LIGN 167 Project

Brian Chu, Julie Hang, Kyle Batalla, Richard Li

3/11/2022

1. Introduction

The study of drug interactions in bio medicine focuses on the actions and side effects of drugs with other drugs and supplements as well. Understanding the behavior of drug interactions is a crucial part of bio medicine, as it allows us to understand the potential side effects and hazards of different drugs. The different results and conclusions of drug interactions are based on many years of studies and a large variety of biomedical papers by doctors, scientists, and researchers over the years. In terms of natural language processing, many models have been built and trained to analyze and extract the different semantic relations of multiple drugs through data sets of text that has been derived from such studies. Based on the prior study and knowledge of drug interactions, these can be used to test accuracy of different models by comparing results of the model to the correct results obtained by the different studies (Wang et al.).

The data sets which will be used in this project will be extracted from well known drug online databases that contain the names and annotated effects of drugs with other different known drugs (more on 3. Datasets). From here, we plan to extract the drugs and different annotations to derive the semantic relations through relation extraction. Relation extraction will be the basis of our project, in which it can be see as extracting relations from text and analyzing their semantic meaning. Importantly, this can be used to build and represent different relations and interactions through graphs on multiple entities (drugs in our case) based on different word embeddings and encodings, all of which are used to train and build predictive models. This is important because it allows us to predict how multiple drugs will interact with each other, without prior interactions.

Many notable studies on relation extractions of drug interactions have been based off training some form of neural network and implementing some form of graph. Knowledge graphs, which are the most common, are important because they are used to store data based from information extraction, and more importantly, used to link relations between multiple entities. Our original approach was to replicate some of the methodology of such previous work implemented by others such as implementing a knowledge graph and using a graph neural network to predict outcomes. Due to challenges however, we decided to test if we were able to get effective results based off text generation by feeding a pre-trained model with a corpus of effects and fine tuning it based on our data.

2. Model

The model we have decided to work with was Huggingface's XLNet Autoregressive Pretrained Model, which is more popularly known for solving tasks such as sentiment analysis, question answering, document ranking, and more (Wolf et al.). We trained the model with the names of the drugs and sentences respectively and added it into the corpus iteratively. We appended an $< eod >< /s >< eos >$ token to the end of the each string. We then ran it through our train_method that uses huggingface transformers pipeline for word generation. We loaded in the pretrained model named "xlnet-base-cased" and ran it through each node adding to the corpus to get generated text results.

```
def train_model():
    PADDING_TEXT = ""
    prompt = "The combined effects of drugs are "
    mlength = 100
    nnodes = 10
    for i in range(nnodes):
        PADDING_TEXT = PADDING_TEXT + list2str(drugbank_epochs[i:i+1])
        mlength = mlength + len(PADDING_TEXT.split())
        generated = train_instance(PADDING_TEXT, prompt, mlength)
        print(generated)
        print("\n\n")
    print(generated)
```

3. Datasets

Our datasets consists of annotated biological XML files containing text information of known interactions and side effects of other drugs, all of which were derived from online drug databases DrugBank and MedLine. These XML files are sorted into two folders, one from DrugBank and the other MedLine. Each file contains information about multiple drugs that are related and includes variations of each drug along with their respective effects. We are taking each training point as a single node which consists of a type of drug. Each file will correspond to an epoch, which we will train the model in iterations of epochs to see results. We will add on the file contents we parsed to the corpus to find the different results the model generates. Since DrugBank and MedLine are very different types of data, we actually end up having two different models based on the same pre-trained data. We used about 10 percent of the data for fine tuning and about 5 percent for testing, since processing the data took a lot longer than expected.

