VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN EFFECTIVE FEATURE INTEGRATION IN IMAGE RICH INFORMATION NETWORKS

M. Maheswari[1] and K. S. Neelu Kumari[2]
[1]Department of Computer Science and Engineering, P.A College of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]Department of Electronics and Communication Engineering, P.A College of Engineering and Technology, Coimbatore, Tamil Nadu, India
E-Mail: maheswaribtechit@gmail.com

ABSTRACT

An image-rich information network is a social media website which contains billions of images uploaded by users which are associated with information about owner, consumer, producer, annotations, and comments are forming heterogeneous image-rich information networks. In feature extraction has been extract both global feature as visual and local feature as color for an image and the relevance feedback approach to increase retrieval accuracy and gives more personalized results for the user. In a hybrid approach which measures the similarity of images to retrieve based on the link based, content based and relevance feedback. The link similarity based on HMoK-SimRank algorithm depends upon the social network information like tags, groups and human annotation over the images. Content based similarity considers both global and local features, for measuring the similarity. In relevance feedback based approaches, a system learns from feedback provided by the user for an efficient way to retrieve the image. The Main advantage is to improve the image retrieval very effective and performance has to be improved.

Keywords: feature extraction, image retrieval, relevance feedback.

1. INTRODUCTION

An image retrieval system is a computer system for querying, retrieving images from a large database of digital images. Most common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or description to the images so that retrieval can be performed over the annotation words.

In social multimedia sharing and hosting websites, such as Flickr [1], Picsearch [2] contain billions of user-submitted images. Famous Internet commerce websites such as Amazon.com is also contains large amounts of product-related images. In an images in such social networks are also to be by metadata, comments, and other information, thus forming different image-rich information networks.

The direct use of link information solely based on human annotations may also lead to unsatisfying results if the annotation is meaningless, too general, or inadequate. In additionally, if the image could not link to other object in the information network, then link information cannot work. Using content similarity may lead to unsatisfying results. Image search is a specialized data search used to find images. For searching an image, a user may provide query terms such as keyword, image file, or click on any image, and the system will retrieve images similar to the query. The similarity used for searching can be metadata, tags, color distribution of images, region/shape attributes. Content based image retrieval is the application of computer vision to the image retrieval. CBIR aims to avoid the use of textual descriptions and instead retrieves images based on similarities in their contents such as textures, colors, and shapes to a user-supplied query image or user-specified image features.

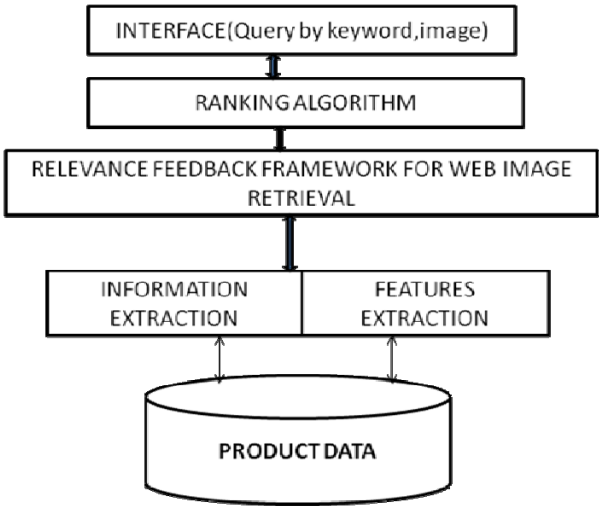

Figure-1. Image retrieval architecture.

Figure-1 describes the system architecture for an image search and recommendation. In product data is a database contains collection of hosted images with their details and also having indexing for each image. During user query, both information and feature is extracted from database and user giving feedback for an image has been extracted for retrieving the top images.

Relevance Feedback mainly for information retrieval [3] and main aim to improve the effectiveness of the information retrieval system and user guiding the system. While retrieval process, the user interacts with the system to provide rates to the relevance of the retrieved images, according to their subjective judgment. In addition to the information, the system dynamically learns the user's intention, and presents better results.

