Topic Modeling for Maternal Health Using Reddit

Shuang Gao
New York University
sg5963@nyu.edu

Shivani Pandya
John Hopkins University
spandya6@jhu.edu

Smisha Agarwal
John Hopkins University
sagarw23@jhu.edu

João Sedoc
New York University
jsedoc@stern.nyu.edu

Abstract

This paper applies topic modeling to understand maternal health topics, concerns, and questions expressed in online communities on social networking sites. We examine latent Dirichlet analysis (LDA) and two state-of-the-art methods: neural topic model with knowledge distillation (KD) and Embedded Topic Model (ETM) on maternal health texts collected from Reddit. The models are evaluated on topic quality and topic inference, using both auto-evaluation metrics and human assessment. We analyze a disconnect between automatic metrics and human evaluations. While LDA performs the best overall with the auto-evaluation metrics NPMI and Coherence, Neural Topic Model with Knowledge Distillation is favourable by expert evaluation. We also create a new partially expert annotated gold-standard maternal health topic modeling dataset for future research.[1]

1 Introduction

Evidence suggests that poor quality maternal and newborn care is responsible for nearly 60% of the estimated 5 million deaths each year globally (Kruk et al., 2018). This situation spouses urgent demands of artificial intelligence. AI has potential applications in automating aspects of service delivery such as basic counselling, thereby reducing the burden on health systems. In this work, we hope to help improving healthcare counselling AI by understanding maternal health related content. What are people concerning about regarding maternal health? And among state-of-the-art techniques, which model is the best at extracting and interpreting those highly-professional topics?

With the development of the online communities and social media, advanced text mining techniques are widely needed when applied to domains. It is necessary to understand a question in order to answer it, and the first step towards understanding the content of a document is to determine which topics that document addresses. Therefore, many successful conversational agents employ topic models to encourage more coherent dialogue (Baheti et al., 2018; Vlasov et al., 2019).

Topic models can provide legible and concise representations of both the entire corpus and individual documents because they represent topics as ranked word lists and documents regarding their probable topics. Therefore, while mainly applied to mining the text content, topic models can also classify the topics for new documents by inferring the latent topic distribution. In classical topic models like Latent Dirichlet Allocation (LDA), distributions over the latent variables are estimated with variational inference algorithms (EM) or Gibbs sampling (Blei et al., 2003). However, recently, with the development of deep learning methods, now there are neural topic models that incorporate additional information and leverage the variational autoencoder (VAE) framework for latent variable inference. Particularly, pre-trained transformer-based language models (e.g., BERT, (Devlin et al., 2019)) is employed to fine-tune on a wide variety of NLP problems; some models combine the advantages of pre-trained transformers and topic models.

This work contributes the following:

- Provide an exemplary application of the three topic models (LDA, Neural topic model with knowledge distillation, and Embedding Topic Model) on texts in the maternity health domain.

- Evaluate the model performances from both quantitative and qualitative perspectives. The assessment may inspire ideas in metric design and model improvement.

---

[1] https://github.com/jsedoc/maternal_health_topics

- Provide a new partially annotated dataset in the maternal healthcare domain for topic modelling.

2 Related Work and Background

There are a multitude of state-of-the-art topic models adapted from the LDA model. Neural topic models (Srivastava and Sutton, 2017) surpass traditional methods in architecture by using various forms of neural networks and can be applied to either labeled or unlabeled data. Hoyle et al. (2020) further improves neural topic models by knowledge distillation (KD). Another thread of innovations improve models in terms of semantic meaning of words, for example, Dieng et al. (2020) incorporates word embedding to topic models. Given that our dataset is unlabeled, this work focus on two state-of-the-art models, the KD model and the Embedding Topics Model (ETM), and compares their performances to LDA. We briefly summarize the details of LDA, KD and ETM in this section.

2.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. The topic probabilities provide an explicit representation of a document. LDA follows the generative process as below:

1. For each document d, draw the topic proportion θ_d from $Dir(\alpha)$;

2. For each word w_n in document d:

    (a) Choose a topic $z \sim Multinomial(\theta)$;

    (b) Choose a word w_n from p($w_n|z_n$, β), a multinomial probability conditioned on the topic z_n.

In our instance, we adopted MALLET (McCallum, 2002), a machine learning toolkit which implements LDA and conducts variational inference by Gibbs Sampling.

