ColdRoute: Effective Routing of Cold Questions in Stack Exchange Sites

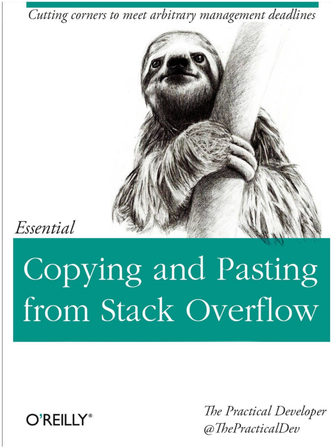
Jiankai Sun [1]
Abhinav Vishnu [2] Aniket Chakrabarti [3] Charles Siegel [2]
Srinivasan Parthasarathy [1]

[1]The Ohio State University

[2]Pacific Northwest National Laboratory [3]Microsoft

ECML PKDD, Sep 2018

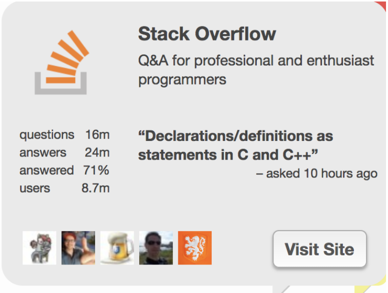
OOP vs SOP



[0] Left pic: `https://goo.gl/5vKuaR`
[0] Right pic: `https://goo.gl/XAG4DP`

Motivation

- What if you could not find answers for your questions in Stack Overflow? 4.8m unanswered
- Find the right experts to answer cold questions



0 Pic: `https://goo.gl/3VEsV9`

Find the right experts to answer cold questions

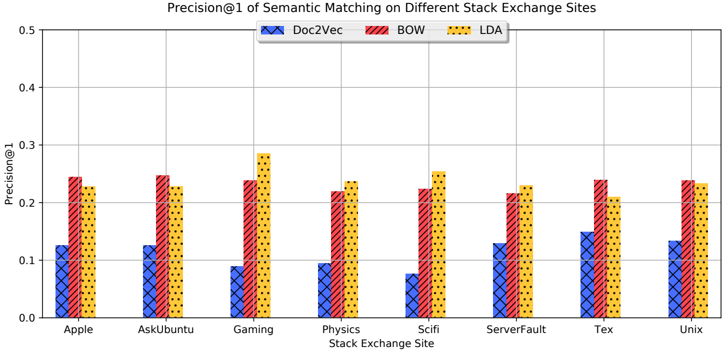
- Cold questions: newly posted questions (no answers) asked by new registered or existing askers
- Cold-start problem: routing cold questions to right experts
- A common problem in recommender systems [Cheng et al. WWW'17, Wang et al. TIST'14, CIKM'12, Sun et al. CIKM'12]

Semantic Matching Methods

- Leverage textual information to route cold questions
- How to model textual information?
  - BOW : bag of words [Zhou et al. 2012; 2013; Figueroa and Neumann 2013;]
  - LDA: Latent Dirichlet allocation, Topic Modeling [Guo et al. 2008; Ji et al. 2012]
  - Doc2Vec: Distributed Representations of Sentences and Documents [Le and Mikolov 2014; Dong et al. 2015]
- Users with highest semantic similarity will be selected as the best answerers

Experts finding by semantic matching (questions and best answers)

- $Precision@1$ computes the average number of times that the best answer is ranked in top-1: $Precision@1 = \frac{\{q \in Q | r^q_{best} == 1\}}{|Q|}$



Challenge I:

Find the indicator of the best answerer

Up Vote

3

Down Vote

1. MLP is sensitive to feature scaling. Have you normalzied your data?
2. Modify your network structure: add more hidden layers and change number of perceptrons of each layer
3. change activation function to sigmod/tanh/relu etc.

