Using Generative AI to Extract Emphasis Frames

Washington University in St. Louis

Gechun Lin ✉ lingechun@wustl.edu

Motivation

Emphasis frames, which present the same things selectively to highlight some aspects, can be defined in relation to a specific issue (Chong and Druckman, 2007). However, two commonly used methods for identifying emphasis frames in textual data have limitations:

- *topic modeling* (DiMaggio et al., 2013):
  1. traditional ones (e.g., LDA and STM) are fully unsupervised;
  2. resulting groups of keywords (which are "topics") lack semantic contexts for exploring frames.
- *dictionary-based approach* (Hitt and Searles, 2018):
  1. existing dictionaries would miss novel frames;
  2. creating new dictionaries is labor intensive.

Solution: Generative AI

- a theory-driven data annotation tool;
- sophisticated abilities of text summarization and information extraction to synthesize high-level concepts (Lam et al., 2024).

Highlights & Findings

Generative AI	*versus*	Existing Methods
1. leverage on knowledge from extensive pretraining to extract frame features;		1. require predefined lists of frame terms or coding schemes;
2. produce more semantically interpretable frame features—phrases describing how things are framed in a specific context.		2. rely on co-occurent words or predefined terms stripped of context to identify frames;

- Re-examine Gilardi et al. (2021) using Generative AI
  1. consistent with their main conclusion: policy frames tend to be more complex as the policy diffuses;
  2. new findings: discover more nuanced frames and co-existing pattern.

Reference

Chong, D. and J. N. Druckman (2007). Framing theory. *Annu. Rev. Polit. Sci. 10*, 103–126.

DiMaggio, P., M. Nag, and D. Blei (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics 41*(6), 570–606.

Gilardi, F., C. R. Shipan, and B. Wüest (2021). Policy diffusion: The issue-definition stage. *American Journal of Political Science 65*(1), 21–35.

Hitt, M. P. and K. Searles (2018). Media coverage and public approval of the us supreme court. *Political Communication 35*(4), 566–586.

Lam, M. S., J. Teoh, J. Landay, J. Heer, and M. S. Bernstein (2024). Concept induction: Analyzing unstructured text with high-level concepts using lloom. *arXiv preprint arXiv:2404.12259*.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems 35*, 24824–24837.

Wu, T., M. Terry, and C. J. Cai (2022). Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pp. 1–22.

Ying, L., J. M. Montgomery, and B. M. Stewart (2022). Topics, concepts, and measurement: A crowdsourced procedure for validating topics as measures. *Political Analysis 30*(4), 570–89.

Identifying Frames via Generative AI

- First, leveraging *chain-of-thoughts* prompting (Wei et al., 2022), I instruct the AI assistants (LLM models) to code all information about broadly-defined categories of frames following three steps: 1) quote, 2) summarize, and 3) name;

Prompt Template

Role: You are an expert in framing analysis.
General Instruction: There are {N} generic frames—{a set of initial frames based on prior work or research questions} may be used in discussions about {description of an issue, event, or policy}.
Input: A document about the issue, event, or policy.
Output: JSON functions include three main parameters for each generic frame

Quote	Summarize	Name
relevant text for the generic frame used in the document	how quoted text is framed	a phrase to describe context-specific use

- Second, I create a *human-in-the-loop* (Wu et al., 2022) LLM pipeline to discover substantially meaningful subcategories of generic frames;