2. RELATED WORK

J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih [1] define a new image feature called the colour correlogram and use it for image indexing and comparison. Image Retrieval (IR) systems adopt the following two-step approach to search image databases: (1) indexing for each image in a database, a feature vector capturing certain important properties of the image is computed and stored in a feature base, and (2) searching based on a query image, its feature vector is calculated, compared to the feature vectors in the feature base, and also images most similar to the query image are retrieved to the user. The correlogram robustly tolerates large changes in appearance and shape caused by changes in viewing positions, camera zooms. In these features has focused on (1) It includes the spatial correlation of colours, (2) It can be used to describe the global distribution of local spatial correlation of colours, (3) The size of the feature is fairly small.

S. Chakrabarti, B. Dom, P. Raghavan, and S. Rajagopalan [6] introduced Automatic Resource Compilation (ARC) for automatically compiling a list of authoritative web resources on any topic. ARC is to compile resource list to operate fully automatically based on a combination of text and link analysis for authoritative Web resource. ARC obtains a faster coverage of the available resources and of the topic space than a human can achieve. Next proposed the algorithm has three phases: (1) a search and growth phase algorithm first gathers a collection of pages from among it will distil ones that it considers the best for the topic, (2) a weighting phase, the topic is sent to a term-based search engine and the particular root set returned by the search engine is determined by its own scoring function and,(3) an iteration-and-reporting phase has each iteration reflects the notion that good hub pages it contains a large number of links to pages containing information about the topic point to good authority pages contains a lot of information about the topic and describe them as being relevant to the topic text [Kleinberg, 97].

L. Page, S. Brin, R. Motwani, and T. Winograd. [7] proposed a novel approach of PageRank. PageRank is a method for rating Web pages objectively and effectively measuring the human interest for every web page based on the graph of the web. Applications of PageRank have in search, browsing, and traffic estimation. The World Wide Web (WWW) is hypertext and provides considerable auxiliary information on top of the text of the web pages, such as link structure and link text [WVS 96, SVE 97]. In the link structure of the Web has a global ranking of every web page. This is called PageRank helps search engines and users fastly make sense of the vast heterogeneity of the World Wide Web [GE, 96].

D.M. Squire, W. Muler, H. Muler, and T. Pun. [9] developed a techniques inspired by text retrieval research to content-based image retrieval. The text retrieval can be applied to the content-based query of image databases and RF. Next, the use of inverted files, coupled with an appropriate choice of discrete features such as set of color and texture features is proposed. It allows feature spaces of extremely high dimensionalities to be searched efficiently. The Relevance Feedback (RF) offers (1) First, augmenting the query with features from relevant images produces a better representation of the user desires (2) The second is unique to CBIR.

G. Jeh and J. Widom [10] introduced a new algorithm called SimRank that measuring similarity of objects arises in many applications, and many domain-specific measures. It works under object-to-object relationships for measuring similarity of the structural context in the objects, based on their relationships with other objects. A similarity measure can be used to cluster objects, in a recommender system of "similar" users and items are grouped based on the user preferences.

X. Jin, J. Luo, J. Yu, G. Wang, D. Joshi, and J. Han. [16] considered the problem of performing image retrieval in image-rich information network. The proposed algorithm is MoK-SimRank to significantly improve the speed of SimRank. In a computing node has similarity in information networks. Next, it propose an algorithm called Simlearn to extend Mok-SimRank to heterogeneous image-rich information network, and account for both link-based and content based similarities by seamlessly integrating reinforcement learning with feature learning.

In the survey mainly focused on retrieve the images as effective manner. Many techniques has discussed such as IDB, Indexing, CBIR, Page Rank and algorithm SimLearn, MoK-SimRank. Common drawback are (1) Storage, manage, retrieve large size of images difficulty, (2) Visual feature extraction such as shape, texture, color, face to retrieve the image as consider either global or local feature. So retrieving has not a effective because of single feature consideration, (3) Semantic based extraction is based on if semantic does not provide such as keywords, or descriptions of the images so that retrieval can be performed correctly, (4) Ranking has provided for web page based on PageRank or HIT Rank or Count based. In a page rank has infinity of page link are available, it difficult to provide rank, (5) SimRank does not work under image rich information networks (6) ARC for automatic compilation of ranking instead of manual but to provide clear semantic description about each topic. (7) SVM used for efficiently classified the texture in feedback.

