EN.625.692.81.SP22 PROBABILISTIC GRAPHICAL MODELS PROJECT TOPIC PROPOSAL

NICHOLAS LINES

1. Proposed Topic: Reporting Trust

One field of research rich with applications of Probabilistic Graphical Models is knowledge management and fusion. Knowledge graphs are used in many influential and meritorious ways, from improving Google searches to predicting opportunities for medical advancement [1]. One corner of this field of particular interest to me is that region dealing with trust and certainty problems within knowledge systems. For my project I'd like to propose an investigation into modeling a system featuring observers whose trustworthiness varies by observer and also by topic.

2. A Suggested Toy Setting and Relevant Problems

2.1. **Introduction.** While the details of what I investigate will no doubt change throughout the next few weeks as my understanding of the tools and theory readily available grows, here is a first-pass at a simple application and the relevant problems it inspires.

In this setting, we imagine a newspaper reporter B who is required to stay abreast of current events in each of $N_E \in \mathbb{N}$ news topics. To aid in this, the reporter has established $N_O \in \mathbb{N}$ distinct sources that observe some subset of world events and pass on their opinions to the reporter. The reporter has no independent verification system outside of this set of observers $O = \{O_1, O_2, \ldots, O_{N_O}\}$. The overall goal of the reporter is to understand the world events as well as possible using the available sources. The first task toward this end will be for the reporter to identify consensus in each topic and decide on a set of sources that are most trustworthy regarding that topic. The second task will be to introduce $N_n \in \mathbb{N}$ new sources of unknown trustworthiness, and determine their trustworthiness and assign them to informative groups as appropriate.

2.2. **Details.** We'll now begin to provide assumptions that will limit the scope of the problem to enable us to perform meaningful experiments. Let $N_E = 3$, that is, we limit ourselves to three topic sets, specifically P for Politics, E for Entertainment, and T for Technology. We will assume hard boundaries between these topics, i.e. knowledge about Politics will be independent of knowledge about Entertainment, etc. Moving forward from instantiation, we will assign to each topic set a new value $P_t \sim D_P, E_t \sim D_E, T_t \sim D_T$ for $t = 1, 2, 3...$, sampled from $\{0, 1\}$ according to a binomial probability distribution unique to each topic. This value represents a fact related to that topic area at time t. Note that the value held by each topic variable is binary, and each value is independent from those held at other time steps and by other variables within the same time step.

Date: February 14, 2022.

