Targeting abbreviated medication names with NLP

Mara Alexeev, MD, MPH [1,]
@MaraAlexeev
mara.alexeev@childrens.harvard.edu

Amir Kimia, MD[2] Marvin B. Harper, MD[2] Assaf Landschaft, M.Sc.[3] Al Ozonoff, PhD, CPPS[4, 5]

[1] Department of Pediatrics, Boston Children's Hospital
[2] Division of Emergency Medicine, Boston Children's Hospital
[3] Boston Children's Hospital
[4] Precision Vaccines Program, Boston Children's Hospital
[5] Department of Pediatrics, Harvard Medical School

Introduction

According to The Joint Commission, medication names should not be abbreviated as misinterpretation may lead to administration of incorrect medication.

Objectives

1. Identify abbreviated medication names in a first step towards elimination.

Methods

Retrospective chart review of pediatric ED consult notes at a tertiary pediatric center in 2019. We targeted consult notes due to potential differences in expertise between the documenting and reading providers. Abbreviated medication names were identified using 2 Natural Language Processing methods: a) named-entity recognition (NER) and b) Regular Expressions (RegEx). The NER model was a pre-trained model called MED7, identifying 7 categories: drug names, route of administration, frequency, dosage, strength, form, duration. We fine-tuned the model on a small sample of annotated documents from our hospital. RegEx was used to identify strings likely to be medications given surrounding text context. The abstracted lists were then matched against both generic and commercial medication names using the following lists: RxNorm, National Drug Code Directory (NDCD). The remaining terms were then sorted by frequency, and the top 2,295 were reviewed.

Abbreviated medication names are present in many notes and can be ambiguous.

We can identify shortened medication names with natural language processing (NLP) tools.

Boston Children's Hospital
Until every child is well™

Results

There were 29,877 consult notes available for review. We narrowed the corpus of documents to services more likely to prescribe medication (endocrinology, cardiology, neurology) and eliminated those services where medication treatment is a smaller part of their practice (orthopedics, urology) leaving 16,010 notes for review.

We identified 8,288 unique medication terms using NER and 2,671 using RegEx. The union of the two lists of medications had 9,541 unique medication terms. After cross referencing against RxNorm and NDCD 7,248 unique medication terms remained. After reviewing all terms occurring 2 or more times, a subset of terms, which were identified by an author as likely abbreviated medication names, was curated and is presented with frequency counts in Table 1.

Term	Count	Potential Meaning(s)
vanc	101	vancomycin
ctx	98	ceftriaxone, Cytoxan
vanco	67	vancomycin
midaz	62	midazolam
ceftaz	39	ceftazidime
lzp	35	lorazepam
amox	32	amoxicillin
norepi	30	norepinephrine
tazo	20	tazobactam
oxc	20	oxcarbazepine, ofloxacin, oxycodone
oxcarb	18	oxcarbazepine
tacro	16	tacrolimus
vgb	15	vigabatrin
ivmp	14	intravenous methylprednisolone
mmf	14	mycophenolate mofetil, maxillomandibular fixation
acei	14	angiotensin converting enzyme inhibitor, acetylcholinesterase inhibitors
phb	13	phenobarbital
ara-c	11	cytarabine
ruf	11	rufinamide, rectourethral fistula
zns	10	zonisamide, zolmitriptan nasal spray
ino	10	inhaled nitric oxide, internuclear ophthalmoplegia, inhalation, inositol, inotuzumab ozogamicin
clb	8	clobazam, chlorambucil
pip	8	peak inspiratory pressure, piperacillin
mero	8	meropenem
lr	6	lactated ringers, low risk
hts	6	hypertonic saline
vigab	5	vigabatrin
cbz	4	clobazam, carbimazole, carbamazepine
mdz	4	midazolam, metronidazole

Conclusions

With natural language processing tools we can create libraries of abbreviated medication names. Future studies should include domain-expert champions to help interpret domain-specific expressions. These libraries should be incorporated into documentation tools within electronic medical records.