3. MOTIVATION

Conducting information retrieval in a large image-rich information networks is a very useful but also very difficult task, because it contains more number of information like text, image feature, user, group, and most importantly the network structure. The objective of this paper is to extract the similarity between image rich social resources using content, relevance feedback and social annotation of online social media.

System design

A. Architecture

In a system design architecture Figure-2 describes the user posting images, user annotations and feedback for images are called social image rich database. Based on user query to extract information from social image rich database as feedback, annotations and feature extraction and calculate weight for each individual has to be combined by integration technique. Based on ranking of images to provide top images in an efficient way the link similarity depends upon the social network information like tags, groups and human annotation over the images [16]. Content based similarity considers the image content properties colour histogram, for measuring the similarity. In relevance feedback-based approaches, a CBIR (content based image retrieval) system learns from feedback provided by the user for efficient way to retrieve the image. The combined score of these measures is used to integrate the social resources and helps to classify the images in image-rich information networks.

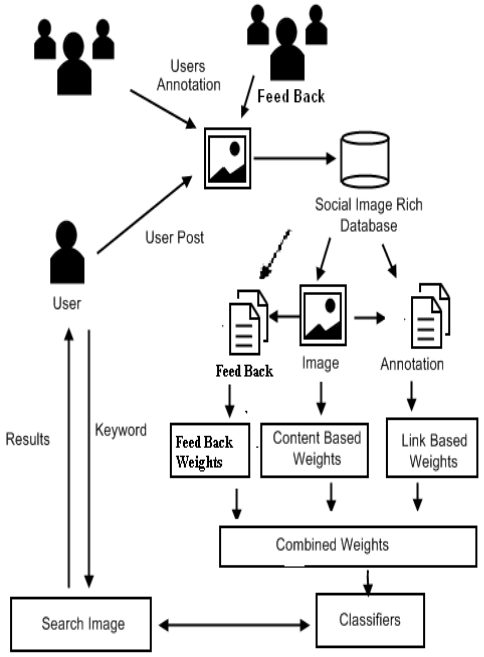

Figure-2. System design.

B. Modules

Social community development

This module is used to construct the basic social multimedia platform similar to the flicker through which the image and image related data are collected and pre-processes step to construct the required data model for our Similarity Integration. In Figure-3 for social community development it contains user registration and login process, image posting, sharing and human based annotation like tagging of images and image pre-processing step.

C. Hybrid algorithm

Link-based similarity

Similar images are likely to link to similar tags and groups, so we define the link-based semantic similarity between images as a combination of similarity of a group and similarity of tags. It is defined as follows This module iteratively calculate the similarity between image pairs, similarity between group pairs of images and similarity between tag pairs of image until the convergence is reached.

LINK-BASED SIMILARITY MEASURES
Input: social information
Step 1. Extract the social information such as annotation ā.
Step 2. Perform normalization from ā for keyword extraction.
Step 3. Stop word removal technique can be used to extract the keyword.
Step 4. Find keyword frequency fi
Step 5. Top k results from fi stored in database.
Output: weight for information

Content-based image similarity

The image vector information is extracted from the image content based on color and histogram and this vector information is used by the cosine similarity function to measure the similarity. Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of 0° is 1, and if it is less than 1 for any other angle. Because its a judgment of orientation and not magnitude: two vectors with the same orientation have a Cosine similarity of 1, two vectors at 90° has a similarity of 0, and two vectors diametrically opposed has a similarity of -1, independent of their magnitude.

CONTENT-BASED SIMILARITY MEASURES
Input: images
Step 1. Extract the feature of image I
Step 2. Extract image vector K, from images. This can be done by using RGB color histogram.
Step 3. Calculate histogram by using following equation 1
$$\sum \text{hist}(R,G,B) = \Theta \quad (1)$$
Step 4. Calculate similarity function for an image by using equation 2
$$\text{Cos}(\Theta) = 0, 1, -1 \quad (2)$$
Output: requested images.

Relevance feedback (RF)

RF based on user gives the feedback in the form of 'relevance judgments' expressed over the retrieval results [4]. The relevance judgments calculate the results

based on a three value assessment. These three values are relevant, irrelevant and don't care. Relevant means the image relevant to the user, non-relevant means the image is definitely not relevant, and don't cares mean the user does not say anything about the image. If the user giving feedback is relevant, then feedback loop come to an end otherwise it continues until the user get satisfied with their results.