4. Results

We started by taking each of the effects of the drugs parsed from the XML files and adding them into a corpus by sentences. We also used segment_and_tokenize from Homework 4 code dealing with text processing. Our results are shown below with the corpus we used printed out and ran into the pre-trained model for text generation.

```
No drug nutritional supplement food or herb interactions have yet been reported No formal drug/drug interaction studies with Plenaxis were performed Plenaxis Cytochrome P-450
 is not known be involved metabolism of Plenaxis Plenaxis Plenaxis is highly bound plasma proteins ( 96 99 % ) Plenaxis Laboratory Tests Response Plenaxis should be monitored
 by measuring serum total testosterone concentrations just prior administration on Day 29 every 8 weeks thereafter Plenaxis testosterone Serum transaminase levels should be o
btained before starting treatment with Plenaxis periodically during treatment Plenaxis Periodic measurement of serum PSA levels may also be considered Formal drug interaction
 studies have not been conducted with ORENCIA ORENCIA Population pharmacokinetic analyses revealed that MTX NSAIDs corticosteroids TNF blocking agents did not influence abata
cept clearance MTX NSAIDs corticosteroids TNF blocking agents abatacept The majority of patients RA clinical studies received one or more of following concomitant medications
 with ORENCIA : MTX NSAIDs corticosteroids TNF blocking agents azathioprine chloroquine gold hydroxychloroquine leflunomide sulfasalazine anakinra ORENCIA MTX NSAIDs corticos
teroids TNF blocking agents azathioprine chloroquine gold hydroxychloroquine leflunomide sulfasalazine anakinra Concurrent administration of TNF antagonist with ORENCIA has b
een associated with increased risk of serious infections no significant additional efficacy over use of TNF antagonists alone TNF antagonist ORENCIA TNF antagonists Concurren
t therapy with ORENCIA TNF antagonists is not recommended ORENCIA TNF antagonists There is insufficient experience assess safety efficacy of ORENCIA administered concurrently
 with anakinra therefore such use is not recommended ORENCIA anakinra Formal drug interaction studies with Abciximab have not been conducted Abciximab Abciximab has been admi
nistered patients with ischemic heart disease treated concomitantly with broad range of medications used treatment of angina myocardial infarction hypertension Abciximab Thes
e medications have included heparin warfarin beta-adrenergic receptor blockers calcium channel antagonists angiotensin converting enzyme inhibitors intravenous oral nitrates
ticlopidine aspirin heparin warfarin beta-adrenergic receptor blockers calcium channel antagonists angiotensin converting enzyme inhibitors nitrates ticlopidine aspirin Hepar
in other anticoagulants thrombolytics anti platelet agents are associated with increase bleeding Heparin anticoagulants thrombolytics anti platelet agents Patients with HACA
titers may have allergic or hypersensitivity reactions when treated with other diagnostic or therapeutic monoclonal antibodies diagnostic monoclonal antibodies therapeutic mo
noclonal antibodies The concomitant intake of alcohol Acamprosate does not affect pharmacokinetics of either alcohol or acamprosate alcohol Acamprosate alcohol acamprosate Ph
armacokinetic studies indicate that administration of disulfiram or diazepam does not affect pharmacokinetics of acamprosate disulfiram diazepam acamprosate Co-administration
 of naltrexone with Acamprosate produced 25 % increase AUC 33 % increase Cmax of acamprosate naltrexone Acamprosate acamprosate No adjustment of dosage is recommended such pa
tients The pharmacokinetics of naltrexone its major metabolite 6-beta-naltrexol were unaffected following co-administration with Acamprosate naltrexone 6-beta-naltrexol Acamp
rosate Other concomitant therapies : In clinical trials safety profile subjects treated with Acamprosate concomitantly with anxiolytics hypnotics sedatives ( including benzod
iazepines ) or non-opioid analgesics was similar that of subjects taking placebo with these concomitant medications Acamprosate anxiolytics hypnotics sedatives benzodiazepine
s non-opioid analgesics Patients taking Acamprosate concomitantly with antidepressants more commonly reported both weight gain weight loss compared with patients taking eithe
r medication alone Acamprosate antidepressants
Input length of input_ids is 979, but ``max_length`` is set to 250.This can lead to unexpected behavior. You should consider increasing ``config.max_length`` or ``max_length`
`.
The combined effects of these drugs are described
```

We then added commas and period in some locations to see if there is a large difference. But there was very little change and it seemed like our model did not work at all when the only issue was the small warning of too small of a max_length size input.

