



AUTOMATIC IDENTIFICATION OF BEST ANSWERS IN ONLINE ENQUIRY COMMUNITIES

GRÉGOIRE BUREL, YULAN HE AND HARITH ALANI
Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom

Extended Semantic Web Conference 2012.
Heraklion, Crete, GR. 2012

OUTLINE

- Enquiry Communities
 - A Definition
 - Structure of Enquiry Communities
 - Needs and Motivation
 - Contributions
- Computing and Mapping Features for Content Quality
 - Predictors
 - Datasets
 - Feature Computation: Users, Content and Threads.
 - Feature Mapping
- Identifying Best Answers
 - Prediction Results
 - Feature Ranking
- Findings
- Future Work
- Conclusion

ENQUIRY COMMUNITIES

“Enquiry Communities are communities composed of *askers* and *answerers* looking for *solutions* to particular issues.”

ENQUIRY COMMUNITIES


“Enquiry Communities are communities composed of **askers** and **answerers** looking for **solutions** to particular issues.”

Question Thread	Question
	Answer #1
	Answer #2
	...
	Answer #n



ENQUIRY COMMUNITIES

“Enquiry Communities are communities composed of **askers** and **answerers** looking for **solutions** to particular issues.”

Question Thread	Question
	Answer #1
	Answer #2 
	...
	Answer #n



ENQUIRY COMMUNITIES



	StackExchange	serverfault	Cooking	SAP SCN
Users	51,727	4,941	32,718	
Questions	71,962	3,064	95,017	
Answers	162,401	9,820	421,873	

- Enquiry communities have different shapes:
 - Size (e.g. small vs. large)
 - Topics (e.g. programming vs. cooking)
 - Users (e.g. business vs. recreational)
 - Features (e.g. community ratings vs. author ratings)

Feature		SAP Forums	StackExchange
Type	Name		
User	Profile	●	●
	Bookmarks	—	●
	Accepts	○	●
	Points	○	●
	Classes	○	○
	Levels	○	○
	Achievements	—	●
	Abilities	○	●
	Rewards	○	●
	Leading Board	○	●
	Thread	Views	●
Votes		○	●
Lock		○	●
Sticky		●	—
Bookmarks		—	●
Categories		●	—
Tags		—	●
Question	Status	●	●
	Votes	—	●
	Comments	○	●
	Modification	○	●
Answer	Accepted	●	●
	Votes	○	●
	Comments	○	●
	Modification	○	●

Abbreviations: ● = Yes. ○ = Limited/Partial. — = No.

ENQUIRY COMMUNITIES

- Enquiry Communities Needs:
 - Community Managers:
 - Make sure that the community is “happy” (questions are solved).
 - Identify and implement features that help users goals.
 - Askers:
 - Find answers related to a given issue.
 - Identify best answers for a given question.
 - Answerers:
 - Identify unanswered questions.
 - Provide the best answers

ISSUES

- Enquiry Communities Needs:
 - Questions can have many answers (~ 9 answers, up to > 100):
 - Difficulty to identify best answers
 - Best answers are not always annotated by the askers ($\sim 50\%$ of the questions of our datasets do not mention best answers):
 - Some questions can be solved but lack such information.
 - Platform features can support the identification of best answers, but which ones?
 - What features help best answer identification? What features should be implemented?

IDENTIFYING BEST ANSWERS

How user, content, thread and platform features affect quality content identification?

1. Identifying Best Answer (helping user to find relevant content)

- Normalise three different communities/datasets in a SIOC based Q&A Vocabulary
- Generate 19-23 features of each dataset
- Assess the impact of features groups on best answer identification

2. Analysis of Quality Predictors Across Communities (helping community manager to identify important community and platform features)

- Rank each features for each communities
- Estimate the importance of particular features within each communities

CONTRIBUTIONS

How user, content, thread and platform features affect quality content identification?