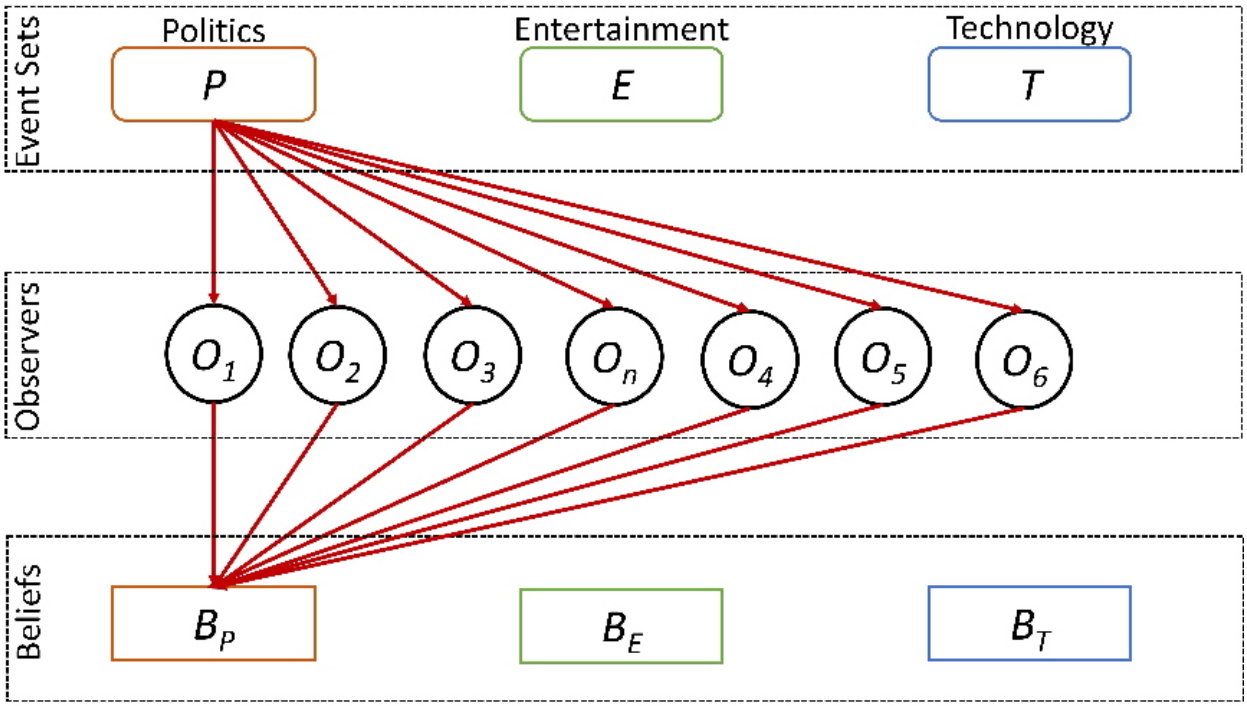

FIGURE 1. An illustration of the trust network described in the Details subsection. The directed edges from the Politics variable to the observers and from the observers to the Reporter's current belief about the Political variable are shown. In the full model, directed edges like these flow from each of the Event Sets to each Observer and from each Observer to each Belief.

Let's discuss the observers. Let $N_O = 6$. We define three Dirichlet distributions

$$\begin{aligned} D_{O,P} &= Dir\left([0.8, 0.1, 0.1]\right) \\ D_{O,E} &= Dir\left([0.1, 0.8, 0.1]\right) \\ D_{O,T} &= Dir\left([0.1, 0.1, 0.8]\right) \end{aligned}$$

so that each is biased towards a single topic, and sample from these the multinomial parameters for each observer node

$$\begin{aligned} p_{O_1}, p_{O_2} &\sim D_{O,P} \\ p_{O_3}, p_{O_4} &\sim D_{O,E} = Mult\left(n_{O,E}, p_{O,E}\right) \\ p_{O_5}, p_{O_6} &\sim D_{O,T}. = Mult\left(n_{O,T}, p_{O,T}\right). \end{aligned}$$

For each time step t and each observer node $i = 1, 2, ..., 6$, we draw a sample

$$O_{i,t} \sim Mult\left(n_{O_i}, p_{O_i}\right)$$

which will serve tell us how the observer will pass on their observations to the reporter. We'll start with $n_{O_i} = 5$ for all nodes, and crank this up as needed to make learning faster. Each observer takes note of the three event values and

retransmits them to the reporter according to the following system:

For each topic $j = 1, 2, 3$:
If $O_{i,t}[j] > 0$, pass on the true event value
Otherwise, pass on 1 or 0 with equal probability.

Task 1 is to learn the trustworthiness of each node O_i with respect to each topic by forming clusters that agree by consensus.

For Task 2 after Task 1 is complete at time step t_k we introduce a new observer O_n whose bias has been sampled randomly according to one of the three strategies described above. The task will be to learn the bias of O_n in as few time steps as possible.

2.3. **Related work.** Some elements of this proposal were inspired by [2]. I also referred heavily to the course text, [1].

References

1. Daphne Koller and Nir Friedman, *Probabilistic graphical models: principles and techniques*, MIT press, 2009.
2. Georgiy Levchuk and Erik Blasch, *Probabilistic graphical models for multi-source fusion from text sources*, 2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), IEEE, 2015, pp. 1–10.
3. Larry Wasserman, *All of statistics: a concise course in statistical inference*, Springer Science & Business Media, 2013.

E-mail address: nicholasalines@gmail.com