Extracting Number of Trial Participants From Abstracts of Randomized Controlled Trials

- A collaboration project between
Centre for Health Informatics, University of New South Wales, and
Department of Health Science and Technology, Aalborg University

Marie Juul Hansen, mjha03@hst.aau.dk
Nana Rasmussen, nora04@hst.aau.dk
Grace Chung

TTeC 2008
10th of June 2008, Tromsø, Norway

Motivation

- Telemedicine - Aid to decision-making in healthcare

- The amount of information available to answer questions are constantly increasing

  → The use of keyword-based search methods to locate the answers

- These methods tend to overload with a lot of irrelevant information

  → The precision in the search result should be improved

PICO Framework

- Patient/Problem: What is the primary disease?
  What are the characteristics of the patient?
- Intervention: What is the main intervention?
- Comparsion: What is the main intervention compared to?
- Outcome: What is the effect of the intervention?

- Patient/Problem
  - Trial Participants
  - Age
  - Gender

  } Patient

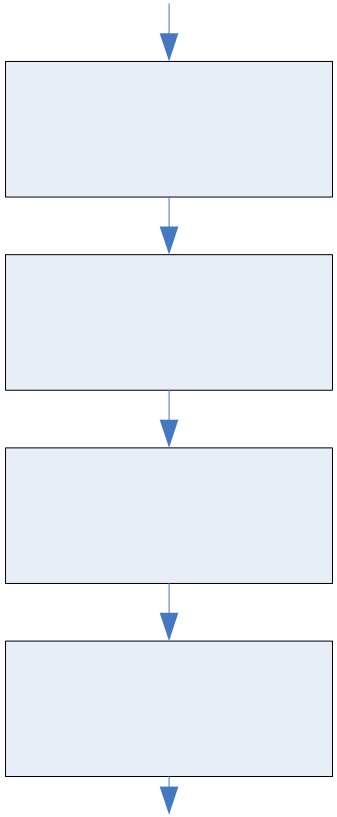
  - Disease
  - Symptoms

  } Problem

Data

- 223 labeled abstracts achieved from PubMed
  - 148 training set
  - 75 test set
- Manually detected number of trial participants for each abstract

Method



- Preprocessing
- Feature Selection and Extraction
- Choice of classification algorithm
- Postprocessing

Abstra

Preprocessing

Example :

We determine whether a continuous infusion of ketamine can decrease the severity of a moderately severe acute asthma exacerbation by a clinically significant 2 points using a 15 point Pulmonary Index scoring scale. 0 STUDYOBJECTIVE A

A double-blinded, randomized, placebo-controlled trial was performed to evaluate patients aged 2 to 18 years who presented to a pediatric emergency department with an acute asthma exacerbation. 2 METHODS P

Exclusion criteria included temperature greater than 39 degrees C (102 degrees F), focal infiltrate on radiograph, or any glucocorticoid use in the last 72 hours. 2 METHODS P

Eligible patients received 3 treatments with albuterol, ipratropium bromide, and a dose of oral or parenteral glucocorticoids. 2 METHODS M

If the Pulmonary Index score remained 8 to 14, enrollment proceeded. 2 METHODS M

All enrolled patients received continuous nebulized albuterol at 10 mg/hour and were randomized to receive an intravenous bolus of 0.2 mg/kg of ketamine, followed by a 2 hour ketamine infusion at 0.5 mg/kg per hour or an equal-volume regimen with normal-saline placebo. 3 METHODS I

A Pulmonary Index score was performed on patients at 0, 30, 60, 90, and 120 minutes. 3 METHODS E

Sixty-eight patients were enrolled, with 33 patients randomized to the ketamine infusion and 35 patients randomized to placebo. 3 RESULTS R

Mean ages of patients enrolled, chronic severity of asthma, and duration of symptoms before presentation were similar between groups. 4 RESULTS R

At enrollment, the mean Pulmonary Index score in the placebo group was 10.3 +- 1.1 versus 10.5 +- 1.5 for the ketamine group (difference of means 0.2,, 95% confidence interval [CI] -0.5 to 0.8). 4 RESULTS R

Sixty-two patients completed the entire 2 hour infusion protocol. 4 RESULTS R

No significant difference between groups was seen in rate of improvement in the Pulmonary Index score at completion. 4 RESULTS R

The mean decrease in the Pulmonary Index scores at the end of the infusion was 3.6 +- 1.3 in the placebo group versus 3.2 +- 2.0 in the ketamine group (difference of means 0.4,, 95% CI -0.4 to 1.3). 5 RESULTS R