Then we tried changing the prompt to see if it makes a large difference. When we ran our code, we noticed that we actually started to get results with having an appropriate max_length text generation length.

```
No drug nutritional supplement food or herb interactions have yet been reported No formal drug/drug interaction studies with Plenaxis were performed Plenaxis Cytochrome P-450
 is not known be involved metabolism of Plenaxis Plenaxis Plenaxis is highly bound plasma proteins ( 96 99 % ) Plenaxis Laboratory Tests Response Plenaxis should be monitored
 by measuring serum total testosterone concentrations just prior administration on Day 29 every 8 weeks thereafter Plenaxis testosterone Serum transaminase levels should be o
btained before starting treatment with Plenaxis periodically during treatment Plenaxis Periodic measurement of serum PSA levels may also be considered Formal drug interaction
 studies have not been conducted with ORENCIA ORENCIA Population pharmacokinetic analyses revealed that MTX NSAIDs corticosteroids TNF blocking agents did not influence abata
cept clearance MTX NSAIDs corticosteroids TNF blocking agents abatacept The majority of patients RA clinical studies received one or more of following concomitant medications
 with ORENCIA : MTX NSAIDs corticosteroids TNF blocking agents azathioprine chloroquine gold hydroxychloroquine leflunomide sulfasalazine anakinra ORENCIA MTX NSAIDs corticos
teroids TNF blocking agents azathioprine chloroquine gold hydroxychloroquine leflunomide sulfasalazine anakinra Concurrent administration of TNF antagonist with ORENCIA has b
een associated with increased risk of serious infections no significant additional efficacy over use of TNF antagonists alone TNF antagonist ORENCIA TNF antagonists Concurren
t therapy with ORENCIA TNF antagonists is not recommended ORENCIA TNF antagonists There is insufficient experience assess safety efficacy of ORENCIA administered concurrently
 with anakinra therefore such use is not recommended ORENCIA anakinra Formal drug interaction studies with Abciximab have not been conducted Abciximab Abciximab has been admi
nistered patients with ischemic heart disease treated concomitantly with broad range of medications used treatment of angina myocardial infarction hypertension Abciximab Thes
e medications have included heparin warfarin beta-adrenergic receptor blockers calcium channel antagonists angiotensin converting enzyme inhibitors intravenous oral nitrates
ticlopidine aspirin heparin warfarin beta-adrenergic receptor blockers calcium channel antagonists angiotensin converting enzyme inhibitors nitrates ticlopidine aspirin Hepar
in other anticoagulants thrombolytics anti platelet agents are associated with increase bleeding Heparin anticoagulants thrombolytics anti platelet agents Patients with HACA
titers may have allergic or hypersensitivity reactions when treated with other diagnostic or therapeutic monoclonal antibodies diagnostic monoclonal antibodies therapeutic mo
noclonal antibodies The concomitant intake of alcohol Acamprosate does not affect pharmacokinetics of either alcohol or acamprosate alcohol Acamprosate alcohol acamprosate Ph
armacokinetic studies indicate that administration of disulfiram or diazepam does not affect pharmacokinetics of acamprosate disulfiram diazepam acamprosate Co-administration
 of naltrexone with Acamprosate produced 25 % increase AUC 33 % increase Cmax of acamprosate naltrexone Acamprosate acamprosate No adjustment of dosage is recommended such pa
tients The pharmacokinetics of naltrexone its major metabolite 6-beta-naltrexol were unaffected following co-administration with Acamprosate naltrexone 6-beta-naltrexol Acamp
rosate Other concomitant therapies : In clinical trials safety profile subjects treated with Acamprosate concomitantly with anxiolytics hypnotics sedatives ( including benzod
iazepines ) or non-opioid analgesics was similar that of subjects taking placebo with these concomitant medications Acamprosate anxiolytics hypnotics sedatives benzodiazepine
s non-opioid analgesics Patients taking Acamprosate concomitantly with antidepressants more commonly reported both weight gain weight loss compared with patients taking eithe
r medication alone Acamprosate antidepressants
The combined effects of drugs Abarelix and Testosterone Therapy The benefits of A/C therapy : Comparison of A/C and T/C treatments from 1  to 3 ( 8 to 24 <unk>),A/C therapy t
o 12 <unk> of T/C therapy. 2 <unk> of A/C therapy to 3 <unk> of T/C therapy 2 <unk> of A/C therapy to 6 <unk> of A/C therapy to 12 <unk> of T/
```