share improve this answer

answered Jun 5 '17 at 18:07

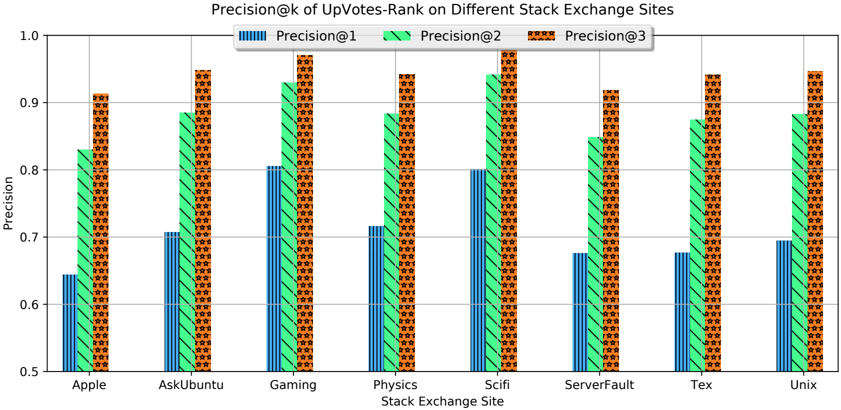
zhenv5

103 • 4

4. Change learning rate: `learning_rate`, `learning_rate_init`. 5. Toggle `early_stopping` – ijoseph Apr 10 at 22:13

How about voting score (up-votes - down-votes)?

UpVotes-Rank: select the answerer with the highest voting score as the best answerer

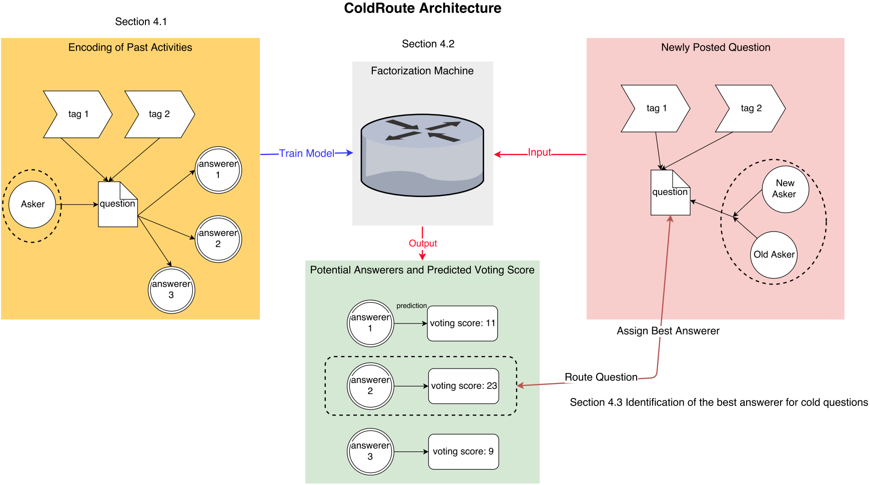


A Regression Problem

- How to identify the best answerer for a newly posted question?

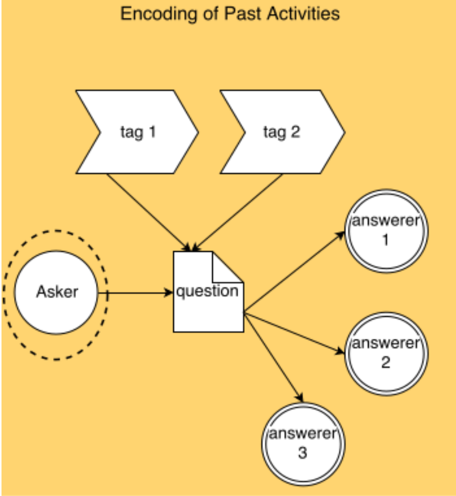
- Predict voting score for each question-answerer pair
- Select the user who has the highest voting score as the best answerer

Architecture of ColdRoute



Challenges II:

How to encode past activities?



Encode Past Activities Factorization Machine [Rendle et al. 2010, 2012]

Table 2: Illustration of FM, the main component in our ColdRoute. Each row represents a feature vector $\mathbf{x}^{(i)}$ and its corresponding target (voting score) $y^{(i)}$. The first 4 columns (orange) represent one-hot encoding of questions (ids); the next 4 (yellow) represent one-hot encoding of answerers (ids); The next 4 columns (blue) hold the one-hot encoding of corresponding askers (ids); The last 4 columns (green) are indicator variables for question tags.

