Recommending Treatments for Comorbid Patients Using Word-Based and Phrase-Based Alignment Methods

Elie Merhej[1], Steven Schockaert[2], T. Greg McKelvey[4⋆], and
Martine De Cock[1,3]

[1] Ghent University, Ghent, Belgium
{elie.merhej,martine.decock}@ugent.be
[2] Cardiff University, Cardiff, United Kingdom
schockaerts1@cardiff.ac.uk
[3] University of Washington Tacoma, Tacoma, USA
mdecock@u.washington.edu
[4] KenSci, Seattle, USA
Greg@KenSci.com

Abstract. The problem of finding treatments for patients diagnosed with multiple diseases (i.e. a comorbidity) is an important research topic in the medical literature. In this paper, we propose a new data driven approach to recommend treatments for these comorbidities using word-based and phrase-based alignment methods. The most popular methods currently rely on combining specific information from individual diseases (e.g. procedures, tests, etc.), then aim to detect and repair the conflicts that arise in the combined treatments. This proves to be a challenge especially in the cases where the studied comorbidities contain large numbers of diseases. In contrast, our methods rely on training a translation model using previous medical records to find treatments for newly diagnosed comorbidities. We also explore the use of additional criteria in the form of a drug interactions penalty and a treatment popularity score to select the best treatment in the case where multiple valid translations for a single comorbidity are available.

1 Introduction

The number of comorbid patients keeps rising [7]. Since clinical guidelines focus on treating every disease separately, simply combining these guidelines to find treatments for comorbid patients may introduce unwanted conflicts in the combined treatments. For example, a patient diagnosed with Duodenal Ulcer (DU) is required to stop the use of any anti-inflammatory medicine, including aspirin. On the other hand, a patient diagnosed with a Transient Ischemic Attack (TIA) is required to take aspirin as part of the treatment. When a patient is diagnosed with both DU and TIA, combining the individual treatments of these

[⋆] University of Washington Occupational and Environmental Medicine Fellow diseases introduces a conflict around the use of aspirin. Therefore, a system that can automatically suggest treatments for comorbid patients is a valuable aid for clinicians.

Existing algorithmic approaches aim to combine the important information found in the clinical guidelines of the individual diseases of a comorbidity. First, they make use of computer interpretable guidelines with dedicated languages [6,3] that capture all the essential information of clinical guidelines, such as tests, procedures, etc. Then, they try to detect the different types of conflicts that arise between the procedures inside the combined treatment. Finally, they provide techniques to resolve these detected conflicts in order to find a valid treatment for a given comorbidity.

In this paper, we tackle the problem of recommending treatments for comorbid patients from a very different angle, namely by treating it as a Statistical Machine Translation (SMT) problem. The main idea is to look at a set of diagnoses (i.e. a comorbidity) as a source sentence that we want to translate into a set of procedures (i.e. a treatment), considered as a target sentence. In other words, every diagnosis in a comorbidity is considered as a “word” in a “sentence”, and similar to the methods used to automatically translate a text sentence from one language to another, we aim to translate a comorbidity into a treatment. In order to do that, we train the translation models on a corpus of medical records. As we will show in Section 4, we also take advantage of the ability of the translation system to find multiple treatment recommendations for a single comorbidity, which allows us to make use of additional criteria, such as a drug interactions penalty and a treatment popularity score to recommend the most optimized treatment.

The remainder of this paper is structured as follows: In Section 2, we discuss the advantages and drawbacks of various approaches that aim to find treatments for comorbidities. In Section 3, we provide some background on statistical machine translation systems. In Section 4, we propose a number of different algorithms used for treatment recommendations. We first demonstrate how a nearest neighbour baseline method (1-NN) can be used to solve our problem, then we present our methods based on word-based and phrase-based alignment. In Section 5, a comparative experimental evaluation on approximately 45K patient records from the MIMIC-III database, shows that our SMT approaches comfortably outperform 1-NN. Finally, we conclude with directions for future work in Section 6.

2 Related Work

Several *expert knowledge driven* approaches for recommending treatments for comorbid patients have already been proposed. Mainly, these approaches try to combine treatments of individual diseases, and repair the conflicts that arise in the combined treatments. In [16], general models for representing computer interpretable guidelines that express evidence as causation beliefs are presented. These models aim to detect the different types of interactions found in these guidelines and specify their severities. In our previous work [11] based on [17,15], mitigation operators are used to detect and repair conflicts in combined treatments. When applied, these mitigation operators offer alternative treatments based on the information that they encode.

