Information Extraction From Medical Notes

Scott Kraus

Catherine Blake

Suzanne West

University of North Carolina at Chapel Hill, USA

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE

UNC SCHOOL OF PUBLIC HEALTH

Motivation

- Medical Notes
  - Thousands of medical notes in any hospital system
  - Unstructured text not amenable for data mining
- Goal :
  - Accurately extract information from each note
    - Drug
    - Dosage
    - Method of delivery

Materials

- Data Set
  - Transcribed dictated notes
  - 12,222 patients in UNC Hospital
  - Anonymized
- Part of UNC DEcIDE project
  - Identify over-dosage of diabetes patients
- Gold standard
  - structured data sets
  - Physician annotated

Method Summary

Method	Trigger	Output
1	Drug name	± 50 chars from drug
2	Drug name or 'Insulin' or 'Diabetes	•Sentence •Trigger
3/4	Word > 3 letters & Number & Measurement Unit	•Drug •Amount •Method of delivery

Example – Method 1

- Trigger: drug name

Trigger: Drug name

We will begin the patient on her home dose of **Glucotrol**, 5 milligrams PO QD and we will give sliding scale insulin as needed.

- Output: 50 characters either side of drug

We will begin the patient on her home dose of **Glucotrol**, 5 milligrams PO QD and we will give sl

50 Character limit

Example – Method 2

- Trigger: drug, ‘diabetes’, ‘insulin’

Trigger: ‘diabetes’

The other diabetes related medications are Actos 45 mg and Lantus at bedtime 20 units.

Trigger: Drug name

- Output: Entire Sentence

Example – Method 3/4

- Trigger: word > 3 letters, dosage, measurement

Trigger: word > 3 letters

Trigger: dosage

He continued on Metformin 500 mg p.o. b.i.d.

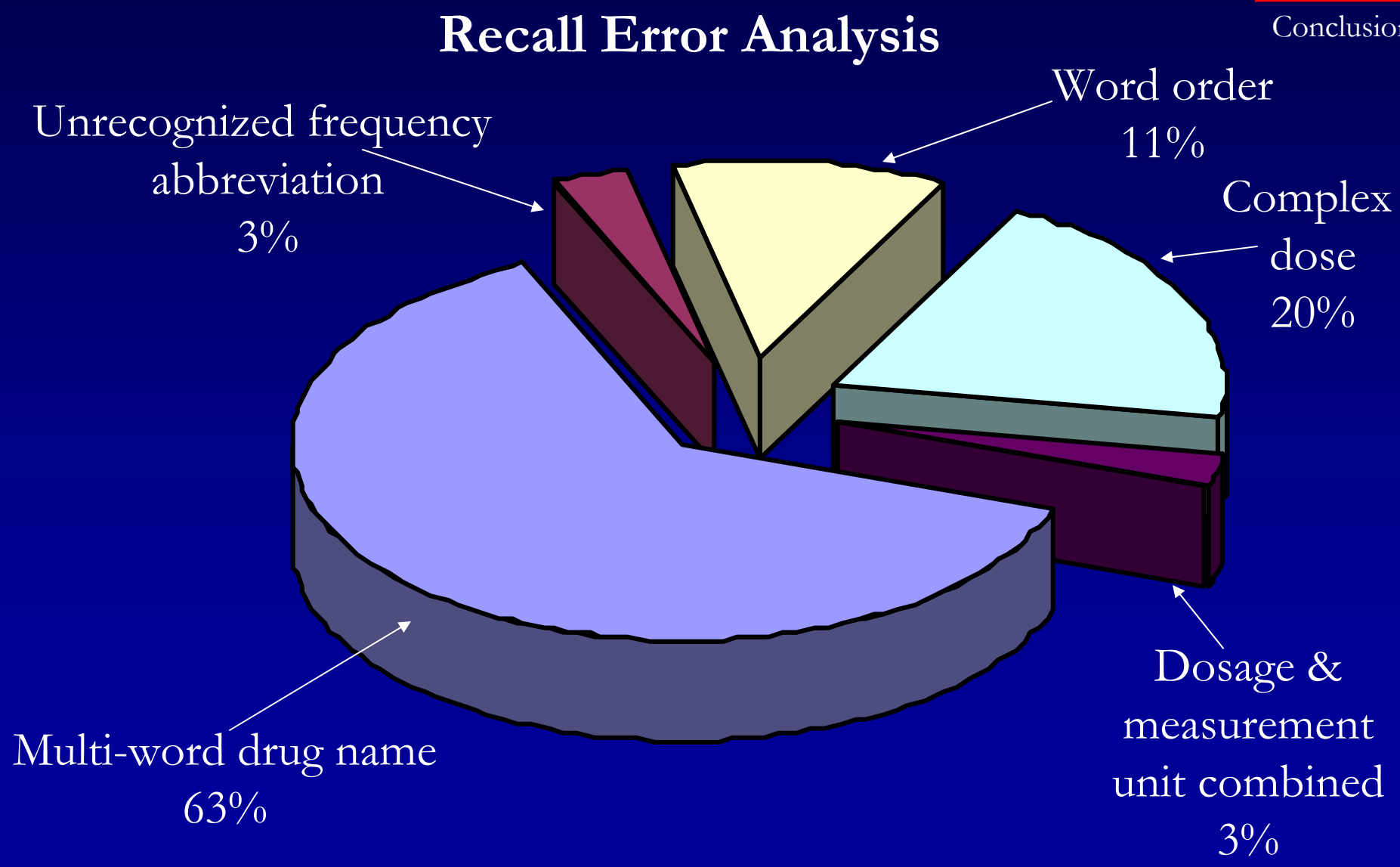
Trigger: measurement

- Output: Drug, dosage, unit, delivery, frequency

Method 3/4 – Advantages

- Processing Speed
  - 24-26 patient notes/sec
  - Pentium 4 2.0GHz processor
- Drug name not used in trigger
  - Identifies new drugs
  - Missing or obscure drugs do not lower performance

Method 4 Limitations



Results of Method 4

	Training Set		Test Set	
	Precision	Recall	Precision	Recall
Drug	86.96%	82.76%	86.81%	82.78%
Amount	100%	91.39%	100%	96.08%
Method of Delivery	100%	62.03%	100%	91.89%
Prescribed Frequency	100%	41.73%	100%	48.15%
Average	**96.74%**	**69.48%**	**96.70%**	**79.72%**

Conclusion

- Average F-Measure (Test Set)
  - Drug 84.80%
  - Amount 98.04%
  - Method of Delivery 95.95%
  - Prescribed Frequency 74.06%
  - **Average 88.21%**
- Scales well : 24-26 patient notes/sec

Questions or Comments ?

Scott Kraus (scott@unc.edu)
Catherine Blake (cablake@email.unc.edu)
Suzanne West (sue_west@med.unc.edu)