- Contributions:
 - Performance study of best answers identification across different communities and platforms.
 - Analysis and Introduction of thread based features
 - Influence of user, content and thread features on predictions
 - Impact of platform features on content quality
- Literature:
 - Single domain analysis
 - Non Thread features
 - No Cross-domain/platform impact analysis

FEATURES

1. User Features:

- Represents the characteristics and reputation of authors (e.g. reputation, number of best answers, normalised activity entropy...).

FEATURES

1. User Features:

- Represents the characteristics and reputation of authors (e.g. reputation, number of best answers, normalised activity entropy...).

2. Content Features:

- Questions and answers features representing importance and quality (e.g. readability, ratings, number of views...).

FEATURES

1. User Features:

- Represents the characteristics and reputation of authors (e.g. reputation, number of best answers, normalised activity entropy...).

2. Content Features:

- Questions and answers features representing importance and quality (e.g. readability, ratings, number of views...).

3. Thread Features:

- Represents relation between answers within a particular thread. (e.g. score ratio, topical reputation ratio...).

FEATURES

		Features Set	
Type	Core Features Set (19)	Extended Features Set [†] (23)	
User	<i>Reputation, Post Rate, Normalised Activity Entropy, Number of Answers, Answers Ratio, Number of Best Answers, Best Answers Ratio, Number of Questions, Questions Ratio, Normalised Topic Entropy, Topical Reputation.</i> (10)	<i>Reputation, Age, Post Rate, Normalised Activity Entropy, Number of Answers, Answers Ratio, Number of Best Answers, Best Answers Ratio, Number of Questions, Questions Ratio, Normalised Topic Entropy, Topical Reputation.</i> (11)	
Content	<i>Answer Age, Number of Question Views, Number of Words, Gunning Fog Index, Flesch-Kinkaid Grade Level.</i> (5)	<i>Score, Answer Age, Number of Question Views, Number of Comments, Number of Words, Gunning Fog Index, Flesch-Kinkaid Grade Level.</i> (7)	
Thread	<i>Number of Answers, Answer Position, Relative Answer Position, Topical Reputation Ratio.</i> (4)	<i>Score, Number of Answers, Answer Position, Relative Answer Position, Topical Reputation Ratio.</i> (5)	

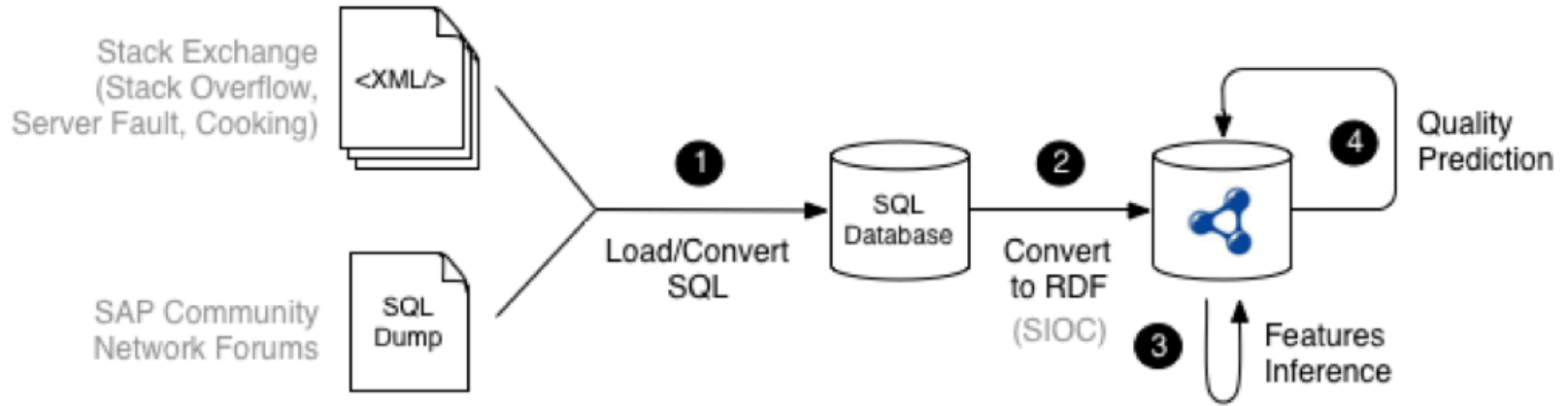
[†]Only valid for the *Server Fault* and *Cooking* datasets.