The existing expert knowledge driven methods require the availability of clinical guidelines encoded in a machine readable way. Such computer interpretable guidelines are not readily available except for of a few of the most common diseases. This is reflected by the fact that, in the literature, the knowledge driven methods are only evaluated for a handful of specific comorbidities. The data driven approach that we propose in this paper can be directly applied to any given comorbidity, without the need for any expert knowledge encoded in a knowledge representation language. Hence, our approach is far more widely applicable than existing methods. In particular, to the best of our knowledge, there is no existing method that can be directly applied to recommend treatments for all the comborbidities that we evaluate our approach on in Section 5. The comorbidities considered in Section 5 are not hand picked, as is usually the case in work on recommending treatments for comborbid patients. Instead, we apply our approach to *all* the combordities that occur in a database of hospital discharge records.

The analysis of medical records has shown to have potential in developing and optimizing clinical treatment regiments. With the large amounts of available clinical data, there is a growing need to develop methods for automatically mining and analyzing this data. In [12], a method is proposed that aims at exploiting the rich information in doctor orders to improve clinical treatments. In [13], a system is implemented to use previously collected medical data to automatically identify the co-occurence of patient events. These prior works are not aimed at recommending treatments for comorbid patients. To the best of our knowledge, we are the first to propose a fully *data driven* approach to this end.

Concerning the specific technique we employ (see Section 3), recently, neural network based models for machine translation have emerged as a popular alternative for SMT [2,5]. Such models avoid the need for an explicit alignment between the source and target sentences, and generally consist of an encoder network, which derives a vector representation for the source sentence, and a decoder network, which maps that vector onto a sentence from the target language. While such models have achieved state-of-the-art performance, they have two drawbacks, which are important for our purposes. First, they tend to need a large amount of training data, which means that they would not be suitable, in our context, for generating recommendations for rare (combinations of) diseases, whereas it is precisely in such rare cases that a recommendation system might be most helpful to a doctor. Second, alignment based models allow us to generate explanations as to why a certain treatment is proposed (e.g. normally disease A is treated using procedure P, but because disease B is also present, procedure Q is preferred). Generating supporting explanations from neural network approaches, on the other hand, is known to be a challenging problem.

3 Word-Based and Phrase-Based Alignment

The field of Machine Translation (MT) is concerned with automatically translating the meaning of a sentence (i.e. sequence of words) $s = [s_0, \dots, s_i]$ of a source language to another sentence $t = [t_0, \dots, t_j]$ of a target language. Statistical Machine Translation (SMT) systems learn how to make such translations in a purely data driven fashion, by comparing a large number of sentences from the source language with their corresponding translation in the target language. This only requires access to a sentence-aligned corpus, i.e. SMT systems figure out automatically which words or phrases from each source sentence correspond to which words or phrases from each target sentence. This process is called alignment, and plays a crucial role in our approach for recommending treatments.

The problem of translating a sentence s to t can be expressed, using the Bayes rule, as follows:

$$\text{argmax}_t \ \text{Pr}(t|s) = \text{argmax}_t \ \text{Pr}(t)\text{Pr}(s|t) \tag{1}$$

In other words, given any sequence of words s in the source language, we want to find the sequence of words t in the target language which maximizes $\text{Pr}(t|s)$. In (1), the prior probability $\text{Pr}(t)$ is the language model probability, which models how natural the sequence t is in the target language, irrespective of the source sentence. On the other hand, the probability $\text{Pr}(s|t)$ is the translation model probability, which models how likely s is as a translation of t.

3.1 Word-Based Alignment

As mentioned above, a key challenge for SMT is to identify, in a given sentence-aligned training corpus, which words from each sentence correspond to which words from their translation. In word-based alignment models, the probability that a given word s_i from the source sentence matches the word t_j from the target sentence is assumed to be independent of the other words in these sentences. Different ways of modelling the alignment probability can be found in [4,14].