RELEVANCE FEEDBACK MEASURES

Input: User Feedback

Step 1. Begin

Step 2. Obtain the initial retrieval results of CBIR by using following equation 3

$$\overrightarrow{Fopt} - \overrightarrow{Fini} + \frac{\propto}{NRel} \sum_{I \in Rel} \overrightarrow{Fi} - \frac{\beta}{Nnon - rel} \sum_{J \in Non-Rel} \overrightarrow{Fj} \quad (3)$$

Step 3. Repeat until user satisfaction or result remains same

Step 4. From user interaction, obtain the feedback from the users on prior results of weighting scheme by using the following equation 3 and 4. Feedback is in the form of relevant or irrelevant to request.

$$\overrightarrow{F^T} = (w1, \ldots, wL) \quad (4)$$

$$Wi = tf_{i.ln(N|ni)} \quad (5)$$

Step 5. If results found to be not satisfied

Step 6. Learn the system through a feedback algorithm and hence results are refined

Step 7. End repeat

Step 8. End

Output: Retrieve relevant images

In Equation (3) where $0 < \alpha < 1$. We let α be a constant and gradually increase its value from 0 to 0.9 with each value of α corresponds to an algorithm. In Equation (4) ω1 to ωv is the weight of individual textual features. In Equation (5) where tf_i is the total number of frequency of word, N is the number of images and n_i is the number of images that contains the textual features. The accuracy of the algorithms after 2-5 iterations of feedback is shown in Figure-4.

Similarities integration

To present novel algorithm to integrate link-based and content-based similarities: First performs HMok-SimRank for computing the link-based similarities and second perform feature learning considering the relevance feedback to update the feature weights, and then update the node similarities based on the new content similarity.

Formation of clustering

The final weight calculated after integration of link based similarity; relevance feedback and content similarity are used to classify the images based on their similarity. From this clustered information users can retrieve their needed images by using a keyword.

INTEGRATION ALGORITHM

Input: G, the image-rich information network, feedback.

Step 1. Construct kd-tree over the image features.
Step 2. Find top k similar candidates of each object.
Step 3. Initialize similarity scores
Step 4. Iterate {
Step 5. Calculate the link similarity for image pairs via HMok – SimRank.
Step 6. Perform feature learning to update for either global or local feature learning.
Step 7. Calculate the result from feedback.
Step 8. Search for new top k similar image candidates based on the new similarity weighting.
Step 9. Update the new image similarities global or local feature learning by using the following equation

Sm+1 (i, i')

Step 10. Compute link-based similarity for all group and tag pairs via Hmok-SimRank
Step 11. Compute new Feedback is in the form of relevant or irrelevant to request.
Step 12. } until converge or stop criteria satisfied.

Output: retrieve top ranked images.

4. EVALUATION AND RESULT

Effective way to retrieve the image is implemented through the JSP (Java Server Pages) is server side development technology and to store, update, retrieve the image in MySql Database. Tomcat server 6.0 for running web page in the browser. The Flicker dataset contains 1800 images for image retrieval and related metadata information, such as groups and tags. To compare Visual is for the visual similarity, Text based on textual similarity, Link based on HMok-SimRank which only uses the link similarity, IWSL_L for local feature weight learning, and IWSL_G for global feature weight learning, Hybrid for the combination of textual, visual, link and relevance feedback.

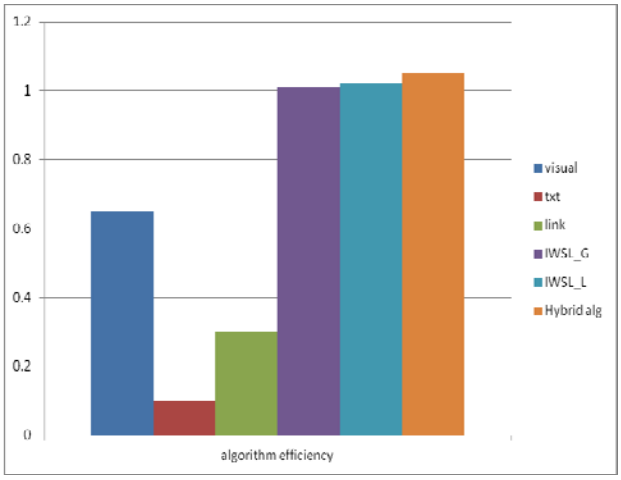

Figure-3. Map the algorithm for Flickr data. X-Axis denotes the Algorithm and Y-Axis denotes the efficiency.