In the dataset, we treated the each drug name as a node and used this for creating epochs for training. After training it, we noticed that there were

```
train_model()
8m 36.6s                                                                                                                                                    Python

The combined effects of drugs are tested by measuring serum total testrone concentrations just prior to administration and every 8 weeks thereafter. The body should respond to the drug
immediately on Day 29 and every 8 weeks thereafter. The body must respond to the drug immediately on Day 29 and every 8 weeks thereafter for a suitable time thereafter, the body must respond
to the drug immediately on Day 29 and every 8 weeks thereafter for a suitable time thereafter for a suitable time



The combined effects of drugs are shown on a RA clinical trial in RA by looking at blood levels and plasma mass index as well as on other RA clinical studies. Multiple RA clinical trials



The combined effects of drugs are controlled by controlling the levels of TNF blocking agents on the patient's blood tests. TNF blocking agents in RA use many of the same parameters as in RA
clinical studies. TNF blocking agents in RA use many of the same parameters as for RA clinical studies and are administered in any way that provides high, effective, and controlled levels of
TNF (to be used with a RA clinical plan). RA clinical studies are conducted in
```

After further iterations in our trained model we got very interesting results that started to deviate away from our own dataset, and followed the pre-trained data.

```
The combined effects of drugs are a ------- drugs ------------------------------------------------------------------------
128



The combined effects of drugs are very similar compared with their counterpart steroids, which include the pharmacological effects of drug - - - - corticosteroids TNF blocking agents which enhance the pharmacological effects of steroid - - -
206



The combined effects of drugs are usually measured by the level of testosterone in their serum. The highest level of testosterone in patients who received a RA clinical trial and received the most reaction to the Drugs is 9/14/18/19/20/20/21/18/20/20/22/24/20/20/20/25/20/23/25/25/23/25/25/25/25/26/25/25/25/25/25/25/25
304



The combined effects of drugs are described in EA III (Supplemental Table). The complete list of drug effects is presented in Table III (Supplemental Table).<eop><eod> 2. In order to use this software, you may first want to access Windows 9.2.1 or Windows 10.1.2.2. The installation must include some programs such as Microsoft Word and Excel, Microsoft Business Builder or File System Manager (File System Manager) or the like. This should be available without the installation of Windows 8.4.1 or the like.<eop> 3. To download the latest version of Windows 9.2.1 or the like, you may need to install Windows 7.2.1 or the like.<eop> 4.
437



The combined effects of drugs are shown to be reversible in control of those controlling. It also appears that the inhibition of amino acids is reversible in the control of other factors and can produce significant results.<eop><eod> The page links to the new website.<eop> This is a sample of the new website. You should also look at the other versions of the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website.<eop> The article links to the new website. The article links to the new website. The article links to the new website.<eop><eod> The image is a part of a computer made out of glass, ceramic, or metal with a similar texture, it may not be the same color or the same size, and that it may be any other metal or ceramic or glass in some way.<eop><eod><eod> This is a list of all
606



The combined effects of drugs are shown to be reversible in control of those controlling. It also appears that the inhibition of amino acids is reversible in the control of other factors and can produce significant results.<eop><eod> The page links to the new website.<eop> This is a sample of the new website. You should also look at the other versions of the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website. The article links to the new website.<eop> The article links to the new website. The article links to the new website. The article links to the new website.<eop><eod> The image is a part of a computer made out of glass, ceramic, or metal with a similar texture, it may not be the same color or the same size, and that it may be any other metal or ceramic or glass in some way.<eop><eod><eod> This is a list of all
```

