Towards Semantic Search for Community Question Answering for Mortgage Officers

Amir R. Rahmani, Linwei Li, Brian Vanover, Shourabh Rawat, Colin Bertrand
Zillow Group

KDD Document Intelligence Workshop, August 15, 2021

Introduction

- Mortgage industry is dynamic and complex
- A Community Question Answering (CQA) system for mortgage officers can
  - facilitate interactive, organized, and dynamic knowledge sharing
  - help ensure the consistency of the services offered
  - on-board and train new members
- We built a CQA system for mortgage officers at Zillow. It consists of 2 major parts:
  1. a question-answering platform which allows the users to post questions, answer their peers' questions, up-vote and approve answers, and track their peers activities
  2. a state of the art search engine that can search public mortgage resources as well as internal contents that our mortgage team produces using the QA platform (**subject of this paper**)
- Requirements on the search engine
  - High relevancy results (top 5 retrieved results should contain the most relevant information)
  - Users should be able to search via natural language queries as well as keywords
  - Relatively low latency

Zillow

Datasets

1. Mortgage database
   - ~6000 FAQ-answer pairs (documents) from major mortgage organizations
   - Queries will be searched against this database
   - Example: **FAQ**: What is the difference between a modular home and a manufactured home?

     Answer: The main difference between manufactured and modular homes is that manufactured homes are built to the national HUD code, while modular homes are built to all applicable state and local building codes...

2. Human labeled search evaluation dataset
   - 16 queries that loan officers usually search for
   - ~110 labeled results (FAQ-answer pair) for each query
   - 0 (irrelevant), 1 (somewhat relevant), 2 (relevant)
   - Example query:
     - What's the maximum DTI (Debt To Income) ratio for a conventional loan?

Solution Formulation

- Observations
  - The question part of an FAQ-answer pair is often a good summary of the answer part
  - FAQs are often grammatically complete sentences
  - Similar concepts can be expressed in various lexical forms, e.g. loan and mortgage

- Proposed approach
  - Hybrid search engine to retrieve semantic and lexical information
  - Keyword search using TF-IDF and BM25
  - Semantic search using Sentence BERT[1] (SBERT)

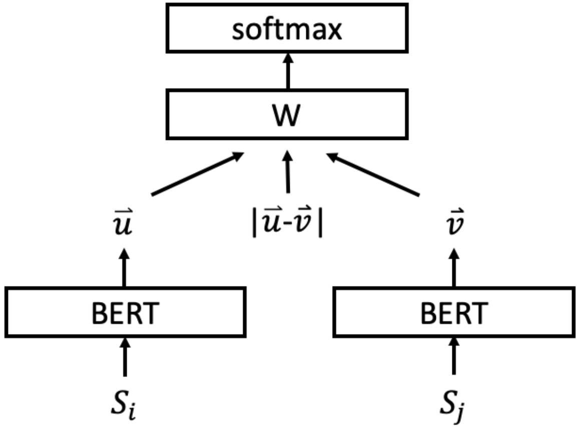
[1] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.

Zillow

SBERT Training and Finetuning

1. Finetune BERT[1] on the mortgage dataset
    - Task: Masked LM

2. Train SBERT on the NLI[2] dataset
    - Target: 3 classes, contradiction, entailment, neutral
    - 3-way softmax classifier with cross entropy loss
    - Mean pooling



SBERT architecture

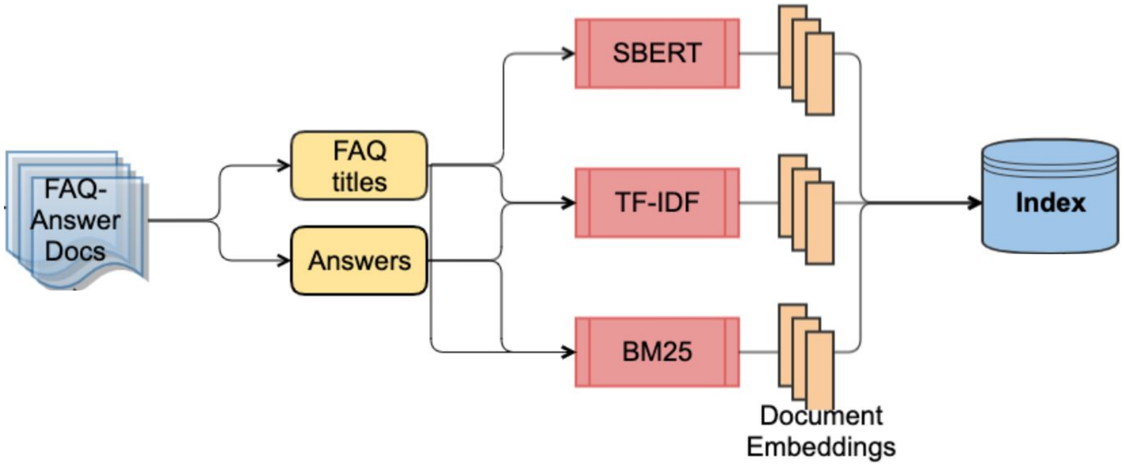
[1] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

