abridge

# Medication Regimen Extraction From Clinical Conversations

Sai P. Selvaraj, Sandeep Konam
{prabhakarsai, san}abridge.com

## Motivation

- Patients forget 40-80% of the medical information provided by healthcare practitioner immediately [1]
- Clinicians spend up to 49.2% of their overall time on EHR causing burnout [2]
- Automatically extracting information from medical conversations can:
  - Improve patient's recall
  - Help patients follow through with their care plan
  - Reduce the documentation burden for doctors

## Task

- Medication Regimen tags (MR tags) extraction from Doctor-Patient conversations
- MR tags:
  - *Medication* name
  - Medication *Dosage*
  - Medication *Frequency*

| (Timestamp) Part of Transcript | Summary | Medication | Dosage | Frequency |
|---|---|---|---|---|
| (1028.3s) So, let's, I'm going to have them increase the mg, uh, Coumadin level, so that, uh, like I said, the pulmonary embolism doesn't get worse here.<br>(1044.9s) Yeah.<br>(1045.2s) Yeah.<br>(1045.4s) Increase it to three point five in the morning and and before bed | Increase Coumadin to 3.5 mg to prevent pulmonary embolism from getting bigger | Coumadin | 3.5 mg | Twice a day |

Table 1: Example of our annotations grounded to the transcript segment.

## Dataset

- De-identified, fully-consented 6,693 real doctor-patient conversations
- Recorded in a clinical setting using distant microphones of varying quality
- Transcribed by medical experts
- Annotated (and grounded) by expert annotators with:
  - *Summaries*
  - *MR tags*

## Approach

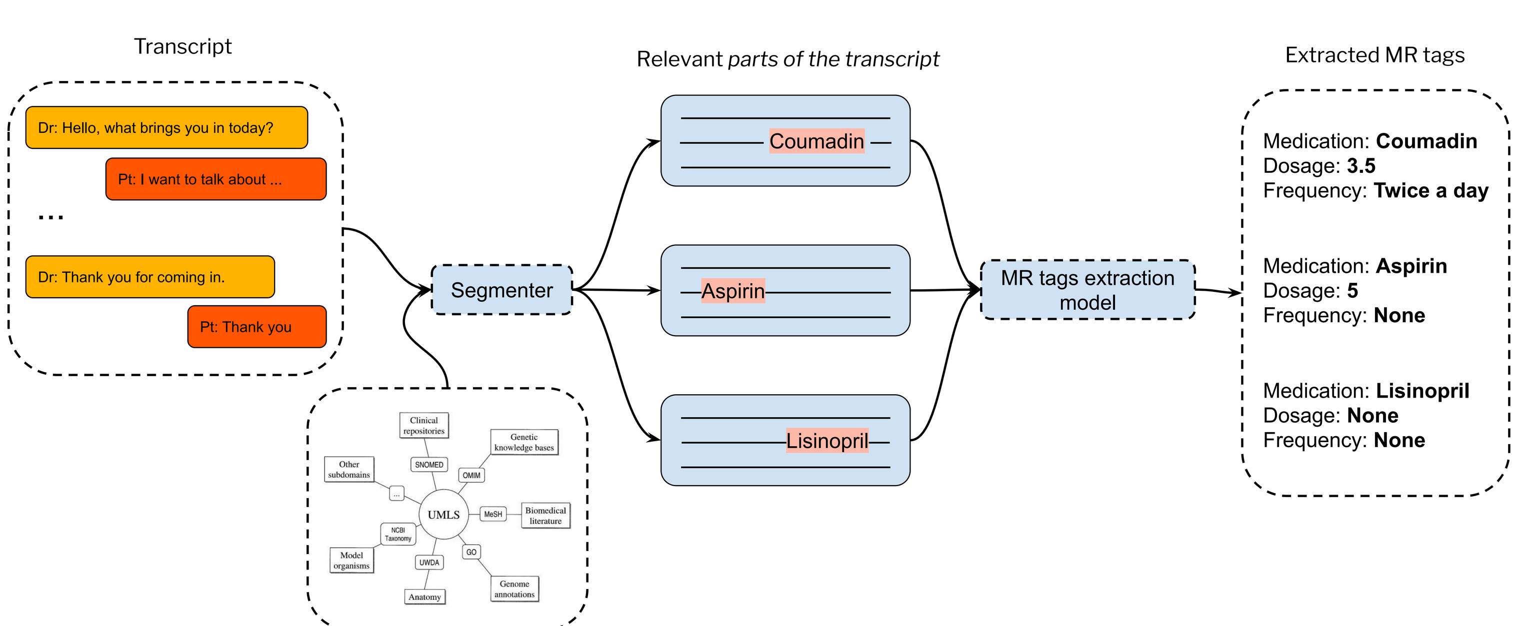


Figure 1: Overview of our medication regimen extraction pipeline

- We identify medications in a conversation using string matching with UMLS ontology [3]
- For each medication occurrence, *Part of the Transcript* is identified
- We extract the medication's *dosage* and *frequency* using a model in one of the following frameworks:
  - Question Answering framework (QA PGNet):
    - Pointer-generator networks [4]
    - Augmented with Co-Attentions [5] to attend on the questions
    - Input: "what is the dosage of Coumadin?"; *Part of the Transcript*
    - Output: 3.5
  - Question Answering + Information Extraction frameworks (MD QA PGNet):
    - Pointer-generator networks
    - With multi-decoder to output *dosage* and *frequency* separately
    - Input: *Part of the Transcript*
    - Output: Dosage decoder: 3.5; Frequency decoder: Twice a day