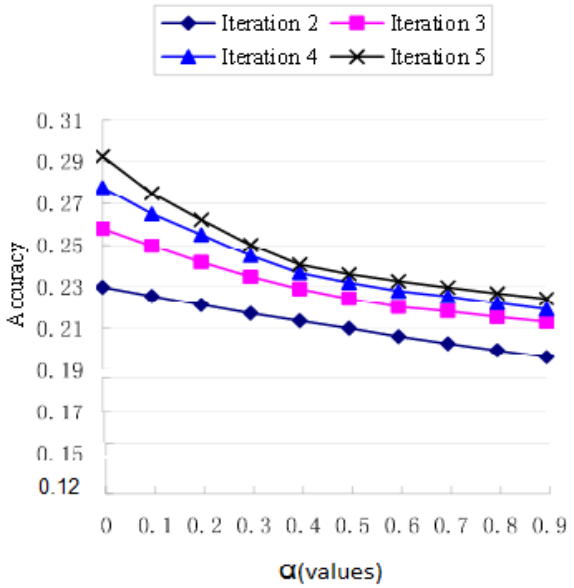

Figure-4. The accuracy of the algorithms after 2-5 iterations of Feedback with different α values.

For each iteration percentage of retrieval by using Rochhio algorithm is presented in the following Table-1.

Table-1. Each iteration % of retrieval in Rochhio Algorithm.

Method	I_1	I_2	I_3	I_4	I_5
Rochhio	73.6	92.7	97.3	99.2	99.4
Simple	73.6	83.2	88.0	91.0	92.9

Figure-5. Social community development.

In an edit profile Figure-6, for editing the personal details and also for searching the images**.** For uploading profile Figure-7. To upload of images give the details about tag, title, visible and description of images. In Figure-8, for retrieving the effective images based on user query on relevance feedback.

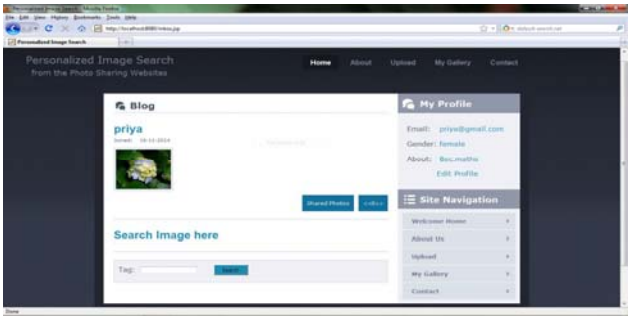

Figure-6. Edit profile.

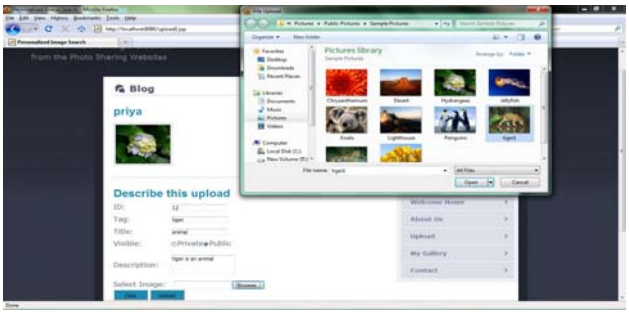

Figure-7. Upload image.

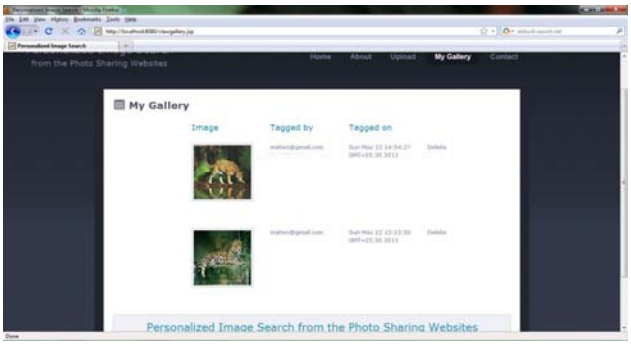

Figure-8. Image retrieval.