[2] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference.

Indexing

1. SBERT embedding only for the FAQ title

2. TF-IDF and BM25 indices for both the FAQ title and the answer/body



Indexing Process

Retrieving

1. The initial ranking score is a weighted average of SBERT and tf-idf similarity. The weight is controlled by a dampening factor.
   Note: Q is the query, (q,a) is the FAQ-answer pair to be searched against.

$$\begin{aligned} F_{\text{initial}}(Q,(q,a)) &= [\alpha + (1-\alpha) * \zeta] * F_{\text{sbert}}(Q,(q,a)) \\ &+ [(1-\alpha) * (1-\zeta)] * F_{\text{tfidf}}(Q,(q,a)) \end{aligned} \quad (1)$$

dampened weight

2. Cosine similarity used. For tf-idf, another weight is set to merge the FAQ title and answer.

$$\begin{aligned} F_{\text{sbert}}(Q,(q,a)) &= \cos(u_Q, u_{FAQ}) \\ F_{\text{tfidf}}(Q,(q,a)) &= w * \cos(e_Q, e^q_{FAQ}) + (1-w) * \cos(e_Q, e^a_{FAQ}) \end{aligned} \quad (2)$$

3. Dampening factor. Favors tf-idf with longer queries.

$$\zeta = exp(\frac{1 - len(Q)}{\beta}) \quad (3)$$

Zillow

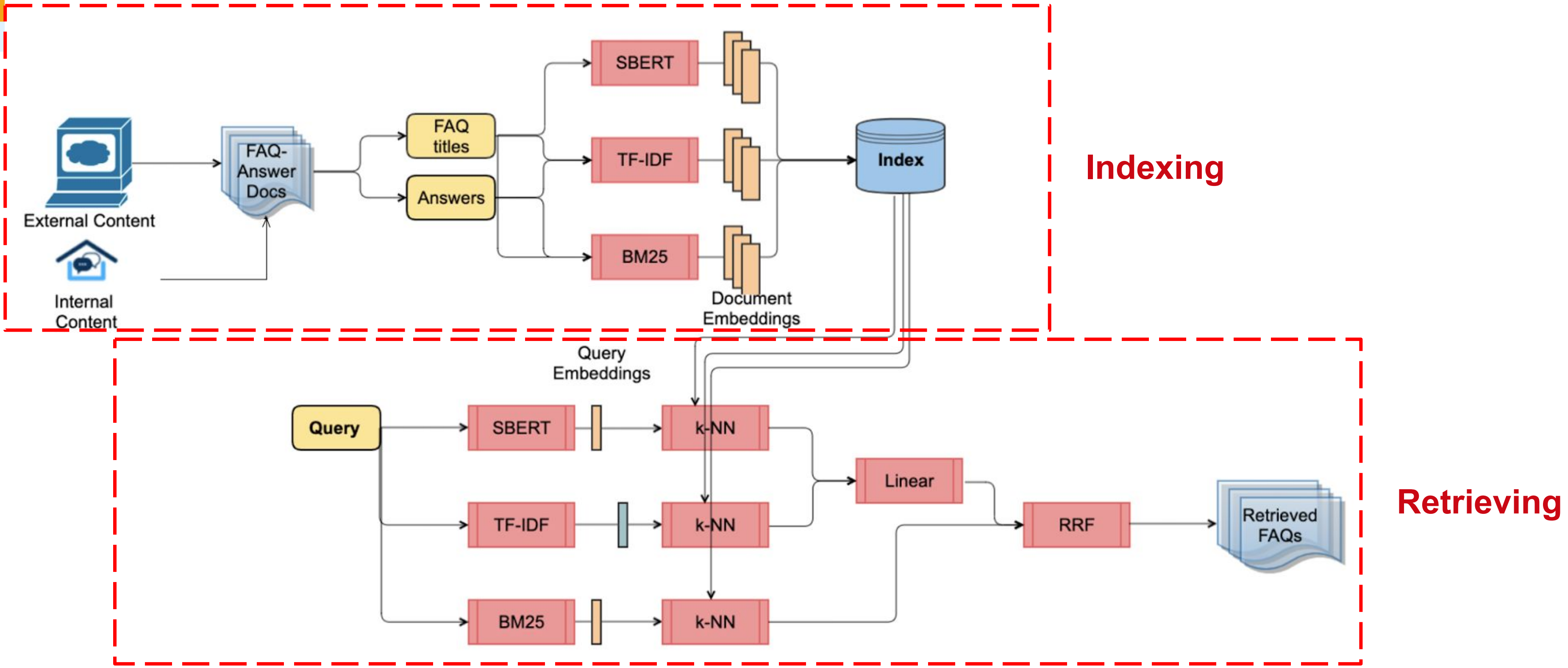
Retrieving - continued

4. BM25 scores for (query, FAQ-title) and (query, answer) are computed separately and combined using the same weight as we did for tf-idf.

5. The final ranking is produced by a reciprocal rank fusion between the initial ranking (SBERT+tfidf) and the BM25 ranking (k is a hyperparameter).

$$\mathrm{RRF}(Q, (q, a)) = \frac{1}{k + R_{\mathrm{SBERT+TFIDF}}} + \frac{1}{k + R_{\mathrm{BM25}}} \tag{4}$$

System Architecture



Zillow

Results

Algorithm	Mean Reciprocal Rank	Recall@5	nDCG@5	MAP@5	MAP
TF-IDF	0.8507	0.2370	0.6160	0.2289	0.4485
BM25	0.8221	0.2431	0.6579	0.2288	**0.4923**
SBERT+TF-IDF	**0.9375**	0.2583	**0.6625**	**0.2518**	0.4439
SBERT+TF-IDF+BM25	0.8828	**0.2589**	0.6600	0.2308	0.4874

- Including SBERT leads to better ranking for the top 5 retrieved documents
- BM25 catches up when ranking extended beyond the top 5