2.2 Neural Topic Model with Knowledge Distillation

Hoyle et al. (2020) combine neural topic modeling with knowledge distillation. It generates and stores the teacher logits (latent topics) $z^{teacher}$ for each document in the training set using DISTIL-BERT (Sanh et al., 2019) as a deterministic auto-encoder and training on document reconstruction as a decoder. Then it entails the framework of the neural topic model SCHOLAR but substitutes the reconstruction loss L_R with L_{KD}:

$$L_R = (w_d^{BoW})^T \log f(\theta_d, B)$$
$$L_{KD} = \lambda T^2 (w_d^{teacher})^T \log f(\theta_d, B; T) + (1 - \lambda)L_R$$
$$f(\theta_d, B) = \sigma(m + \theta_d^T B)$$

where $\sigma(\cdot)$ is the softmax function, $w_d^{teacher}$ is the result of teacher logits $z^{teacher}$ scaled by softmax temperature T, and $f(\cdot; T)$ is the scaled version of $f(\cdot)$. In doing so, the weights were updated to minimize the original reconstruction loss of SCHOLAR and the loss of reconstruction from teacher logits, which are pretrained and deterministic, so the knowledge contained in pretrained transformers is distilled to this student model. In this instance, we adopted the implementation of Hoyle et al. (2020).

2.3 Embedding Topics Model

Embedding Topics Model (ETM; Dieng et al., 2020) is also a generative probabilistic model that combines the useful properties of topic models and word embeddings. As a topic model, it discovers an interpretable latent semantic structure of the texts; as a word embedding, it provides a low-dimensional representation of the meaning of words. In contrast to LDA, the per-topic conditional probability of a term has a log-linear form that involves a low-dimensional representation of the vocabulary. Each term is represented by an embedding, and each topic is a point in that embedding space; the topic's distribution over terms is proportional to the exponentiated inner product of the topic's embedding and each term's embedding. ETM follows the generative process as below:

1. For each document d, draw the topic proportion θ_d from $LN(\alpha)$;
2. For each word w_n in document d:

    (a) Choose a topic $z \sim Multinomial(\theta)$;

    (b) Choose w_n from $softmax(\rho^T \alpha_z)$, a multinomial probability computed by the inner product of word embedding and the embedding of topic z.

In this instance, we adopted the implementation of Dieng et al. (2020).

3 Experimental Details

3.1 Data

3.1.1 Collection

For our project, we deploy data from the popular site Reddit spanning between December 2005 and April 2019 because of the abundance of data in almost any given topic due to the site's specialized "subreddits". The public data on Reddit is accessible, which provides convenience to researchers, and the specialized channel create an online community about maternal healthcare, making it easier to collect and filter a domain corpus. To make the data set more manageable and remove unwanted data, we eliminated all subreddits not in a pre-compiled list of relevant subreddits chosen by a healthcare professional. Finally we end up with 24k comments for training and 6k comments as hold-out samples. Table 1 is a statistical description of the corpus, and we release this maternal health corpus on the github.

# documents	30647
Avg length per doc	81.51
# Vocabulary (over 3 counts)	11537
# Vocabulary (over 100 counts)	1880
# Vocabulary (over 1000 counts)	336
# Vocabulary (over 10000 counts)	37

Table 1: Statistics description of the collected corpus

3.1.2 Preprocessing

Before applying the models on this dataset, we preprocess the data for each model to prepare valid tokens or word embeddings. For LDA and the neural topic model with knowledge distillation (KD), we preprocess the data by removing common stop words in MALLET stoplist. LDA uses MALLET's built-in tokenizer, and the other two models tokenize the text by single words and convert the word lists to bag of words. Except for the KD model which requires pretraining with the teacher model, ETM also requires pretraining the word embedding on the training documents.

3.2 Metrics

3.2.1 Coherence

Mimno et al. (2011) proposed coherence to measure the quality of topics. Coherence measures topic word co-occurrence across documents to detect low quality topics, and show it correlates with

expert topic annotations. The Coherence of topic t is defined as

$$coherence(t) := \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \log \frac{D(v_i^t, v_j^t)}{D(v_i^t)},$$

$D(x)$ represents the number of documents word x appears in, and $D(x, y)$ represents the number of documents x and y co-appear in. Therefore, if the words in the same topic always appear in the same documents, this topic is coherent, and the score approaches zero. Otherwise, if the words always appear in different documents, the coherence score approaches negative infinity.