3.2 Phrase-Based Alignment

One main disadvantage of word-based alignment models is that the context of a word is not taken into account when trying to find a suitable translation. For many words, the translation depends heavily on the surrounding words that occur in the same sentence. For example, when translating a sentence from English to French, the meaning of the word “right” in the English sentence “You are right” is totally different than in the sentence “Turn to the right”. Hence, different translations of the same word are expected. Indeed, the English word “right” in the first sentence is translated to the French word “raison”, while the same English word “right” in the second sentence is translated to the French word “droite”. The basis of phrase-based alignment is to decompose the input sentence from the source language into phrases (i.e. natural sequences of words), find a translation for every phrase, then re-order these phrase translations and combine them to produce the target sentence.

A popular method for finding a phrase-based alignment is to learn the phrase translations from a corpus that has already been aligned using a word-based translation model [10]. In particular, the method relies on two word alignments: from source language to target language and from target language to source language. The two word alignments are then combined by doing an intersection of the aligned words, and then finding additional words that are not present in the intersection using different heuristics. Then, all aligned phrase pairs that are consistent with the combined word alignment are collected. Consistency is defined such that the words of a phrase pair are only aligned to each other, and are not aligned to words that are not inside the phrase pair. The extracted phrases and their translations constitute a phrase translation table. A phrase-based decoder (usually based on beam search) is finally used to generate the output sentence by re-ordering phrase translations. More details about this method can be found in [10].

N-best Phrase-based Alignment A phrase-based alignment model is generally expected to output the most likely translation of an input sentence. However, some applications benefit from having a set of alternative translations. A common method to find the n-best translations of an input sentence is to apply a phrase-based translation model to generate candidate translations. Subsequently, the probability score from the phrase-based translation model can be combined with additional features, where available, to produce a list of the n-best translations [9].

In Section 4, we use all methods described above, i.e. word-based alignment, phrase-based alignment, and n-best phrase-based alignment, to recommend treatments for comorbid patients.

4 Treatment Recommendation Methods

In this paper, we use alignment based translation models to map lists of diagnoses to lists of procedures. To train the translation model, we assume the availability of a large database of medical records, showing for each clinical event of each patient (e.g. a hospital admission) what diagnoses were made, what procedures were proposed by the providers and which drugs were prescribed. Some examples of such records are shown in Table 1. In medical records, the diagnoses and procedures of every admission are usually encoded using a standard encoding. For example in Table 1, the ICD-9 encoding is used, where the diagnosis codes "V3001" and "74783" refer to "Single liveborn, born in hospital, delivered by cesarean section" and "Persistent fetal circulation" respectively, and the procedure codes "9604" and "9671" refer to "Insertion of endotracheal tube" and "Continuous invasive mechanical ventilation" respectively. The patients in Table 1 are considered comorbid because they have a variety of different diagnoses that each require treatment.

Admission_ID	**Diagnoses**	**Procedures**	**Drugs**
...	...	...	...
196807	V3001, 74783, 7700, 7756, 7761, 77181, V053	9604, 9671, 9390, 0331, 9955	Heparin, Ampicillin_Sodium, Gentamicin, Pediatric_Vitamins, Potassium_Chloride, Sodium_Bicarbonate, Sodium_Chloride
100589	3940, 9982, 9971, 49320, 4239, E8788, 2449, 53081, 4019	3596, 370, 3723, 8856, 8872	Acetaminophen, Magnesium_Hydroxide, Atropine_Sulfate, Clonazepam, Hydrochlorothiazide, Levothyroxine_Sodium, Mirtazapine, Morphine_Sulfate, Nitroglycerin, Oxazepam, Oxycodone_Acetaminophen, Pantoprazole, Potassium_Chloride, Prochlorperazine, Simethicone, Venlafaxine, Vioxx, Zebeta
...	...	...	...

Table 1. Extract from medical records containing diagnose, procedure and drug information for every hospital admission.

An important difference between our medical treatment recommendation setting and the standard machine translation setting relates to the role of word ordering. While in standard settings, word ordering plays a critical role, the order in which diagnoses and procedures are ordered in our setting may be arbitrary. However, in practice, these orderings are not completely arbitrary, in the sense that the most important diagnoses and procedures are often listed first. As we will see in Section 5, this can be exploited by the translation model. For evaluation purposes, however, the task that we consider is to predict an (unordered) set of procedures.