- Example queries where the relevant result cannot be recovered unless we use semantic search (minimal lexical overlap between the query and the right FAQ)

Query	Relevant FAQ
What credit counseling advice can we give borrowers?	What resources can I provide to applicants to help improve their credit score?
Can we originate a loan for a home on the market?	Can a property be refinanced if it is currently listed for sale?
Minimum size for manufactured house?	What are the requirements for a living unit?

Zillow

Future Work

- Extend SBERT embeddings to answers
- Utilize CQA signals, such as upvotes or approvals, for more effective retrieval
- Adopt supervised training approaches such as Learning to Rank[1] based on the collected user data
- Use weakly supervised approaches to create more labeled data to fine-tune SBERT to mortgage domain
  - Augmented SBERT to generate more labelled data[2]
  - Create semantically similar sentences using GPT-2 paraphrasing capability to train SBERT[3]
- Unsupervised sentence embeddings (e.g. TSDAE[4], GenQ[5])

[1] Thorsten Joachims. 2006. Training Linear SVMs in Linear Time. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining(KDD) (2006). https://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html
[2] Thakur, N., Reimers, N., Daxenberger, J. and Gurevych, I., 2020. Augmented sbert: Data augmentation method for improving bi-encoders for pairwise sentence scoring tasks. arXiv preprint arXiv:2010.08240.
[3] Mass, Y., Carmeli, B., Roitman, H. and Konopnicki, D., 2020, July. Unsupervised FAQ retrieval with question generation and BERT. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 807-812).
[4] Wang, K., Reimers, N. and Gurevych, I., 2021. TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning. arXiv preprint arXiv:2104.06979.
[5] Thakur, N., Reimers, N., Rücklé, A., Srivastava, A. and Gurevych, I., 2021. BEIR: A Heterogenous Benchmark for Zero-shot Evaluation of Information Retrieval Models. arXiv preprint arXiv:2104.08663.