No short-term adverse effects necessitating discontinuation of the infusion or adverse behavioral impacts at 48 hours after discharge were noted. 5 RESULTS R

We conclude that ketamine given at 0.2 mg/kg followed by an infusion of 0.5 mg/kg per hour for 2 hours provided no incremental benefit to standard therapy in this cohort of children with a moderately severe asthma exacerbation. 7 CONCLUSION O

NLP-preprocessing

Natural Language Processing on the abstracts :

WORD	STEM	POS
68	68	CD
patients	patient	NNS
were	be	VBD
enrolled	enrol	VBN
,	,	,
with	with	IN
33	33	CD
Patients	patient	NNS
randomized	randomize	VBD
to	to	TO
the	the	DT
ketamine	ketamine	NN
infusion	infusion	NN
and	and	CC
35	35	CD
Patients	patient	NNS
randomized	randomize	VBN
to	to	TO
placebo	placebo	NN
.	.	.

- Easier to gain information
- Representation by grammatical class

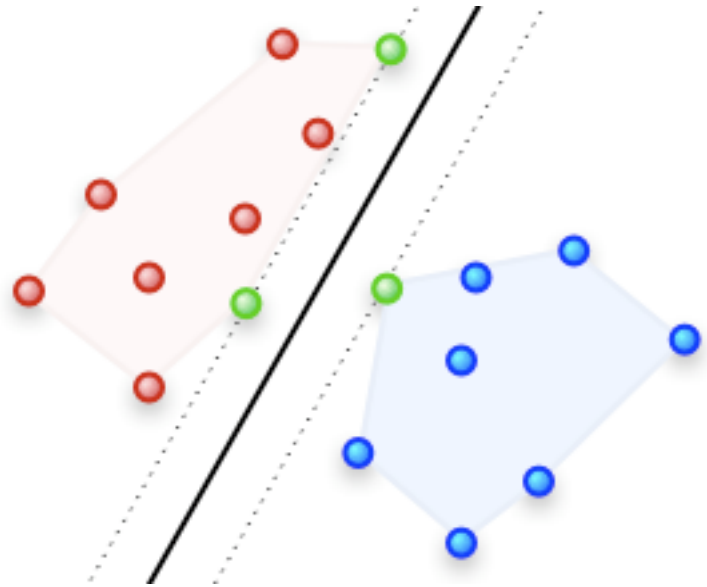
Feature Extraction

- Feature selection based on analysis and former studies
- Characteristic in different context
- Patient words
- Verb carateristics
- [Number][Disorder][Patients]

Features	Centre of Attention
int>10	*Ii*
len(int)<5	*Ii*
otherIntegers	*Sentence*
patientGroup	*Ii[JJ]*[patientGroup]*
patientWindow	*Ii Wi1-Wi4*
N=	*N=Ii*
patientGroupBeforeN=	*[patientgroup]n=Ii*
POSbefore	*POSi-1 Ii*
POSafter	*POSi+1 Ii*
wordBefore	*Wi-1 Ii*
wordAfter	*Wi+1 Ii*
verbInSentence	*Sentence*
sentenceNumber	*Sentence*
label	*Sentence*

Classification

- Binary classification task
- Three classification algorithms are tested
  - Linear Support Vector Machine (SVM)
    - 94th dimensional feature space
    - Linear separable



Postprocessing

- Reduce the number of false positive predictions
- 68 patients were enrolled, with 33 patients randomized to the ketamine infusion and 35 patients randomized to placebo.
- According to definition the highest number is chosen

Results

- Rule-based baseline system returns the largest interger that is:
  - Larger than 10
  - Not followed by an unit
  - Followed by [Adjective]* [Participant-related Noun]

Classifier	Precision	Recall	F-measure	Accuracy
Baseline	0.917	0.603	0.73	58.67%
SVM	0.965	0.743	0.84	97.03%

- 'State of the Art'-algorithm: F-measure = 0.85
  [Demner-Fushman and Lin, 2005]

Predicted / Actual	Trial Participants	Other
Trial Participants	53	19
Other	2	632

Results

- Classification errors mainly due to:
  - No number chosen in the abstract
  - Two numbers of trial participants present in abstract
  - NLP- preprocessing errors

- Correction of NLP-preprocessing errors

Classifier	F-measure	Accuracy
Before	0.84	97.03%
After	0.86	97.45%