3.2.2 NPMI

First introduced by Bouma (2009), Normalized Pointwise Mutual Information (NPMI) is adapted from PMI. Newman et al. (2010) shows that the auto-evaluation of topic-semantic coherence using PMI is highly correlated with human evaluation, and NPMI has been widely used as a quantitative measurement of topic quality (Aletras and Stevenson, 2013; Hoyle et al., 2020). The NPMI of topic t is defined as

$$NPMI(t) := \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\log \frac{P(v_i^t, v_j^t)}{P(v_i^t)P(v_j^t)}}{-\log P(v_i^t, v_j^t)}.$$

3.2.3 Topic Uniqueness

We also use Topic Uniqueness (TU) as a quantitative measurement. As the coherence and NPMI measure how well the top words within a topic share similar context, TU measures to what extend the different topics overlap. The TU for topic t is defined as

$$TU(t) := \frac{1}{N} \sum_{i=1}^{N} \frac{1}{count(v_i^t)}$$

where $count(v_i^t)$ is the total number of times the word v_i^t appears in top N words across all topics. If $TU(t)$ equals to 1, then all top words of topic t don't appear in top words of any other topics.

4 Experimental Results

4.1 Automatic Evaluation

We train each model under 10, 25, 50, 100, and 200 topic settings, and compare them with respect to the three metrics. Table 6, Table 7 and Table 8 in Appendix A show the performances of different models in terms of Coherence, NPMI and Topic Uniqueness respectively.
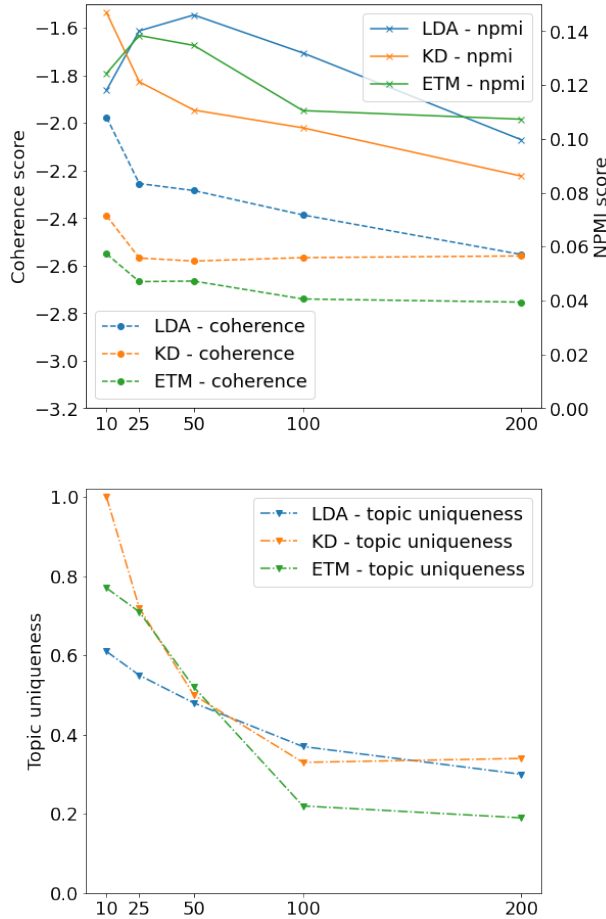
Figure 1: Performances of different models. The top plot shows the coherence scores and the NPMI scores, and the bottom plot shows the topic uniqueness.

4.2 Human Assessment

Two experts in public health identified topics generated by the 10-topic and 25-topic models. By reading through the top words for each topic, one expert annotates effective words and summarize the topic, and the other checks the annotation and topic names, and makes complement. They make an agreement on the final assessment.

ETM dismisses more topics than the other two models. When looking for 10 topics with each model, there are a total of 10 topics identified from all topics generated by the three models, and only 5 of them are identified in all of the three models: "Pregnancy", "Abortion", "Vaginal health", "Menstruation", "Labor and delivery". Besides that, only the KD model generated topics which are identified as "Penile health", "Breastfeeding" and "pregnancy symptoms". ETM also generated the least identifiable topics when looking for 25 topics. Out of the 13 topics identified, ETM dismisses "Sex", "Maternity and paternity leave" and "Stis", and it had 4

topics that cannot be identified as maternity-related.

