

Content Ranking System based on Feedback from User

Kavita Kimmatkar, Kaustubh Alandkar, Sushant Ghodnadikar, Kishor Karbhari

Computer Engineering Department
Zeal College of Engineering & Research
Pune, India

Abstract—Availability of accurate content, when the user demands is a major challenge. To address this challenge present systems make use of learning tasks. Practically it is tedious to obtain a significant number of examples for training in learning tasks. This research proposes a system for effectively training acquired information from the user and using it to classify the content, by using only the essential information in a set of unclassified information. Forming a set of useful keywords of diverse type and ranking content based on the usefulness of reviews or feedback from the user. The committee (set of keywords) is obtained from probable possibilities for committee members by using set selected for training. This paper explains an approach which implicitly determines the significance of the expected facts to obtain from a training example, which makes it easy for implementation. The proposed system can help in providing reliable and quality content to the user with minimum delay using optimal resources.

Keywords—Content ranking, diverse keywords, committee, learning task

I. INTRODUCTION

In today's world, there is a continuous increase in the amount of content on the internet. So it is necessary to get only the useful content from the internet when demanded by any entity. To address this issue classification of content as essential and nonessential is needed to be done. The reason behind this is sorting important information from various sources to get required content when needed. A major challenge in the classification of content is how to sort the important content. To overcome this challenge a system is required, to perform the content ranking which will effectively classify content. This classification of content is possible using learning task.

Learning from the user in an interactive manner and processing the gathered information is done in the learning task. This can be done using active learning method which is efficient in acquiring information from the user. The gathered information is further processed by the diverse type of keywords formed from obtained data. There are many ways to form the committee i.e. set of keywords but our research is focused on the diverse type of members in committee.

This research proposes a system that will rank the content by taking the reviews from the user. When there is need of reliable and quality content this system can efficiently provide the required data to users. The ability to actively select the

most useful training examples is an important approach to reducing the amount of supervision required for effective learning [1]. This research was motivated by following points:

- There is a lack of classification techniques based on feedback from the user.
- Effective ranking using active learning.
- Need for quality, reliable content when required for analysis in real-time.

II. LITERATURE REVIEW

Most of the present data classification systems may solve the problem of data scarcity up to certain degree, but do not produce the whole solution at all situations. Techniques considered for research are briefly discussed below.

A. Active Learning with Heterogeneous Ensembles [2]

Active learning involves tasks such as discovering information, processing the information and applying it. One common approach in active learning is to choose one classifier and select data points that help the training of this classifier, which normally includes choosing data points according to some confidence measure[2]. This approach includes uncertainty sampling[3][4], in which data points that the current classifier is most uncertain about are considered informative[2].

Heterogeneous committee members adapt in different ways and are able to solve different problems. Measuring the competence of committee members helps in making competent and accurate decisions [5].

B. Asking Generalized Queries to Domain Experts to Improve Learning[6]

Often active learning can construct classifier by taking help of domain expert with more accuracy. The previous efforts using active learning only form classifier and query specifically. Considering implementation point of view domain experts proves themselves more useful by answering generalized queries. The significance of generalized query is that they are equivalent to many specific queries. Moreover, a general query is not efficient when answers from domain expert are not effective.

C. Combining Labeled and Unlabeled Data with Co-Training [7]

In this method problem where large unclassified information is used to increase the efficiency of the algorithm when only limited information is available. A setting is considered where the description of each portion of data is partitioned into distinct views and learning task is applied to classify content[7]. The assumption in this method is any view can be used for learning if labeled data is available in sufficient amount.

D. Comparing Performance of Committee Based Approaches to Active Learning[8]

This technique is based on constructing active learning systems based on query by committee. This approach considers the use of different measures of disagreement among component classifiers to select the most informative example to query an oracle for its label [8]. A technique is introduced based on analyzing the neighborhood of examples, which is applied to create a starting training set for generating the ensemble[8]. Results of this technique confirm that by using limited examples a final classifier can be generated.

The advantages and disadvantages of the techniques discussed above are mentioned in TABLE I.

TABLE I. COMPARISON OF TECHNIQUES

Technique Name	Advantages	Disadvantages
Asking Generalized Queries to Domain Experts to Improve Learning.	Generalized query is often equivalent to many specific queries.	General queries are not good as answers from the domain experts can be highly uncertain.
Combining Labeled and Unlabeled Data with Co-Training.	Less number of examples needed to learn.	Even for an optimal pair of functions occasionally inconsistent examples are seen.
Comparing Performance of Committee Based Approaches to Active Learning.	Classifiers are created using less number of queries to label examples.	Applicability and effectiveness of committee based sample do not consider nonprobabilistic contexts.
Active Learning with Heterogeneous Ensembles.	Information is discovered easily and diverse type of ensembles increase the accuracy of classification.	This technique is not efficient when using only one classifier type.

III. PROPOSED SYSTEM

By analyzing the information studied, this research proposes a system for ranking content using reviews or feedback from users.