	Feature Vector $\mathbf{X}$																	Target $\mathbf{y}$
$\mathbf{x}^{(1)}$	0	0	1	$\cdots$	1	0	0	$\cdots$	1	0	0	$\cdots$	0	1	0	$\cdots$	4	$y^{(1)}$
$\mathbf{x}^{(2)}$	0	0	1	$\cdots$	0	1	0	$\cdots$	1	0	0	$\cdots$	0	1	0	$\cdots$	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	0	0	1	$\cdots$	0	0	1	$\cdots$	1	0	0	$\cdots$	0	1	0	$\cdots$	2	$y^{(3)}$
$\mathbf{x}^{(4)}$	0	1	0	$\cdots$	0	0	1	$\cdots$	0	0	1	$\cdots$	0	0	1	$\cdots$	5	$y^{(4)}$
$\mathbf{x}^{(5)}$	0	1	0	$\cdots$	0	1	0	$\cdots$	0	0	1	$\cdots$	0	0	1	$\cdots$	6	$y^{(5)}$
$\mathbf{x}^{(6)}$	1	0	0	$\cdots$	1	0	0	$\cdots$	0	1	0	$\cdots$	1	1	1	$\cdots$	2	$y^{(6)}$
$\mathbf{x}^{(7)}$	1	0	0	$\cdots$	0	0	1	$\cdots$	0	1	0	$\cdots$	1	1	1	$\cdots$	4	$y^{(7)}$
	q_1	q_2	q_3	$\cdots$	u_1	u_2	u_3	$\cdots$	a_1	a_2	a_3	$\cdots$	t_1	t_2	t_3	$\cdots$		
	Question				Answerer				Asker				Question Tags					

Challenges III:
Feature vector is very sparse

- Each feature vector $\vec{x}^{(i)}$ has only $(3 + ||\vec{t_i}||_1)$ ones.
- $||\vec{t_i}||_1$ represents question q_i's number of tags (number of ones in the vector $\vec{t_i}$).
- Average number of tags per question in our experiments is 2.5

Advantages of Factorization Machine [Rendle et al. 2010, 2012]

- can handle sparse settings very well in comparing with other regressors such as linear regression
- give us the flexibility to explore the different features' relative importance in cold question routing (feature selection)
    - asker
    - tags
    - textual descriptions: question head and question body

A Toy Example

- Given a cold question q_4 asked by a new asker a_4 with tags $t = \{t_1, t_2, t_3\}$, Predict voting score of u_3?

- Linear Regression

$$\hat{y}(\vec{x}) = \boxed{w_0 + w_{q_4} + w_{u_3} + w_{a_4} + \sum_{i=1}^{3} w_{t_i}} + \sum_{i \in \boldsymbol{S}} \sum_{j \in \boldsymbol{S}, i<j} \langle \vec{v}_i, \vec{v}_j \rangle$$

- where $w_0 \in \mathbb{R}, \vec{w} \in \mathbb{R}^p, \vec{V} \in \mathbb{R}^{p \times k}$

A Toy Example

- Given a cold question q_4 asked by a new asker a_4 with tags $t = \{t_1, t_2, t_3\}$, Predict voting score of u_3?

- Linear Regression

$$\hat{y}(\vec{x}) = \boxed{w_0 + w_{q_4} + w_{u_3} + w_{a_4} + \sum_{i=1}^{3} w_{t_i}} + \sum_{i \in \boldsymbol{S}} \sum_{j \in \boldsymbol{S}, i<j} \langle \vec{v}_i, \vec{v}_j \rangle$$

