are any good for measuring FOL quality...

On Thu, Nov 24, 2022 at 8:36 AM Sushma Akoju <sushmaakoju@arizona.edu> wrote:

Dear Sir,

It was due yesterday night at 11:59pm.
Do you suggest any changes to the report?
Eduardo mentioned I can make changes to the project as needed.

Best Regards,
Sushma Akoju

On Thu, Nov 24, 2022 at 8:33 AM Mihai Surdeanu <surdeanu@gmail.com> wrote:

External Email

when is it due?

On Thu, Nov 24, 2022 at 6:56 AM Sushma Akoju <sushmaakoju@arizona.edu> wrote:

Dear Sir,

I have attached my course project midterm report.
I included and added all possible relevant works to discuss the reasons for having Natural Language to First Order Logic conversion.

I have added you to the overleaf document if you suggest any review comments, https://www.overleaf.com/1492374186tnvjdvjdcsmm

Can you please let me know your thoughts about the report when you get a chance?

Best Regards,

Sushma Akoju

csc_696_midterm_report_tree_kernel.pdf
273K