COMMUNITY DATASETS

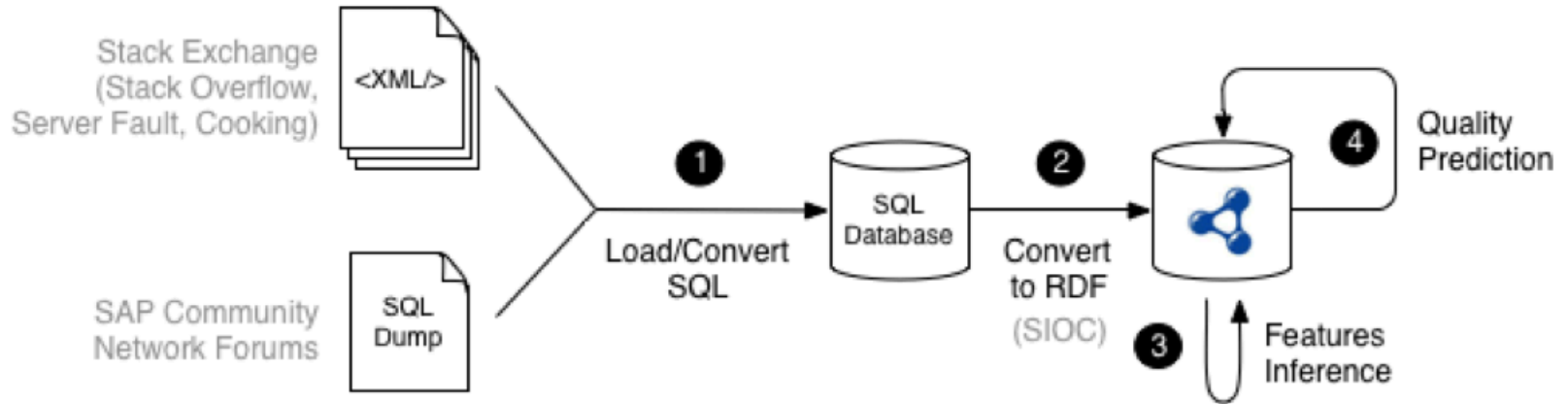
- Stack Exchange (SE, April 2011 Dataset)
 - A set of Q&A websites supporting best answer selection and community ratings.
 - Server Fault (SF): System administration (162k posts)
 - Cooking (CO): Cooking advices (9k posts)
- SAP Community Network Forums (SCN, July 2011 Dataset / 427k posts)
 - A forum based community supporting best answer selection. Provides support on SAP products (software and programming)



COMPUTATION & MAPPING



COMPUTATION & MAPPING



Input Dataset

SCN	SF and CO	RDF Output
User	User	sioc:OnlineAccount/foaf:Person
Thread (first thread Post)	Question	sioc:Question
Post (not in first position)	Answer	sioc:Answer
Post (with 10 points)	Best Answer	sioc:BestAnswer
-	Comment	sioc:Comment
Forum	Tag	sioc:Tag (topic)

BEST ANSWER IDENTIFICATION

- Experimental Setting:

1. Select the **question threads** that have best answers for SF (36,717 questions), CO (2,154 questions) and SCN (29,960 questions).
2. Map data to **SIOC/Q&A** and **compute features**.
3. Use **Multi-Class Alternating Decision Tree** learning algorithm and validate results using **10-folds cross validation**.
4. Compute **Precision (P)**, **Recall (R)**, **F-Measure (F_1)** and area under the **Receiver Operator Curve (ROC)** for different feature groups.