5. CONCLUSIONS

This paper presents a novel and efficient way of finding similar objects (such as photos and products) by modelling major social sharing and e-commerce websites as image rich information networks. The proposed novel approach to measures the similarity based on the link based, content based and relevance feedback. The combined score of these measures is used to integrate the social resources and helps to classify the images in an image-rich information networks. We propose Hmok-Simrank to efficiently compute weighted link-based similarity in weighted heterogeneous image-rich information networks and the relevance feedback have proven to increase retrieval accuracy and give more personalized results for the user. To implement a new search and recommendation system to find both visually similar and semantically relevant products based on our algorithm .In future work relevance feedback approach for

extracting semantic phrases with other techniques to comparing efficiency.

REFERENCES

[1] Flickr Image Search, [Online]. Available: https://www.flickr.Com.

[2] Picsearch Image Search, [Online]. Available: http://www.picsearch. Com.

[3] J. Rocchio. 1971. "Relevance Feedback in Information Retrieval," Upper Saddle River, NJ: Prentice-Hall.

[4] H. Tamura and N. Yokoya. 1984. "Image Database Systems: A Survey," Pattern Recognition. Vol. 17, no. 1, pp. 29-43.

[5] J. Huang, S.R. Kumar, M. Mitra, W.-J. Zhu and R. Zabih. 1997. "Image Indexing using Color Correlograms," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '97), pp. 762-768.

[6] S. Chakrabarti, B. Dom, P. Raghavan and S. Rajagopalan. 1997. "Automatic Resource Compilation by Analyzing Hyperlink Structure and Associated Text," Proc. Seventh World Wide Web Conf. (WWW '97), pp. 65-74.

[7] L. Page, S. Brin, R. Motwani and T. Winogra. 1999. "The Pagerank Citation Ranking: Bringing Order to The Web," technical report, Stanford InfoLab.

[8] Rui, Y., Huang, T. S. 1999. "A Novel Relevance Feedback Technique in Image Retrieval," ACM Multimedia.

[9] D.M. Squire, W. Muler, H. Muler and T. Pun. 2000. "Content-Based Query of Image Databases: Inspirations from Text Retrieval," Pattern Recognition Letters - Selected Papers from 11th Scandinavian Conf. mage, Vol. 21, Nos. 13/14, pp. 1193-1198.

[10] G. Jeh and J. Widom. 2002. "SimRank: A Measure of Structural-Context Similarity," Proc. Eighth Int'l Conf. Knowledge Discovery and Data Mining (KDD '02).

[11] X. S. Zhou and T. S. Huang. 2003. "Relevance feedback in Image Retrieval: A Comprehensive Review," ACM Multimedia Syst., Vol. 8, No. 6, pp.536–544.

[12] D. Fogaras and B. Racz. 2005. "Scaling Link-Based Similarity Search," Proc. 14th Int'l Conf. World Wide Web, pp. 641-650.

[13] T. Deselaers and H. Mller. 2008. "Combining Textual and Content-Based Image Retrieval, Tutorial," Proc. 19th Int'l Conf. Pattern Recognition (ICPR '08), http://thomas.deselaers.de/teaching/ files/tutorial_icpr08/04-combination.pdf.

[14] En Cheng, Feng Jing, and Lei Zhang "A Unified Relevance Feedback Framework for Web Image Retrieval," IEEE Transl. Image Processing.

[15] X. Jin, J. Luo, J. Yu, G. Wang, D. Joshi, and J. Han. 2010. "iRIN: Image Retrieval in Image-Rich Information Networks," Proc. 19th Int'l Conf. World Wide Web (WWW '10), pp. 1261-1264.

[16] Xin Jin, Jiebo Luo. 2013. "Reinforced Similarity Integration in Image-Rich Information Networks," IEEE Trans. Knowledge and Data Engineering, Vol. 25, No. 2, February.