Moreover, the KD model has the most representative top words. The expert was asked to annotate related words among the 20 top words within each topic, and the KD model achieves 17.8 and 15.7 words on average for the 10-topic and 25-topic models respectively, much more than LDA and ETM, as shown in Table 4

4.3 Inference

To infer the topics for the new documents held-out from the training data, we sampled 10 comments from the dev set and inferred their topics one by one. According to the performance reported above, we selected LDA 25 and KD 25 to make the inference. Among the 10 comments displayed in Appendix B, KD had better topic classification. LDA and KD misclassified comment 1 as "penile health" and "vaginal health" respectively, while the comment is related to "pregnancy". Comment 6 was also misclassified by both. Comment 8 was misclassified as "pregnancy" by LDA but classified as "abortion" correctly by KD. Comment 9 was classified as "pregnancy" by LDA, which was correct, but KD more accurately classification as "pregnancy risk". All other comments are correctly classified by both models.

5 Discussion

Measurement Incompatibility There is an incompatibility between human assessment and the auto-evaluation metrics for topic quality. In terms of the auto-evaluation metric Coherence, LDA always performs the best, which is inconsistent with the human assessment. In terms of NPMI, KD dominates in the 10-topic setting, but LDA still performs better in the 25-topic, 50-topic, and 100-topic settings. In terms of Topic Uniqueness, KD is the best in 10-topic, 25-topic setting, which is the same as human evaluation.

Topic Repetitiveness Although fixed values are set for the number of topics when training on the corpus, fewer topics could be identified from the generated word lists because of the overlapping among the lists. For example, when generating 10 topics, KD has 9 topics identified by the expert, while topic 3 and topic 5 are categorized as "vaginal health". For ETM, there are more overlapped generated lists: topic "pregnancy", "vaginal health", "menstruation", and "labor and delivery" all include 2 generated lists.

Method	Topic Words
LDA	**life abortion child** would **fetus** think woman right human people person mother **choice** body **baby** pro women want **abortions** one
KD	**murder alive** human **fetus personhood rights zygote** argument **fetuses womb** begins **killing survive pro clump embryo viable debate development** life
ETM	**abortion life child fetus** people woman human women point person body make mother **abortions baby** medical **support** agree **choice** circumcision

Table 2: Topic words associated with the topic labelled as abortion. Bolded words are marked as topic relevant.

	LDA	KD	ETM
n/a	1 / 1	- / 1	- / 4
pregnancy	1 / 5	1 / 3	2 / 4
birth control	1 / 3	1 / 5	- / 1
abortion	1 / 2	1 / 2	1 / 3
menstruation	2 / 3	1 / 2	2 / 3
vaginal health	1 / 2	2 / 3	2 / 3
penile health	- / 1	1 / 1	- / 1
sex	1 / 1	- / 1	1 / -
breastfeeding	- / 1	1 / 1	- / 1
pregnancy symptoms	- / 1	1 / -	- / 2
labor and delivery	2 / 2	1 / 3	2 / 2
pregnancy risks	- / 1	- / 1	- / 1
maternity leave	- / 1	- / 1	- / -
stis	- / 1	- / 1	- / -

Table 3: Identified topics counts for 10-topics and 25-topic models (separated by '/')

	LDA	KD	ETM
10-topic	6.7	**17.8**	8.7
25-topic	9.5	**15.7**	10.6

Table 4: Average number of effective top words on topics

	LDA	KD	ETM
10-topic	0.1102	0.0132	0.0558
25-topic	0.0881	0.0280	0.0441
50-topic	0.0805	0.0275	0.0424
100-topic	0.0747	0.0250	0.0407
200-topic	0.0662	0.0240	0.0432

Table 5: Average word frequencies for topics generated. The frequency of a word w is represented with the proportion of documents which contain the word w.

recommend applying the neural topic model with knowledge distillation, because it tends to rank professional words that are unique to one topic as top words, and it covers more diverse topics when asked to generate a fixed number of topics. Regarding topic inference on new documents, both LDA and the knowledge-distilled model perform well.

Word Frequency of Top Words As shown in Tables 2 and 5, LDA and ETM include more frequently appeared words such as "feel", "period" and "life" in top word lists, the top words of each topic generated by KD is professional and of low-frequency.

6 Conclusion

We applied and compared three different topic models on the texts related to maternal health and discovered topics with high-quality. By both auto-evaluation and human assessment, we evaluated the topic quality generated by different models and observed incompatibility among metrics. Taking the human assessment as a gold standard, we would

References

Nikolaos Aletras and Mark Stevenson. 2013. Evaluating topic coherence using distributional semantics. *In Proceedings of the 10th International Conference on Computational Semantics*, pages 13–22.