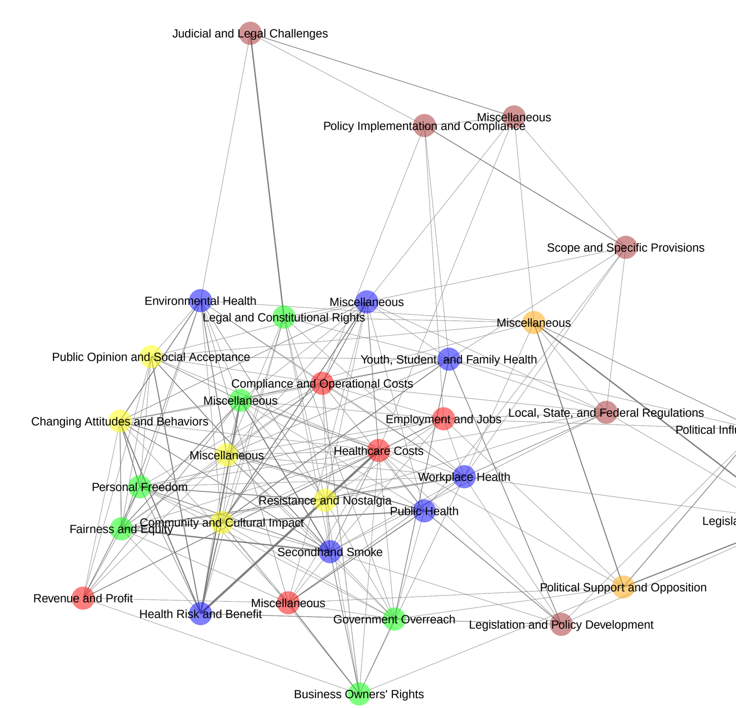
n batches of 100 random phrases
⇓
LLM: Category Generation
⇓
n lists of identified subcategories (definition & examples)
⇓
Human: Category Selection
⇓
frequently appear & substantively meaningful subcategories
⇓
LLM: Category Assignment
⇓
phrases labeled with subcategories

- Finally, I convert the analysis results to a document-frame matrix.

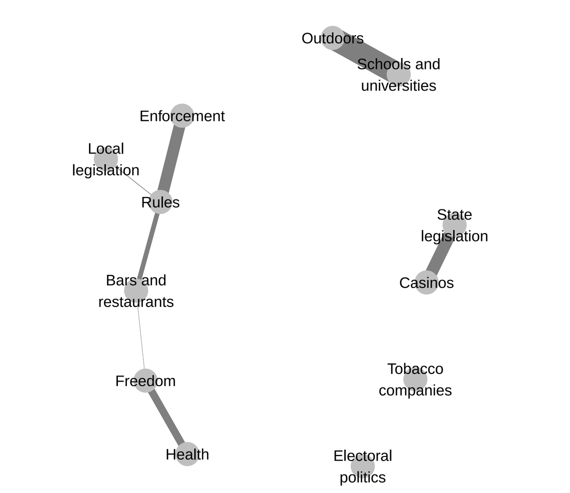
	Frame 1			Frame 2			Frame N		
doc id	*subcat1*	*subcat2*	···	*subcat1*	*subcat2*	···	*subcat1*	*subcat2*	···
1	0	1	···	1	0	···	0	0	···
2	1	1	···	0	0	···	1	0	···
3	0	0	···	1	0	···	0	1	···
···	.	.	···	.	.	···	.	.	···
···	.	.	···	.	.	···	.	.	···

Application: Reanalysis of Gilardi et al. (2021)

- Corpus: news articles covering the smoking ban policy published between 1996 and 2013 in US;
- Methods:
  - 12 topics (policy frames) by structural topic model (STM);
  - 33 subcategories of six generic frames by GPT-4o using the proposed method.
- GPT-4o adds new insights to the frame correlation network originally estimated by STM
  1. Two generic frames, **political** and **legal and regulatory**, are relatively isolated from other groups, whereas complex frames usually concentrate on some subcategories of the other four groups:
     health · economic impact · rights and freedom · social norms
  2. Some subcategories are more strongly correlated (thicker edges), e.g. "Judicial and Legal Challenges" and "Legal and Constitutional Rights", "Health Risk and Benefit" and "Healthcare Costs", suggesting that they are more likely to appear together within news article.



(a) GPT-4o



(b) STM

Validate, Validate, Validate

- Topic validation tasks adapted from Ying et al. (2022):
  multiple-choice questions with four options and one correct answer
  1. R4WSI: choose the word/phrase set below that is most UNRELATED to the other three
  2. OLW: choose the OPTIMAL label for the given word/phrase set
  3. LI: choose the label that is most UNRELATED to the given news article

- Pilot results from MTurk:
  (note: each model has two identical trials per task.)
  1. GPT-4o outperforms STM in R4WSI and OLW, suggesting GPT-4o provides more *distinctive and coherent* features that help identify different frames;
  2. GPT-4o and STM have similarly bad performance in LI, need to improve task design and add attention check.