**For more details on the work please refer to:**

- Sai Prabhakar Pandi Selvaraj and Sandeep Konam. 2020. [Medication regimen extraction from clinical conversations](). In International Workshop on Health Intelligence (W3PHIAI 2020), New York, USA
- Dhruvesh Patel, Sandeep Konam and Sai Prabhakar Pandi Selvaraj. 2020. [Weakly Supervised Medication Regimen Extraction from Medical Conversations](). In Proceedings of the Clinical Natural Language Processing Workshop, EMNLP, 2020

## Experiments

- To address data shortage and improve the model performance, we:
  - Investigated different high-performance contextual embeddings (ELMO, BERT and ClinicalBERT)
  - Pretrained model's encoders on clinical summarization task using the medical *summaries* in our dataset
- Baselines:
  - Rule-based baselines
  - Using embeddings learned from scratch
- Input *Part of Transcript* segmented:
  - Automatically using heuristics (AS)
  - Using human annotations (HS)

## Results

- Human written transcripts:
  - ELMo with MD QA PGNet and BERT with QA PGNet performed the best
  - Pretraining on the summarization task improves the performance
- Automatic Segmentation causes a significant performance drop
  - Row 1 vs Row 2 in Table 2
- The models perform reasonably well in rare cases when the "Part of Transcript" has:
  - Multiple medication mentions
  - Multiple dosage mentions
  - Dosage mentioned before medication

| Transcripts | Models | Dosage | Frequency |
|---|---|---|---|
| | | F1 | F1 |
| HW with HS | BERT + QA PGNet (PT)<br>ELMo + MD QA PGNet (PT) | 85.69<br>85.79 | 49.25<br>43.99 |
| HW with AS | BERT + QA PGNet (PT)<br>ELMo + MD QA PGNet (PT) | 75.43<br>75.71 | 41.62<br>37.83 |
| G-STT with AS | BERT + QA PGNet (PT)<br>ELMo + MD QA PGNet (PT) | 70.51<br>71.75 | 39.90<br>40.13 |
| IBM-STT with AS | BERT + QA PGNet (PT)<br>ELMo + MD QA PGNet (PT) | 73.93<br>78.93 | 30.90<br>36.58 |

Table 2: ROUGE-1 scores of the best performing models on the ASR (Google: G-STT and IBM: IBM-STT) and human-written transcripts (HW) on the test dataset. PT: using pretrained encoder; HS: Human segmentation; AS: Auto segmentation

## Integrating with commercial Automatic Speech Recognition (ASR) systems

- We evaluate our best performing models on the ASR APIs offered by:
  - Google (G-STT) Row 3 in Table 2
  - IBM (IBM-STT) Row 4 in Table 2
- We find that the models are robust to ASR errors with only a small drop in performance
  - Row 2 vs Row 3, 4 in Table 2
- Our best model can correctly extract the *dosage* for 71.75% and *frequency* for 73.58% of the medications with Google Speech-To-Text transcripts

## Deployment and Conclusion

- We investigated and developed different frameworks for MR tags extraction from Doctor-Patient conversations
- The developed models work well in rare cases and on commercial ASR transcripts
- The pipelines presented here have been deployed to over 50,000 users via our mobile app [6]. So far, we extracted and presented cumulatively 93172 medications to users. In addition, they were also presented with definitions for 41478 medications.

## References

[1] Lisa C. Mcguire. 1996. Remembering what the doctor said: Organization and adults' memory for medical information. Experimental Aging Research, 22:403– 428.
[2] Christine Sinsky, Lacey Colligan, Ling Li, Mirela Prgomet, Sam Reynolds, Lindsey Goeders, Johanna Westbrook, Michael Tutty, and George Blike. 2016. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Annals of internal medicine, 165(11):753–760.
[3] Bodenreider, Olivier. "The unified medical language system (UMLS): integrating biomedical terminology." Nucleic acids research 32.suppl_1 (2004): D267-D270.
[4] See A, Liu PJ, Manning CD (2017) Get to the point: Summarization with pointergenerator networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 1073–1083
[5] Xiong C, Zhong V, Socher R (2016) Dynamic coattention networks for question answering. arXiv preprint arXiv:161101604
[6] **https://www.abridge.com**