From our final result for a corpus processed slightly differently using the sentences of all the information of drugs compiled into one very long string.

```
PADDING_TEXT =  """There is insufficient experience assess, safety efficacy of ORENCIA administered concurrently with anakinra therefore such
 use is not recommended ORENCIA anakinra Formal drug interaction studies with Abciximab have not been conducted Abciximab Abciximab has been administered patients with ischemic heart disease
 of medications used treatment of angina myocardial ,infarction hypertension Abciximab. These medications have included heparin warfarin beta-adrenergic receptor blockers calcium channel an
 intravenous oral nitrates ticlopidine aspirin heparin warfarin beta-adrenergic receptor blockers calcium channel antagonists angiotensin converting enzyme inhibitors nitrates ticlopidine a
 anti platelet agents are associated with increase bleeding Heparin anticoagulants thrombolytics anti platelet agents Patients with HACA titers may have allergic or hypersensitivity reactio
 monoclonal antibodies diagnostic monoclonal ,antibodies therapeutic monoclonal antibodies The concomitant intake of alcohol Acamprosate does not affect pharmacokinetics of either alcohol o
 Pharmacokinetic studies ,indicate that administration of disulfiram or diazepam does not affect pharmacokinetics of acamprosate ,disulfiram diazepam acamprosate Co-administration of naltre
 Cmax of acamprosate naltrexone Acamprosate acamprosate No adjustment of dosage ,is recommended such patients The pharmacokinetics of naltrexone its major metabolite 6-beta-naltrexol were u
 6-beta-naltrexol Acamprosate Other concomitant therapies : In clinical trials safety profile subjects treated with Acamprosate concomitantly with anxiolytics hypnotics sedatives ( includin
 subjects taking placebo with these concomitant medications Acamprosate anxiolytics hypnotics sedatives benzodiazepines non-opioid analgesics Patients taking Acamprosate concomitantly with
 compared with patients taking either medication alone Acamprosate antidepressants. <eod> </s> <eos>"""

prompt = "The combined effects of drugs Abarelix and Testosterone "
```

✓ 0.8s Python

```
inputs = tokenizer(PADDING_TEXT + prompt, add_special_tokens=False, return_tensors="pt")["input_ids"]
prompt_length = len(tokenizer.decode(inputs[0]))
outputs = model.generate(inputs, max_length=700, do_sample=True, top_p=0.95, top_k=60)
generated = prompt + tokenizer.decode(outputs[0])[prompt_length + 1 :]

print(generated)
```

✓ 40.7s Python

```
The combined effects of drugs Abarelix and Testosterone inhibit heart rhythm in patients with ischemic heart disease. However, the rate of ischemic heart disease was not significantly affected by both drugs. The percentage of patients who did not feel ischemic heart disease was less than 10%.<eop> "The study was a very small one. The data were very well characterized. Among the patients who did not feel ischemic heart disease on the study we participated in."
```

Furthermore, we ran results from compiling 20 files and got more coherent results.

```
The combined effects of drugs Abarelix and Testosterone are known for their psycho- and pharmaco-dynamic properties. These drugs have been linked to some human deaths and ill
nesses. In the absence of an anti-opioid, no control treatment has been reported. This suggests an end to the vascular problem and a reduction in heart disease risk.<eop> The
 adverse effects of a combination of drugs and opioid and acriceptin may include: Anti-opi
```

In different instances, we got very different results, showing that there is a factor of randomness that scales as we learn the hyperparameters.

```
The combined effects of drugs Abarelix and Testosterone inhibit heart rhythm in patients with ischemic heart disease. However, the rate of ischemic heart disease was not significantly affected by both drugs. The percentage of patients who did not feel ischemic heart disease was less than 10%.<eop> "The study was a very small one. The data were very well characterized. Among the patients who did not feel ischemic heart disease on the study we participated in."
```

5. **Conclusions**

We have learned that just using word generation with prompt as finding the combined effects of the drugs is not enough. Unlike the approach using KGNN and link predictions, we were able to gain verbal results, rather than word embeddings which is what the knowledge graph would learn (Lin et al.). We are unsure of a way to take the word embeddings and the links predicted by the knowledge graph to create actual verbal outputs that other doctors or pharmacists could read and interpret for ascended medical use. Since we do not have an actual answer set that tells us what the combined effects of the drugs are, we are performing text generation based on the info of each drug we are given and the fact that being in the same file makes them related. Furthermore, the drugs in the same file have variations, and have different effects when combined with other drugs as we have tested parts of the information to combine.

We realized that even though we have a pre-trained model for English, a more specific pre-trained model relating to biology and pharmaceuticals would create a much better text generation. The results we have achieved actually seems to be reasonable if we were to combine the drugs for medical or recreational use. We ended up fine-tuning the data with our dataset rather than training it, but the test set seems to make a lot of sense despite not training it initially from this dataset.

Works Cited

Lin, Xuan, et al. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. July 2020. doi:10.24963/ijcai.2020/380.

Wang, Chenguang, et al. "Language Models are Open Knowledge Graphs." 2020. *arXiv:2010.11967*.

Wolf, Thomas, et al. "Transformers: State-of-the-Art Natural Language Processing." *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Oct. 2020, pp. 38–45.