Ashutosh Baheti, Alan Ritter, Jiwei Li, and Bill Dolan. 2018. Generating more interesting responses in neural conversation models with distributional constraints. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3970–3980, Brussels, Belgium. Association for Computational Linguistics.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *In Proceedings of the Biennial GSCL Conference*, 1:31–40.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the*

North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1:4171–4186.

Adji B. Dieng, Francisco J.R. Ruiz, and David M. Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 3:439–453.

Alexander Hoyle, Pranav Goel, and Philip Resnik. 2020. Improving neural topic models using knowledge distillation. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 1752–1771.

Margaret E. Kruk, Anna D. Gage, Catherine Arsenault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, Pierre Barker, Bernadette Daelmans, Svetlana V. Doubova, Mike English, Ezequiel García Elorrio, Frederico Guanais, Oye Gureje, Lisa R. Hirschhorn, Lixin Jiang, Edward Kelley, Ephrem Tekle Lemango, Jerker Liljestrand, Address Malata, Tanya Marchant, Malebona Precious Matsoso, John G. Meara, Manoj Mohanan, Youssoupha Ndiaye, Ole F. Norheim, K. Srinath Reddy, Alexander K. Rowe, Joshua A. Salomon, Gagan Thapa, Nana A. Y. Twum-Danso, and Muhammad Pate. 2018. High-quality health systems in the sustainable development goals era: time for a revolution. *The Lancet Global Health*, 6(11):e1196–e1252.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. *Http://mallet.cs.umass.edu*.

David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. *In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 262–272.

David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. *In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, pages 100–108.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*.

Vladimir Vlasov, Johannes E.M. Mosig, and Alan Nichol. 2019. Dialogue transformers. *arXiv preprint arXiv:1910.00486*.

	LDA	KD	ETM
10-topic	**-1.9756**	-2.3899	-2.5505
25-topic	**-2.2545**	-2.5681	-2.6672
50-topic	**-2.2841**	-2.5807	-2.6649
100-topic	**-2.3867**	-2.5663	-2.7402
200-topic	**-2.5528**	-2.5592	-2.7535

Table 6: Coherence for topics generated

	LDA	KD	ETM
10-topic	0.1181	**0.1469**	0.1243
25-topic	**0.1401**	0.1212	0.1384
50-topic	**0.1460**	0.1107	0.1347
100-topic	**0.13190**	0.1041	0.1105
200-topic	0.0997	0.0862	**0.1073**

Table 7: NPMI for topics generated

A Appendix: Experiment Results

The experiment results are shown in Table 6, Table 7 and Table 8

We trained the LDA model with 4 CPU cores and trained the other two model with a single GPU core. Table 9 shows that LDA implemented by Mallet is quite faster than the other two models, and KD is the most time-consuming one because of the pre-training of teacher model. But once the logits of teacher model has been saved, it is fast to train the student models with various number of topics.

B Appendix: Examples for Inference

Comment 1 I was waiting to see if someone brought this up! Alcohol and tobacco are not the same. Moderate alcohol use (one drink a day or less) hasn't been shown to be harmful to a pregnancy. ANY tobacco use is harmful to a pregnancy. I would feel much more comfortable refusing to serve hooka to a pregnant woman that refusing to serve alcohol. It's actually been shown that moderate alcohol consumption can be beneficial (-drinking-pregnant-women.html)!

	LDA	KD	ETM
10-topic	0.61	**1.00**	0.77
25-topic	0.55	**0.72**	0.71
50-topic	0.48	0.50	**0.52**
100-topic	**0.37**	0.33	0.34
200-topic	**0.30**	0.22	0.19

Table 8: Topic Uniqueness for topics generated

	LDA	KD	ETM
pre-train		teacher model	embedding
		6409 s	10 s
10-topic	111 s	364 s	1112 s
25-topic	131 s	186 s	1121 s
50-topic	148 s	236 s	1110 s
100-topic	167 s	340 s	1062 s
200-topic	204 s	547 s	1086 s

Table 9: Training time for different models

(LDA - penile health, KD - vaginal health)

Comment 2 So will you be back for a follow up pap in a few months or getting the colposcopy right away? A lot of these clear up on their own. Try not to panic!
(LDA - vaginal health, KD - vaginal health)

