

Finding Relevant Answers in Software Forums

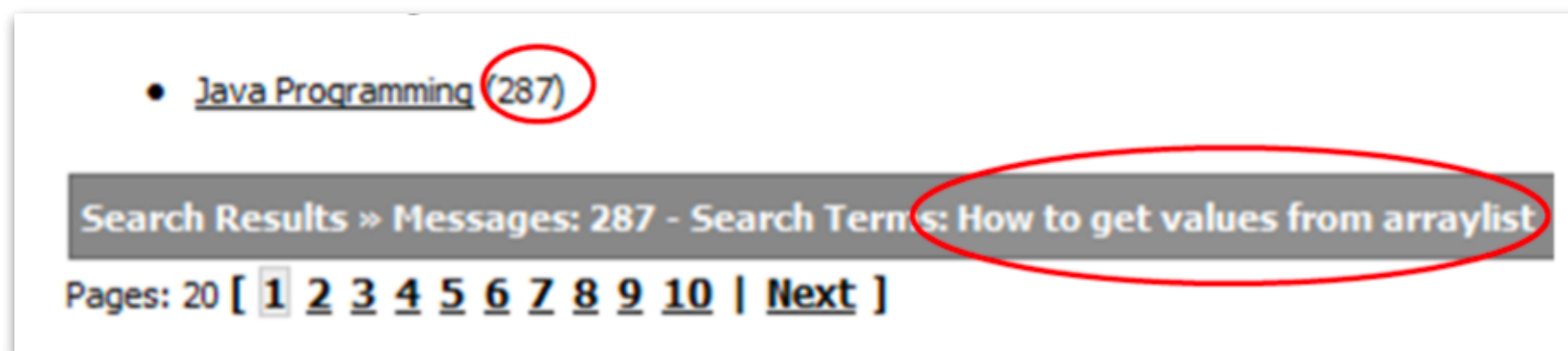
Swapna Gottipati

David Lo

Jing Jiang

背景

- 開発者やユーザーはソフトウェアの問題に直面した時、プロジェクトのフォーラムを利用する
- しかし、フォーラムには多数の投稿があるため目的の投稿を探し出すことは大変な作業となる場合もある
 - 目的の投稿を見つけても解決策が書かれているとは限らない