We now explain the methods we propose for recommending a suitable treatment for a patient given their diagnosed comorbidity. While WBT, PBT and NPBT below are based on SMT approaches, we also consider a nearest neighbour baseline approach (1-NN) that utilizes treatments of previously diagnosed comorbidities in a simple way to find the best treatment for a new comorbidity case. We compare the effectiveness of these approaches in Section 5.

4.1 1-NN: 1-Nearest Neighbor with Jaccard Similarity (Baseline)

This method serves as a baseline approach to find a treatment for a newly diagnosed set of diseases, given a database that contains records of previous comorbidities, as well as their corresponding treatments. The idea behind this method is to find the most similar set of diagnoses across all records in the database, and to use the corresponding treatment for this case as the recommended treatment. To measure similarity, we use the Jaccard measure, which is a standard measure of similarity between sets, and is defined as $J(C, X) = |C \cap X| \ / \ |C \cup X|$ for two sets of diagnoses C and X. In the case where multiple records are equally similar, the first record found is then used.

4.2 WBT: Word-Based Translation

This is the first and most basic alignment method that we use to translate a given set of diagnoses C into a set of procedures T. A detailed explanation about this method is found in Section 3. When using word-based translation, every diagnosis in C is treated as a “source word”. It gets individually translated into a procedure, i.e. a “target word”. Every translation is also given a corresponding translation probability. Essentially, there are two steps to training the translation model: finding the most likely alignment for every source word, and computing a probability table given the alignment. In practice, an Estimation Maximization (EM) algorithm can be used where the probability from a given alignment is initially estimated, and then the alignment is improved based on the new probabilities. This iterative process converges to give us the final probability table [4]. Finding the recommended treatment T then comes down to choosing for each diagnosis, the most likely translation (i.e. procedure) from the final probability table. When using this model, we are effectively ignoring any interactions between diagnoses and procedures. In other words, we are not taking into account conflicts that arise due to comorbidity.

4.3 PBT: Phrase-Based Translation

The main limitation of the word-based model in our context is the fact that interactions between procedures (and between procedures and diseases) are being ignored. This can be addressed by using a phrase-based translation model. Instead of translating every word of a sentence separately, this model groups sequential words together into phrases, then aims to find translations for these phrases. Finding the translation of a sentence consists then of combining these phrase translations.

In our case study, this translation model will solve the problem of arising conflicts between diagnoses (and procedures) of a comorbidity in the following way. When two or more “conflicting” diseases require a special treatment, the model will detect this instance provided that there are enough records in the training data in which these conflicting diseases occur. If the exact set of diagnoses has not been observed before, the model will try to split the set into subsets (i.e. phrases) that it can adequately translate, which contain as many diseases as possible. By doing so, all the conflicts between the diseases within each subset will be avoided effectively. However, any interactions between diseases that belong to different subsets will be unaddressed.

4.4 NPBT: N-Best Phrase-Based Translation With Cost Minimization

When using PBT, the recommended treatment is chosen as the most likely translation of the given list of diagnoses. This strategy is optimal in cases where we have no other information. When it comes to recommending procedures, however, we can take advantage of existing databases that describe the interactions between the drugs used during these procedures [1], among others. To this end, we use the phrase-based translation model to find the n-best translations, and then use the external knowledge to help select the best translation among these. In particular, we look to assign a penalty cost for every possible treatment, and then minimize this cost to find the best treatment.

NPBT+Drug: Drug Interactions Penalty The first type of information that we use to score treatments is the number and severity of drug interactions that are present in them. In previous work, we already found the usefulness of taking into account such interactions when recommending treatments [11]. In order to create a drug interactions penalty, we take into account the number of drug interactions as well as the severity of every drug interaction that is found inside a treatment.

In order to find the drug interactions penalty of a treatment, we need to translate it into a set of drugs. To translate procedures into drugs, we can proceed in the same way as for translating diagnoses into procedures, by learning a phrase-based alignment model from the procedures and drugs in each record of our database (cfr. Table 1). Now, given a comorbidity C, we can translate C into a set of procedures T, which can in turn be translated into a set of drugs R. Note that when translating from procedures to drugs, we also use the phrase-based translation model to find the n-best translations.