BEST ANS. PREDICTION RESULTS

Feature	SCN Forums				Server Fault				Cooking			
	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>AUC</i>	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>AUC</i>	<i>P</i>	<i>R</i>	<i>F</i> ₁	<i>AUC</i>
Words	0.536	0.732	0.619	0.616	0.592	0.621	0.537	0.567	0.671	0.705	0.644	0.651
Answer Score	-	-	-	-	0.643	0.656	0.625	0.673	0.751	0.760	0.753	0.797
Answer Score Ratio	-	-	-	-	0.808	0.809	0.806	0.848	0.866	0.868	0.866	0.916
Users	0.716	0.746	0.687	0.752	0.637	0.651	0.626	0.664	0.687	0.714	0.681	0.686
Content	0.712	0.740	0.659	0.678	0.647	0.659	0.628	0.679	0.708	0.727	0.707	0.754
Thread	0.820	0.827	0.817	0.865	0.753	0.756	0.749	0.809	0.765	0.772	0.751	0.785
All	0.833	0.839	0.831	0.880	0.770	0.769	0.760	0.827	0.777	0.784	0.767	0.816
Users+	-	-	-	-	0.637	0.651	0.626	0.664	0.687	0.714	0.681	0.686
Content+	-	-	-	-	0.700	0.707	0.699	0.761	0.788	0.793	0.789	0.842
Thread+	-	-	-	-	0.844	0.845	0.844	0.910	0.867	0.869	0.867	0.919
All+	-	-	-	-	0.848	0.847	0.844	0.912	0.870	0.872	0.870	0.919

BEST ANS. PREDICTION RESULTS

- Best Answer Identification (F_1 0.83-0.87):
 - Baseline Models:
 - Contrary to previous studies, content length is not a good predictor (Agichtein et al, 2008; Jeon et al. 2006).
 - Scores and scores ratio are very good features. Unfortunately, they are not available in SCN.
 - Core Features:
 - SCN performs better on core features compared to the other datasets.
 - Thread features are the best due to their ability to represent features “comparisons” within a thread.
 - Extended Features:
 - Extended features show better results than core features.

FEATURES RANKING

- Features Ranking:

1. For each feature, the **Information Gain Ratio (IG)** is computed
2. The **features are sorted by decreasing IG** so they can be compared across communities.

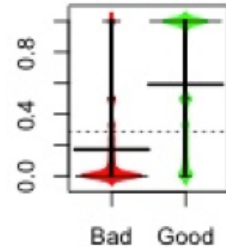
FEATURES RANKING RESULTS

R.	SCN		Server Fault		Cooking	
	IG	Feature	IG	Feature	IG	Feature
1	0.217	<i>Topic. Rep. Ratio (T)</i>	0.332	<i>Score Ratio (T)</i>	0.430	<i>Score Ratio (T)</i>
2	0.196	<i>No. Answers (T)</i>	0.275	<i>No. Answers (T)</i>	0.190	<i>Score (C)</i>
3	0.108	<i>Bests Ratio (U)</i>	0.126	<i>Answer Position (T)</i>	0.164	<i>No. Answers (T)</i>
4	0.105	<i>Questions Ratio (U)</i>	0.117	<i>Topic. Rep. Ratio (T)</i>	0.120	<i>Answer Position (T)</i>
5	0.105	<i>Answers Ratio (U)</i>	0.097	<i>Relative Position (T)</i>	0.083	<i>Topic. Rep. Ratio (T)</i>
6	0.104	<i>Relative Position (T)</i>	0.070	<i>Score (C)</i>	0.074	<i>Bests Ratio (U)</i>
7	0.097	<i>Reputation (U)</i>	0.056	<i>Q. Views (C)</i>	0.070	<i>No. Bests (U)</i>
8	0.093	<i>Topic. Rep. (U)</i>	0.046	<i>Bests Ratio (U)</i>	0.069	<i>Reputation (U)</i>
9	0.090	<i>No. Bests (U)</i>	0.037	<i>No. Comments (C)</i>	0.065	<i>Answer Age (C)</i>
10	0.089	<i>Activity Entropy (U)</i>	0.022	<i>Topic Entropy (U)</i>	0.055	<i>Topic Entropy (U)</i>
11	0.064	<i>Answer Position (T)</i>	0.021	<i>Answer Age (C)</i>	0.054	<i>No. Comments (C)</i>
12	0.048	<i>No. Answers (U)</i>	0.019	<i>Post Rate (U)</i>	0.054	<i>No. Words (C)</i>
13	0.035	<i>Topic Entropy (U)</i>	0.018	<i>Reputation (U)</i>	0.053	<i>No. Answers (U)</i>
14	0.033	<i>Q. Views (C)</i>	0.017	<i>No. Bests (U)</i>	0.045	<i>Relative Position (T)</i>
15	0.027	<i>No. Words (C)</i>	0.016	<i>No. Answers (U)</i>	0.039	<i>Topic. Rep. (U)</i>