A. BLOCK DIAGRAM OF SYSTEM

A block diagram is given below showing general architecture of the proposed system:

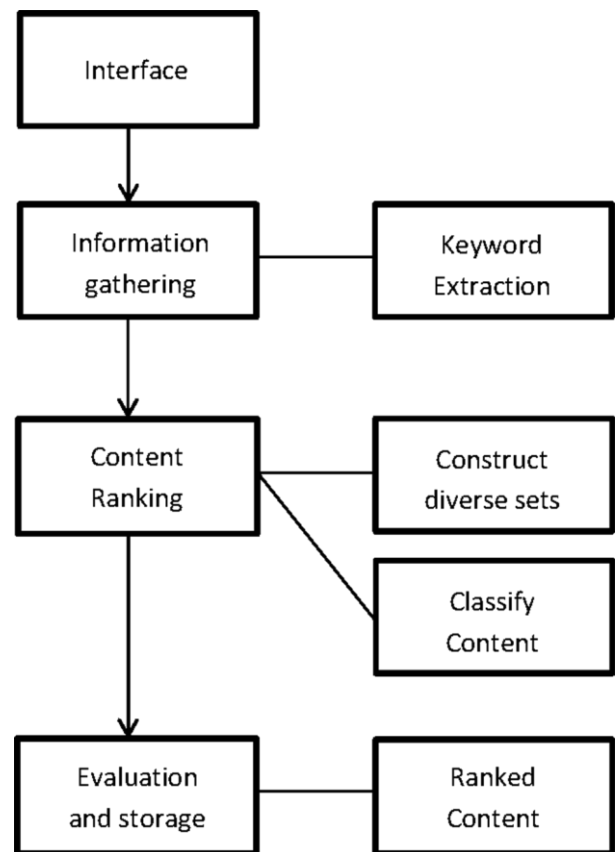


Fig. 1. Block diagram for Content Ranking System

B. WORKING OF THE SYSTEM

1) Information gathering

In this model, information is gathered using interactive frontend for certain content. Information of different categories is available to the user. Following categories would be considered as an example, technical document, articles, images, etc.

The system's approach is based on keywords in reviews or feedback. In our setting the ranking score gives the usefulness of the review, so we look for keywords to extract that are not easily observable using simple techniques. Our interest is in analyzing reviews that are most useful for ranking and affect the content credibility. The reviews may consist description of content, features, personal comments, etc.

2) Content Ranking

This part will process the obtained information from the user and categorize it by applying a certain label to it. The labeled data is further processed by the system and the content is ranked and displayed to the user. For understanding the usefulness of the text in reviews, we relied on active learning and heterogeneous ensembles. The active learning part acquires the information required for classification of reviews and different types of sets of keywords i.e. heterogeneous ensembles, are used for categorizing review as useful or not useful. Based on useful reviews the content is ranked. For constructing useful set of keywords, we consider

random selection of keywords and forming a set i.e. ensemble of different types of keywords.

After constructing the ensemble, we classify each review by assigning it labels and keeping count of its classification as useful. Hence, each review now contains a count of usefulness. Based on this information, we estimate the rank of the content. Based on this fact the system can provide quality content to the users.

3) Storage of information

The labeled information is stored in a database for further use. This information is used for constructing diverse sets and using them for content classification. The content is displayed to a user by retrieving the information stored in the database.

C. APPLICATIONS OF SYSTEM

The system is applicable where content ranking is required for providing quality and reliable data to the user. Following applications are possible:

- This system can be used as an interactive response gathering mechanism from people having diverse opinions.
- Besides this, it can provide most relevant and informative content to users by acquiring reviews or feedback from users.
- It can be used as a tool to rank content and control information on the internet.
- Also, this system can be used in training robots to learn useful information and improve better understanding.

IV. CONCLUSION

We contribute to the research in content ranking based on user reviews. This paper looks the text in user reviews in a unique way of how simple interaction between a machine and human can help in providing essential information for classification.

We also find that different type of reviews helps more by providing necessary data to the system for performing better. In terms of usefulness for classification, we observed diverse types of keywords perform better.

Based on our findings, proposed system can rank the content and process it to display the relevant and reliable content to the user using cost-effective resources.

REFERENCES

- [1] P. Melville and R. Mooney, Diverse Ensembles for Active Learning, Proc.Intl Conf. Machine Learning (ICML), pp. 584-591, 2004.
- [2] Z. Lu, X. Wu, and J. Bongard, Active Learning with Adaptive Heterogeneous Ensembles, Proc. IEEE Ninth Intl Conf. Data Mining (ICDM), pp.327-336, 2009.
- [3] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in Proceedings of Research and Development in Information Retrieval, 1994, pp. 3-12.
- [4] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in Proceedings of the 11th International Conference on Machine Learning, 1994, pp. 148-156.

- [5] N. Jankowski and K. Grabczewski. Heterogenous committees with competence analysis. In HIS '05: Proceedings of the Fifth International Conference on Hybrid Intelligent Systems, pages 417-424, Washington, DC, USA, 2005. IEEE Computer Society.
- [6] J. Du and C.X. Ling, Asking Generalized Queries to Domain Experts to Improve Learning, IEEE Trans. Knowledge and Data Eng., vol. 22, no. 6, pp. 812-825, June 2010.
- [7] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in Proceedings of the Workshop on Computational Learning Theory, 1998.
- [8] J.Stefanowski, M.Pachocki, Comparing Performance of Committee based Approaches to Active Learning. In: M.Kopotek, A.Przepirkowski, S.Wierzcho, K.Trojanowski (red.) Recent Advances in Intelligent Information Systems, Wydawnictwo EXIT, Warszawa, 2009, 457-470.