arraylistから値を取得する方法を検索した例

目的

- ソフトウェアフォーラムから関連する回答を見つけるためのフレームワークを提案する

典型的なフォーラムの投稿パターン

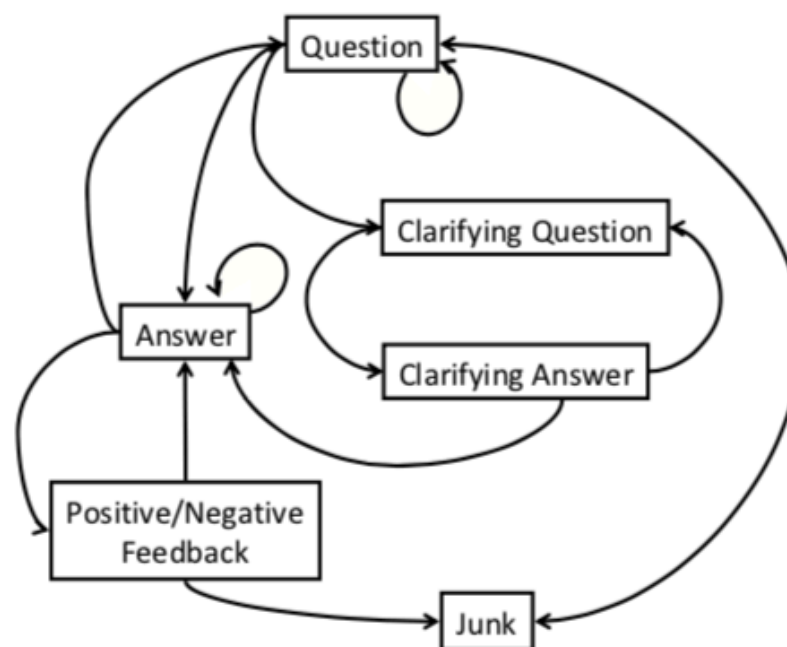


Fig. 3. Question-Conversation Graph

TABLE I
A TYPICAL CONVERSATION

Post Type	Author	Post Content
Question	Lisa	I have a JTable with 4 columns. The first column is a jCheckBox which is working fine. The other three are JComboBoxes. They work fine when you click them and select from the list, but our end users only want to navigate with the keyboard. Below is the code for my table. Any help .. code snippet..
Clarifying questions	Christiaan	When you tab to the component, does it become active
Clarifying answer	Lisa	When I tab to the combobox it doesn't seem to have focus on the component unless I hit the F2 key.
Answer	Christiaan	Hi Lisa, I think your best bet is to add KeyListener to the table addKeyListener(KeyListener l) and listen to certain keys and next call JTable.editCellAt(int row, int column) to put the cell in edit mode...
Clarifying question	Lisa	but Since I have four columns, the first being a checkbox which is working fine and the next three being ComboBoxes, would I need to check which column that is selected
Answer	Christiaan	Hi, I think the following code should help you: some code lines....
Positive Feedback	Lisa	THANK YOU, THANK YOU, THANK YOU!!!!!!!!!!!!!!!!!!!!!! It works great. You are a lifesaver. Lisa
Junk	Linda	I'm a newbie from Kansas. Really like this forum so far and am looking forward to contributing!

アプローチ

- 機械学習を用いて投稿に, PostTypeの7つのタグをつけることで目的の回答を探しやすくする

TABLE I
A TYPICAL CONVERSATION

Post Type	Author	Post Content
Question	Lisa	I have a JTable with 4 columns. The first column is a jCheckBox which is working fine. The other three are JComboBoxes. They work fine when you click them and select from the list, but our end users only want to navigate with the keyboard. Below is the code for my table. Any help .. code snippet..
Clarifying questions	Christiaan	When you tab to the component, does it become active
Clarifying answer	Lisa	When I tab to the combobox it doesn't seem to have focus on the component unless I hit the F2 key.
Answer	Christiaan	Hi Lisa, I think your best bet is to add KeyListener to the table addKeyListener(KeyListener l) and listen to certain keys and next call JTable.editCellAt(int row, int column) to put the cell in edit mode...
Clarifying question	Lisa	but Since I have four columns, the first being a checkbox which is working fine and the next three being ComboBoxes, would I need to check which column that is selected
Answer	Christiaan	Hi, I think the following code should help you: some code lines....
Positive Feedback	Lisa	THANK YOU, THANK YOU, THANK YOU!!!!!!!!!!!!!!!!!!!! It works great. You are a lifesaver. Lisa
Junk	Linda	I'm a newbie from Kansas. Really like this forum so far and am looking forward to contributing!

アプローチ概要

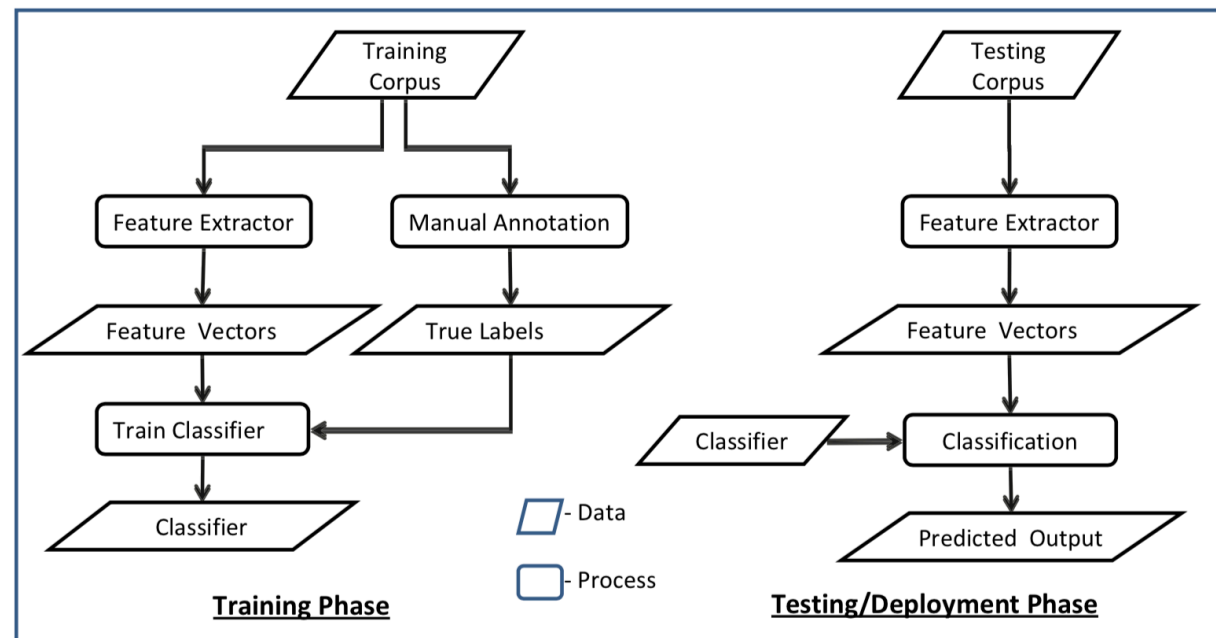


Fig. 5. Proposed Tag Inference Engine

- 特徴量をBoWにより抽出
 - 1.STOP：ストップワード
 - 2.STOPA：ストップワード＋著者情報
 - 3.STEMA：ステミング＋著者情報
 - 4.SSA：ストップワード＋ステミング＋著者情報
- 分類器
 - ・ サポートベクターマシン
 - ・ 隠れマルコフとサポートベクターマシンの組み合わせ