Comment 3 I was on Loestrin24Fe for about a year but at my February yearly gyno appointment, I asked my doctor about a pill to give me less periods because mine, like yours, were *brutal.* He switched me to Lo Loestrin Fe and I haven't had a period since! It's AWESOME. I've also noticed that I am not as emotional as I was on L24Fe. Don't get me wrong - For about the first month, I was super sensitive to everything anyone said to me. But since then, I've been fine. My sex drive is sky high as always - but I don't think ANYTHING could kill my libido! Haha. I have not gained any weight since starting this pill. Hope this helps!
(LDA - birth control, KD - birth control)

Comment 4 Personally, I don't remember distinctly having that feeling (I had an epidural), but I obviously would have no idea if I would have otherwise. I think it was mostly just because I was way too exhausted to feel anything but relief (over 30 hours of labor, and no sleep), also if anything, I think the pitocin interfered with it since that's synthetic oxytocin. However, I totally felt an instant bond with my son and would not put him down those first few hours, I didn't even go to sleep.
(LDA - labor and delivery, KD - labor and delivery)

Comment 5 The only reason I ever heard about not being able to eat is in case of an emergency c-section where you are put under general anesthesia. I guess it depends on where you are and what they do, but general anesthesia is not the common practice for emergency c-section anymore. If you were already having an epidural they would work

with that, or they would give you a spinal in an emergency. That way you are still awake for the operation and there is minimal risk. When getting an epidural, a common side effect is a drop in blood pressure so an IV of fluids is introduced to offset this. That would be one reason to have IV fluids during labour. I plan on eating and drinking during my labour because, like you,I think how the hell are you supposed to be able to push out a baby when you have been starved for a day! Thankfully my midwives are of the same mind. So I will be packing snacks and a few sandwiches to bring with me for sure.
(LDA - labor and delivery, KD - labor and delivery)

Comment 6 Colon cancer can be *prevented* by regular colonoscopies. They remove precancerous polyps while they're up there so you won't ever develop it in the first place. It honestly has to be one of the most preventable cancers, and too many people know nothing about it. If you're over 40, get the pooper scope! It's a good decision.
(LDA - vaginal health, KD - vaginal health)

Comment 7 They make some very small models (check a comparison chart for the actual sizes), and keep in mind that cups are easier to insert than tampons for a lot of people (like me), because tampons are dry. It's pretty easy to moisten cups with a bit of water (or lube, even), since you're in the bathroom anyway. [Here's the latest comparison chart I could find.] For me, they've been a lifesafer since I can't use tampons, pads and my sensitive skin don't get along, and because I have a very heavy flow. They're not for everyone, but they're definitely the most comfortable alternative I've tried.
(LDA - menstruation, KD - menstruation)

Comment 8 Well, kinda yeah. there are other ways to keep yourself from having a baby besides abortions, and tbh, abstinence isn't the end of the world, lol. It isn't fair of anyone to think its okay to destroy the potential of an unborn child because they simply don't want to deal with the consequences of a certain lifestyle. Change your lifestyle, or use better protection, or have the child anyway, but don't just write off the life of a fetus like it is worth nothing, because it isn't.
(LDA - pregnancy, KD - abortion)

Comment 9 My cycles were off, so I had a decent wait. I went to the health dept to get confirmation and health insurance, then scheduled a drs

appointment which was a week or two away, the doctor guessed I was about 6-7 weeks, though my lmp said I was 14 weeks because he couldn't find a heartbeat with Doppler and he was feeling for my uterus. Then I was called the next day with a date for an ultrasound about two weeks away and was told to go back to get the rest of my blood drawn, as they only took one vial the day before. When I should have been 16 weeks, I was 12 weeks, as it showed on the ultrasound. Ultimately, you may as well wait. Waiting is something you deal with all of pregnancy, and your wait is really short, so if anything, I'd wait for the appointment and express your angst to find out how far along she is in hopes of getting things moving.
(LDA - pregnancy, KD - pregnancy risks)

Comment 10 I've had my Mirena for nearly 5 years (getting it replaced this summer) and I'm one of the lucky ones where insertion didn't hurt at all and compared to when I was on the pill, my sex drive went waaaaay up. Not sure if it returned to "normal" or not since I had been on the pill for years and years before I became sexually active so I have nothing to compare it to, but my boyfriend is a lot happier now :p I also haven't had a "real" period in about 4 years, as compared to every two to three weeks even on the pill. Needless to say, I'm a huge personal Mirena fan :)
(LDA - birth control, KD - birth control)