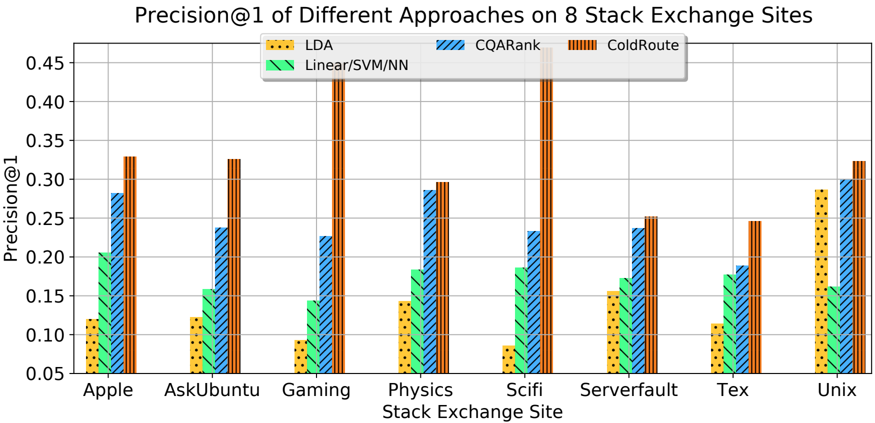
- Interactions among question, asker, answerer, and tags $\boldsymbol{S} = \{q_4, u_3, a_4, t_1, t_2, t_3\}$

- where $w_0 \in \mathbb{R}, \vec{w} \in \mathbb{R}^p, \vec{V} \in \mathbb{R}^{p \times k}$

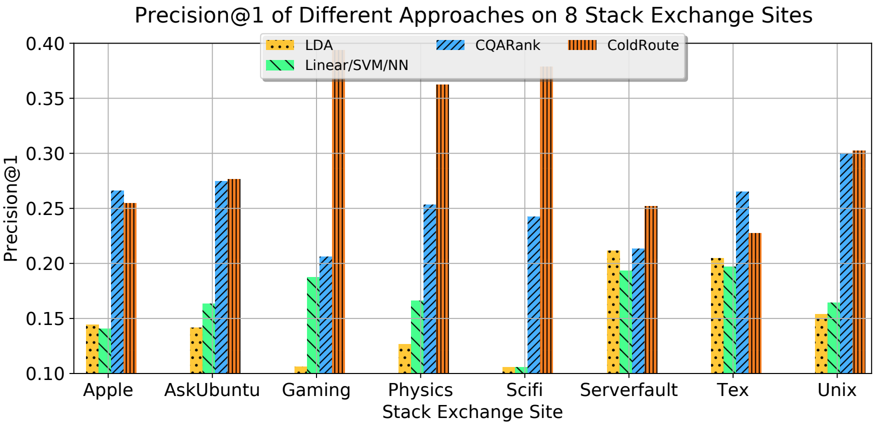
Experiments

- Datasets: 8 Stack Exchange Sites (Apple, Physics, Gaming, etc)
- Evaluation metrics: Mean Reciprocal Rank (MRR), Precision@k, and Accuracy
- Two different kinds of cold questions:
  - asked by new registered users
  - asked by existing users (have asked questions before)

Routing Cold Questions Asked by Existing Askers (ColdRoute vs state-of-the-arts)



Routing Cold Questions Asked by New Askers (ColdRoute vs state-of-the-arts)



Feature Selection for Routing Cold Questions Asked by New Askers (Evaluated by $Precision@1$)

	Apple	Ask.	Gaming	Physics	Scifi
BOW	0.1293	0.1285	0.0625	0.1004	0.0745
Doc2Vec	0.1331	0.1285	0.0625	0.1048	0.0621
LDA	0.1445	0.1416	0.1063	0.1266	0.1056
ColdRoute-HB	0.1483	0.1649	0.1438	0.1441	0.0994
ColdRoute-B	0.1825	0.1734	0.1813	0.1703	0.1429
ColdRoute-H	0.1711	0.1852	0.1875	0.1572	0.1429
Linear/SVM/NN	0.1407	0.1634	0.1875	0.1659	0.1056
CQARank	0.2662	0.2745	0.2062	0.2533	0.2422
ColdRoute-T	0.2548	0.2767	0.3938	0.3624	0.3789
ColdRoute-TA	0.2471	0.2789	0.3688	0.3537	0.3727