評価実験

- データセット（各フォーラムから投稿を取得）
 - SoftwareTipsandTricks : 4020件
 - DZone : 680件
 - Oracle : 1368件
 - 手動で投稿にタグ付けを行う
- 訓練データ
 - SoftwareTipsandTricks : 2000件
 - DZone : 300件
 - Oracle : 620件

分類の正確さ

SVM+マルコフモデル

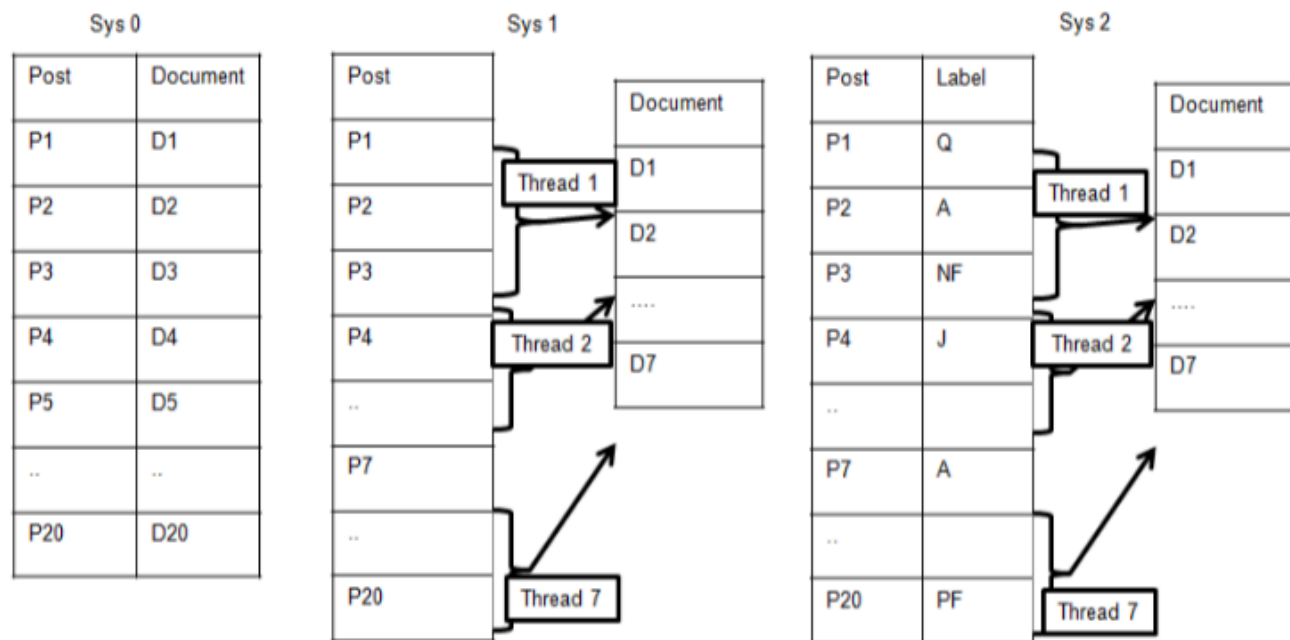
SVM

Dataset		Classifiers							
		HSSA	HSTEMA	HSTOPA	HSTOP	MSSA	MSTEMA	MSTOPA	MSTOP
Q	Prec	89	87	88	88	90	90	88	87
	Recall	82	83	82	74	68	68	70	64
	F-measure	85	85	85	80	77	77	78	74
A	Prec	80	78	74	68	68	67	69	60
	Recall	71	70	80	60	63	63	58	40
	F-measure	75	74	76	64	65	65	63	48
CQ	Prec	64	64	65	65	44	57	18	4
	Recall	63	55	70	68	57	57	71	55
	F-measure	64	59	67	67	50	57	29	8
CA	Prec	28	40	48	48	9	13	15	4
	Recall	40	36	45	39	13	13	22	33
	F-measure	32	37	46	43	10	13	18	7
PF	Prec	54	51	54	42	0	0	1	3
	Recall	65	69	56	58	0	0	0	0
	F-measure	59	59	55	48	0	0	0	0
NF	Prec	6	57	33	21	0	0	2	3
	Recall	57	46	59	62	0	0	37	0
	F-measure	11	51	42	31	0	0	4	0
J	Prec	77	76	76	79	67	67	70	76
	Recall	65	63	60	60	53	54	60	51
	F-measure	70	69	67	68	59	60	65	61