To obtain the drug interactions penalty of a treatment, we need to use a drug interactions database. Let $I = \{(a_0, b_0, \beta_0), \ldots, (a_m, b_m, \beta_m)\}$ be a drug interactions database where a_i and b_i are two drugs that are known to interact, and β_i is an integer that represents the severity level of that drug interaction. The drug interaction penalty p of a treatment T is then calculated as $p = \sum\{\beta \mid (a, b, \beta) \in I \wedge \{a, b\} \subseteq R\}$ i.e. the sum of the severity levels of all drug interactions found in the set of drugs R obtained by translating T. Since we consider n plausible translations for every treatment T, we get n drug interactions penalties. The translation that yields the lowest drug interactions penalty is then considered the best procedures-to-drugs translation of T. In the case where multiple translations have the lowest drug interactions penalty, the most probable translation is then considered.

The treatment that has the lowest drug interactions penalty from the n best diagnoses-to-procedures translations of a comorbidity is then considered the best treatment. In the case where multiple treatments have the same lowest drug interactions penalty, the one that is found to be the most probable translation is then considered the best treatment.

NPBT+Drug+Pop: Drug Interactions Penalty With Procedure Popularity In practice, there may be multiple valid procedures to treat a given disease. Some of these tend to be more popular than others, e.g. because they are cheaper or have a lower risk. While we do not have access to any explicit knowledge about the popularity of different treatments, such popularity scores can be estimated from the database of patient records.

This method of selecting the best valid treatment extends NPBT+Drug in the following way. In the case where multiple treatments have the same lowest drug interactions penalty cost, the treatment with the highest popularity score is now preferred as the best treatment. The popularity score of a treatment is calculated by adding the popularity scores of all the procedures in the treatment. The popularity score of a procedure is calculated by counting the number of times this procedure has been prescribed in the database of medical records. Similar to NPBT+Drug, when multiple treatments have the same lowest drug interactions penalty and the same highest treatment popularity score, the one that resulted from the more probable phrase-based translation is then considered the best treatment.

5 Experiments and Results

To train and evaluate our methods, we use the MIMIC-III database [8] of medical records, from which we extract the diagnoses, procedures and drugs used per hospital admission. The diagnoses and procedures are in the form of ICD9 codes, while the drugs are referred to by their short name. Note that the drug names have been inspected and cleaned manually to be consistent with the drug names in the drug interactions database. We only consider the admissions that contain comorbidities (i.e. at least 2 diagnoses), and all the information needed for the treatment recommendation methods (no empty fields for diagnoses, procedures or drugs). From the MIMIC-III database, we get 44,223 different admissions. The comorbidity size (i.e. the number of diagnoses) in every admission varies from 2 to 39. A graph containing the number of admissions for every comorbidity size is shown in Fig. 1.

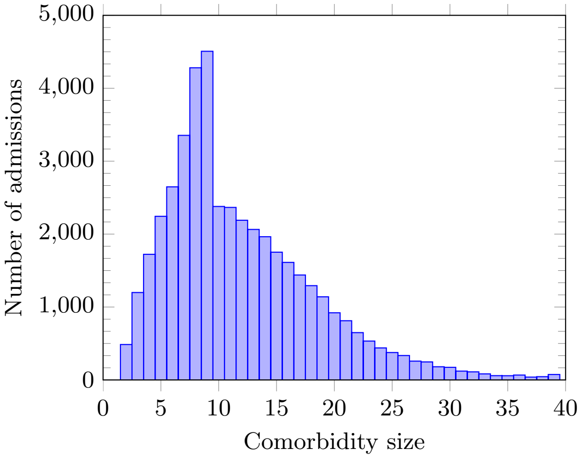

Fig. 1. The number of admissions for every comorbidity size taken from the MIMIC-III database.

We divide this database into two disjoing sets: a training set and a testing set. The training set consists of 80% of the total number of admissions, while the testing set consists of 20% of the total number of admissions. From the MIMIC-III database, we get 35,378 admissions in the training set and 8,845 in the testing set. The task now becomes: using the training set of admissions, find treatments for the comorbidities in the testing set. When applying a treatment recommendation method, the training dataset is used to train the translation model and the language model. These models are then applied to translate the testing dataset. In our implementation, we use the Moses decoder [9], one of the most popular state-of-the-art decoders to create the word-based and phrase-based alignment models.

