MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Advanced Care Planning Content Encoding with Natural Language Processing

Ben Knoll

Research Programmer,
University of Minnesota

Introduction

- Aging Population, many feel under-prepared for difficult medical decisions.
- Advance Care Planning (ACP): Discussions and preparations for future decisions about medical care.
- ACP can help align treatment intensity with patient preferences.
- Real-world ACP participation remains low.
- Automated analysis of ACP notes may help extract information that is clinically useful.

Methods – Themes

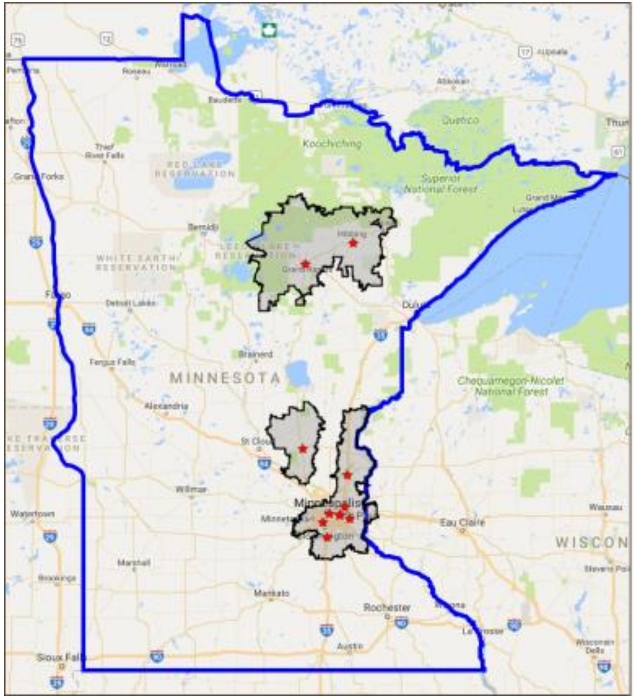
- Developed a set of 5 keys themes from an established content framework which defines core dimensions of serious illness conversations.
  Bernacki RE, Block SD, American College of Physicians High Value Care Task Force. Communication about serious illness care goals: a review and synthesis of best practices. JAMA Intern Med. 2014.

Methods – Themes

Theme	Definition
Critical Abilities, Goals, and Tradeoffs	Goals of care, patient treatment priorities, quality-of-life tradeoffs.
Decision-Making	Patient decision-making capability, decision-making preferences.
Legal Documentation	Health care directive, DNR/DNI, power of attorney, etc.
Prognosis	Health care personnel prediction of likely course of condition/disease/outcomes.
Understanding	Assessment of health literacy, patient and family understanding of diagnosis, treatment options, and prognosis.

Methods – Data

- Palliative care consult notes 2014 – 2021 in the MHealth Fairview system (12,711 notes).



Methods – Annotation

- Created annotation guidelines and an example book.
- Four human annotators.
- Each annotated a random sample of 100 notes (50 overlap).
- 250 annotated notes total (disagreement resolved by consensus).

Methods – Automated Detection

- We used publicly available pre-trained Huggingface Transformers models *bert-base-uncased and Bio_ClinicalBERT*
- Fine-tuned with manually annotated notes for sentence-level detection.
- Multi-label classification output layer.
- Increased batch size, gradient acc. steps, and learning rate.
- Examples weighted inversely proportional to support levels.

Methods – Training and Testing

- 5-fold cross validation using a 200 train - 50 test document split.
- Each fold was trained for 15 epochs and tested on the (50) hold-out notes.

Results – Annotated Notes Class Support

Theme	Quantity	Percent	Class Weight
Crit. Abil., Goals, Tradeoffs	923	1.20%	82.11
Decision-Making	434	0.57%	175.76
Legal Documentation	1,007	1.31%	75.18
Prognosis	338	0.44%	225.96
Understanding	249	0.32%	307.08
Any Theme	2,705	3.53%	27.36

Results – Agreement

- Agreement calculated at sentence-level.
- Overall inter-rater reliability (IRR) Fleiss's Kappa: 0.75.
- Per-theme IRR (Fleiss') in table to right.
- "Any Theme" is the presence of any theme in a sentence.

Theme	Fleiss's Kappa
Crit. Abil., Goals, Tradeoffs	0.58
Decision-Making	0.75
Legal Documentation	**0.91**
Prognosis	0.59
Understanding	0.52
Any Theme	0.73

MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Results – Mean Sentence-Level F1

Theme	bert-base-uncased	Bio_ClinicalBERT
Crit. Abil., Goals, Tradeoffs	.59 (*SD* = .03)	**.60 (*SD* = .05)**
Decision-Making	.72 (*SD* = .04)	**.73 (*SD* = .04)**
Legal Documentation	**.88 (*SD* = .02)**	.87 (*SD* = .03)
Prognosis	**.58 (*SD* = .06)**	.56 (*SD* = .07)
Understanding	**.50 (*SD* = .04)**	**.50 (*SD* = .07)**
Any Theme	**.80 (*SD* = .02)**	**.80 (*SD* = .01)**

Results – Mean Document-Level F1

Theme	bert-base-uncased	Bio_ClinicalBERT
Critical Abilities, Goals, and Tradeoffs	**.90 (*SD*= .04)**	.89 (*SD*= .03)
Decision-Making	**.95 (*SD*= .02)**	.94 (*SD*= .03)
Legal Documentation	**.94 (*SD*= .04)**	**.94 (*SD*= .04)**
Prognosis	**.83 (*SD*= .04)**	**.83 (*SD*= .06)**
Understanding	.73 (*SD*= .04)	**.78 (*SD*= .03)**
Any Theme	**.95 (*SD*= .02)**	**.95 (*SD*= .03)**

Results – Wider corpus Estimates

Theme	Document-level Presence
Critical Abilities, Goals, and Tradeoffs	85.8%
Decision-Making	80.8%
Legal Documentation	84.8%
Prognosis	73.1%
Understanding	66.6%
Any Theme	93.6%

Discussion

- Natural Language Processing (NLP) methods provide promise.
- Variability in performance levels, low accuracy for detecting the "Prognosis" and "Understanding" themes.
- Limited by use of only two test models using the same architecture and the relatively small size of our training and test dataset.
- First step towards more robust analysis of ACP information.

Further Work

- Expand the training data set.
- External validation.
- Leverage structured data.
- Eventually utilize to deliver patient-centered care.

Conclusions

- Can achieve decent performance in automated extraction of ACP themes.
- Shows variability in completeness of ACP discussion and documentation.
- Further development of these approaches has broad applicability for improved quality of care in key patient populations including older adults.

Acknowledgements

- Co-Authors: Melissa Gunderson, Geetanjali Rajamani, Elizabeth C. Wick, Alexis Colley, Elizabeth Lindemann, Rubina Rizvi, Molly Diethelm, Gretchen Hultman, Logan Pierce, Rui Zhang, and Genevieve B. Melton
- University of Minnesota Natural Language Processing-Information Extraction Program.
- University of Minnesota Center for Learning Health System Sciences.
- Funding provided by NIH (UL1TR000114) and AHRQ (R01HS24532).

Contact and Questions

- Me (benknoll@umn.edu)
- NLP/IE Group website: healthinformatics.umn.edu/natural-language-processing
- Group GitHub: github.com/nlpie