SVM+マルコフモデルの方が分類の精度が良い
ストップワード＋著者情報＋ステミングが良い

提案手法の検索エンジンの評価

3つの検索手法で比較



sys2が提案手法

- sys0 : 全文検索
- sys1 : スレッド単位で
- sys2 : スレッド単位で回答とタグづけされたもの

- ・ 投稿の上位20を取得
- ・ 検索はツールlemur[19]を使用

検索クエリ

What is javax telephony InvalidArgumentException?
How to calculate Pi?
How to change this clock object to display 12 hour time instead of 24?
How to execute internet explore or windows word in java program?
What are the differences between a hash map and a hash table?
What is an invalid argument?
How to access CVS files?
How to set a combo box with different fonts?
How to generate random numbers or random booleans?
How to read files in Java?
How to generate prime numbers?
How to remove duplicates in an ArrayList?
How to get values from an ArrayList?
How to calculate the difference between two dates?
How to create a two dimensional array or 2D vector?
How to encrypt?
How to read a PDF file or PDF document?

提案手法の検索エンジンの評価

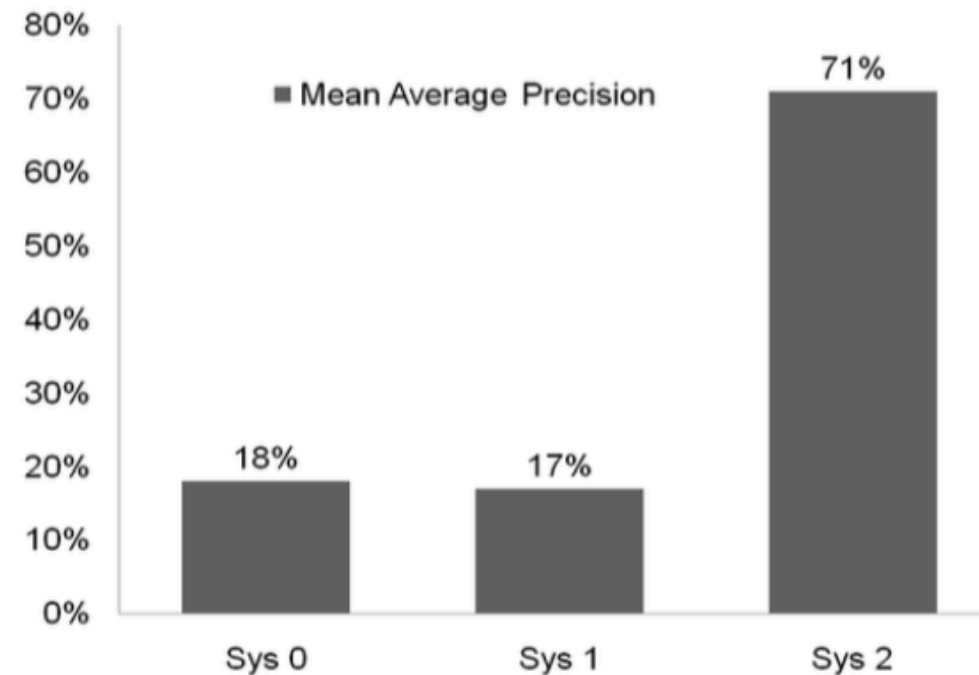


Fig. 12. MAP comparison for 17 queries.

- Mean Average Precision
 - 各クエリでの検索の平均適合率の平均
- 平均適合率
 - 各検索クエリでの検索の再現率ごとの適合率の平均

提案手法が最も精度が良い

まとめ

- フォーラムの投稿に機械学習を利用して7種類のタグづけを行なった
 - 分類器の評価を行なった
 - 提案手法搭載の検索エンジンの評価を行なった
- 特徴抽出はSSA, 分類器はSVM+マルコフモデルの精度が良い
- 提案手法の検索エンジンでは, 全文検索に比べてMean Average Precisionが53%高い

所感

- 約6000件の投稿の手動タグづけはかなり大変そう
- 提案手法の検索エンジンの評価方法は勉強になった
 - やはり自分で正解集合を作っている
 - レコメンドシステムの評価方法を勉強したい
 - ▶ nDCG (normalized Discounted Cumulated Gain) とか