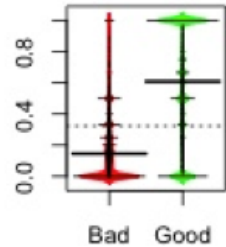
FEATURES RANKING RESULTS

- Features Impact Comparison:
 - Core Features:
 - Thread features are the most important.
 - SCN quality estimation is more affected by user features than the other datasets.
 - Historical user success is more important in SCN.
 - SF and CO show that first answers tend to be the best (due to the wiki nature of SE).
 - Extended Features:
 - Thread features are still the most important.
 - Platforms offering scores show high IG for such features. As a consequence community votes policies improve quality content identification.

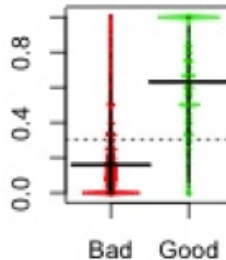
SCN Topic. Rep. Ratio



SF Score Ratio



CO Score Ratio



FINDINGS

- Different communities have different characteristics, goals and behaviours:
 - Results are not always community independent (e.g. importance of answer length in quality content)
 - However, our results seems transferable across the community we analysed.
- Community votes can be reinforced using additional features such as topical reputation and ratio models
 - Community votes alone do not identify perfectly best answers.
- Platform policies and feature influence the identification of best answers (e.g. community votes)
 - SAP as acknowledged the importance of community votes and has integrated community feedback to its new platform.

FUTURE WORK

- Evaluate regression models for estimating the quality of an answer independently of the best annotation (e.g. identify bad answers).
- Evaluate the overall quality of the content posted in different communities.
- Evaluating our models on thread that strictly have more than one answers.

CONCLUSION

- ~50% of questions do not have best answer annotations
- Accurate automatic identification of best answers can be achieved with good accuracy (F_1 0.83-0.87)
- Contrary to previous work (Agichtein et al, 2008; Jeon et al. 2006), answer length is uncorrelated with best answers.
- Community-based answer ratings are highly correlated with best answers ($F_1 > 0.8$ when using score ratios alone)
- Thread-based features are influential predictors

QUESTIONS?

Web: <http://evhart.online.fr>

Email: g.burel@open.ac.uk

Twitter: [@evhart](https://twitter.com/evhart)



REFERENCES

- Chai, K., Potdar, V., Dillon, T.: Content quality assessment related frameworks for social media. In: Proc. Int. Conf. on Computational Science and Its Applications (ICCSA), Heidelberg (2009)
- Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: First ACM Int. Conf. on Web Search and Data Mining, Palo Alto, CA (2008)
- Jeon, J., Croft, W.B., Lee, J.H., Park, S.: A framework to predict the quality of answers with non-textual features. In: SIGIR, Washington, ACM Press (2006)
- Rowe, M., Angeletou, S., Alani, H.: Predicting discussions on the social semantic web. In: Proc. 8th Extended Semantic Web Conf. (ESWC), Heraklion (2011)