Conclusion and Future Work

- present ColdRoute for tacking cold questions routing in CQAs
- leverage Factorization Machine on one-hot encoding of critical features (question tags and askers)
- find that question tags play a more important role than information of askers, question body and title
  - 70% of askers have only asked only 1 question and the average number of questions per asker has asked is only 2.5
  - with CQAs growing and information of askers becoming dense, *ColdRoute-TA* will be more robust and efficient.
- Combine the power of deep neural network and factorization machines

Q & A

THE OHIO STATE UNIVERSITY

Appendix

FM as a regressor

Consider a 2-way FM ($d = 2$) as an example:

$$\hat{\vec{y}}(\vec{x}) = w_0 + \sum_{i=1}^{p} w_i x_i + \sum_{i=1}^{p} \sum_{j=i+1}^{p} x_i x_j < \vec{v}_i, \vec{v}_j > \quad (1)$$

where the model parameters that have to be estimated are:

$$w_0 \in \mathbb{R}, \vec{w} \in \mathbb{R}^p, \vec{V} \in \mathbb{R}^{p \times k} \quad (2)$$

And $< \cdot, \cdot >$ is the dot product of two vectors of size k:

$$< \vec{v}_i, \vec{v}_j >= \sum_{f=1}^{k} v_{i,f} v_{j,f} \quad (3)$$

where a row $\vec{v}_i \in \vec{V}$ describes the i-th variable with $k \in \mathbb{N}_0^+$ factors. k represents the dimensionality of the factorization.

Gradient descent to update parameters

$$\frac{\partial}{\partial\theta}\hat{\vec{y}}(\vec{x}) = \begin{cases} 1, & \text{if } \theta \text{ is } w_0 \\ x_i, & \text{if } \theta \text{ is } w_i \\ x_i \sum_{j=1}^{p} v_{j,f} x_j - v_{i,f} x_i^2, & \text{if } \theta \text{ is } v_{i,f} \end{cases} \quad (4)$$

Evaluation Metric

MRR.

The MRR measure is given by

$$MRR = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{r^{q}_{best}} \tag{5}$$

where r^{q}_{best} is the position of question q's best answerer in the predicted ranking list. It's worth mentioning that MRR is equivalent to Mean Average Precision (MAP) since the number of correct elements (the best answerer) in the predicted ranking list is just 1.

Evaluation Metric

Precision@k.

The $Precision@k$ is applied to measure the average number of times that the best answerer is ranked on top-k by a certain algorithm.

$$Precision@k = \frac{\{q \in Q | r_{best}^{q} <= k\}}{|Q|} \tag{6}$$

Accuracy. The Accuracy is used to measure the ranking quality of the best answerer, given by

$$Accuracy = \frac{1}{|Q|} \sum_{q \in Q} \frac{|R^q| - r_{best}^{q}}{|R^q| - 1} \tag{7}$$

Where $Accuracy = 1$ (best) means that the best answerer returned by an algorithm always ranks on top while $Accuracy = 0$ means the opposite.

Statistics of Stack Exchange Sites (Ask., Ser. are short for AskUbuntu and Serverfault respectively)

	Apple	Ask.	Gaming	Physics	Scifi	Ser.	Tex	Unix
# Questions	80,466	257,173	75,696	93,529	38,026	238,764	129,182	111,505
# Answers	119,878	337,198	130,294	137,258	78,652	398,470	169,354	171,016
# Unique Users	65,851	189,955	51,192	41,115	26,673	130,951	48,049	65,279
# Questions having Best Answers	29,765	85,843	45,798	38,094	21,740	117,275	76,862	53,856
# Unique Tags	1,048	3,020	4,437	876	2,349	3,514	1,525	2,438
Avg # Tags per Question	2.824	2.6982	1.2823	2.9634	2.1967	2.882	2.2752	2.7868
# Askers	40,206	137,171	25,153	31,415	12,413	93,739	42,819	45,773
# Asker (asked only 1 question) (%)	76.74%	75.88%	74.23%	63.26%	74.71%	64.04%	62.55%	68.48%
Avg # Questions per Asker	1.9758	1.8557	2.9689	2.8849	3.0031	2.4411	2.9851	2.4022