In NPBT, we use the drug interactions database described in [1]. We cleaned this database by cleaning the drug names and removing duplicated interactions. We extracted from this database two metrics that describe the severity of a drug interaction: severity level and contraindication. There are 5 possible severity levels (1-5), which increase the drug interactions penalty of a treatment by 1-5 respectively. Additionally, a contraindication is a boolean which, when set to true, indicates that the drug interaction should be avoided at all costs. To reflect this severity measure, the drug interactions penalty of a treatment is increased by 100 when a drug interaction with a contraindication is detected. In NPBT+Drug and NPBT+Drug+Pop, we start from the best n=10 translations for both translations needed in these methods (from diagnoses to procedures and from procedures to drugs).

To evaluate every recommended treatment, we use the F_1 score metric, which we calculate as follows. Let A be the set of recommended procedures by a given model and B the actual set of procedures from the testing database. We write $|A|$ and $|B|$ for the number of procedures in the sets A and B respectively. The F_1 score is given by $F_1 = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$, with precision $= |A \cap B| / |A|$ and recall $= |A \cap B| / |B|$. The precision and recall can also be expressed using true positives (TP), false positivies (FP) and false negatives (FN) in the following way: precision $=$ TP / (TP + FP) and recall $=$ TP / (TP + FN). Since we want to evaluate every method based on all the recommended treatments, we use the micro-average and macro-average of the F_1 score. The micro-average F_1 score consists of individually averaging the TP, FP and FN of all the sets, then calculating the F_1 score using these averages. The macro-average F_1 score consists of averaging the precision and recall of all the sets, then calculating the F_1 score using these averages.

	Average Precision	**Average Recall**	**Mirco-average F1 score**	**Macro-average F1 score**
1-NN	0.334	0.303	0.281	0.318
WBT	0.255	0.531	0.323	0.345
PBT	0.268	0.564	0.348	0.364
NPBT+Drug	0.304	0.595	0.389	0.403
NPBT+Drug+Pop	0.321	0.607	0.414	0.420

Table 2. The average precision, average recall, micro-average and macro-average F_1 scores for every treatment recommendation method.

The results are shown in Table 2. From the table, we notice that the word-based translation approach (WBT) already performs better than the baseline approach. However, using a phrase-based translation model (PBT) leads to a further improvement of the results. Using NPBT gives the most accurate treatment recommendations, and in particular NPBT+Drug+Pop, where the procedure popularity score is used in addition to the drug interactions penalty to select the best treatment out of the n-best valid translations. Note that the Moses decoder allows up to 100-best translations for every input sentence. We also evaluated method NPBT+Drug+Pop when using n=20 and n=100, but we found the differences in F_1 score to be negligible.

As previously mentioned, word ordering plays an important role during the training of the alignment model. In the MIMIC-III database, the diagnoses and procedures are ordered based on their priorities, from highest to lowest. In other words, the primary diagnosis and the primary procedure of an admission are listed first, then the remaining ones are listed from the most important to the least important one. We compare this default setting (ordered by priority) with two different word ordering settings: Keep Primary Then Sort (KPTS) and Keep Primary Then Random (KPTR). In KPTS, the primary diagnosis and procedure of the source and target sentences in the training dataset are listed first, but the remaining ones are sorted alpha-numerically in increasing order. In KPTR, the primary diagnosis and procedure are also listed first, but the remaining ones are randomly shuffled. We train two new phrase-based translation models using KPTS and KPTR respectively and apply PBT to find the treatment recommendations. The results are shown in Table 3. From the table, we notice that the recommendation method is less accurate when using KPTS and KPTR compared to the default word ordering (ordered by priority) found in the medical database.

	Average Precision	**Average Recall**	**Mirco-average F1 score**	**Macro-average F1 score**
Ordered by Priority	0.268	0.564	0.348	0.364
KPTS	0.261	0.565	0.347	0.357
KPTR	0.263	0.536	0.338	0.354

Table 3. The average precision, average recall, micro-average and macro-average F_1 scores for PBT using different word orderings during the training process.

6 Conclusion

In this paper, we presented the first fully data driven method to find treatments for patients that are diagnosed with comorbidities. Instead of trying to combine clinical guidelines and trying to repair the conflicts that arise in the combined treatments, we used word-based and phrase-based alignments to find direct mappings from a comorbidity to a recommended treatment. To improve this translation based approach, we also detect drug interactions and calculate procedure popularity scores to select the best treatment out of different valid translations. Contrary to manual approaches that aim to prescribe treatments in an evidence based way, this approach allows us to take advantage of previous medical records to recommend treatments for newly diagnosed comorbidities. From our experimental results, we found that using the method NPBT+Drug+Pop, which uses a drug interactions penalty and a procedure popularity score to find the best translation, give the most accurate treatment recommendations. The next step in this research would be to do an error analysis by looking into divergences between recommended treatments and treatments that were actually prescribed, and use this information to improve the treatment recommendation methods. There can be many reasons for why a recommended treatment diverges from a prescribed treatment: (1) the recommendation is wrong, (2) the recommendation is slightly different from the prescribed treatment, in the sense that the recommended procedures are very similar to the prescribed procedures, though not identical. A more coarse grained evaluation, where related procedures are grouped together (i.e. using ICD-grouping software) would bring this to light. (3) The recommendation is a valid alternative, and possibly even better than the treatment that was given in practice. This could be an indication of inefficiency, error, or even fraud from the provider.

References

1. Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, N.P., Vilar, S., Brochhausen, M., Samwald, M., Rastegar-Mojarad, M., et al.: Toward a complete dataset of drug–drug interaction information from publicly available sources. Journal of biomedical informatics 55, 206–217 (2015)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
3. Boxwala, A.A., Peleg, M., Tu, S., Ogunyemi, O., Zeng, Q.T., Wang, D., Patel, V.L., Greenes, R.A., Shortliffe, E.H.: Glif3: a representation format for sharable computer-interpretable clinical practice guidelines. Journal of biomedical informatics 37(3), 147–161 (2004)
4. Brown, P.F., Pietra, V.J.D., Pietra, S.A.D., Mercer, R.L.: The mathematics of statistical machine translation: Parameter estimation. Computational linguistics 19(2), 263–311 (1993)
5. Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)
6. Fox, J., Johns, N., Rahmanzadeh, A.: Disseminating medical knowledge: the proforma approach. Artificial intelligence in medicine 14(1), 157–182 (1998)
7. Jakovljevic, M., Ostojic, L.: Comorbidity and multimorbidity in medicine today: challenges and opportunities for bringing separated branches of medicine closer to each other. Psychiatr Danub 25(Suppl 1), 18–28 (2013)

8. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.w.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Scientific data 3 (2016)
9. Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al.: Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions. pp. 177–180. Association for Computational Linguistics (2007)
10. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. pp. 48–54. Association for Computational Linguistics (2003)
11. Merhej, E., Schockaert, S., McKelvey, T.G., De Cock, M.: Generating conflict-free treatments for patients with comorbidity using asp. In: 8th International workshop on Knowledge Representation for Health Care (KR4HC'16); held in conjunction with HEC 2016: Health-exploring complexity: an interdisciplinary systems approach. pp. 93–100 (2016)
12. Sun, L., Liu, C., Guo, C., Xiong, H., Xie, Y.: Data-driven automatic treatment regimen development and recommendation. In: KDD. pp. 1865–1874 (2016)
13. Titus, A., Faill, R., Das, A.: Automatic identification of co-occuring patient events. In: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. pp. 579–586. ACM (2016)
14. Vogel, S., Ney, H., Tillmann, C.: Hmm-based word alignment in statistical translation. In: Proceedings of the 16th conference on Computational linguistics-Volume 2. pp. 836–841. Association for Computational Linguistics (1996)
15. Wilk, S., Michalowski, W., Michalowski, M., Farion, K., Hing, M.M., Mohapatra, S.: Mitigation of adverse interactions in pairs of clinical practice guidelines using constraint logic programming. Journal of biomedical informatics 46(2), 341–353 (2013)
16. Zamborlini, V., Hoekstra, R., Da Silveira, M., Pruski, C., ten Teije, A., van Harmelen, F.: Generalizing the detection of internal and external interactions in clinical guidelines. In: HEALTHINF. pp. 105–116 (2016)
17. Zhang, Y., Zhang, Z.: Preliminary result on finding treatments for patients with comorbidity. In: Knowledge Representation for Health Care, pp. 14–28. Springer (2014)