

Negative Correlation Discovery for Big Multimedia Data Semantic Concept Mining and Retrieval

Yilin Yan, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33146, USA
Email: y.yan4@umiami.edu, shyu@miami.edu

Qiusha Zhu
Senzari, Inc.
601 Brickell Key Drive
Miami, FL 33131, USA
Email: q.zhu2@umiami.edu

Abstract—With massive amounts of data producing each day in almost every field, traditional data processing techniques have become more and more inadequate. However, the research of effectively managing and retrieving these big data is still under development. Multimedia high-level semantic concept mining and retrieval in big data is one of the most challenging research topics, which requires joint efforts from researchers in both big data mining and multimedia domains. In order to bridge the semantic gap between high-level concepts and low-level visual features, correlation discovery in semantic concept mining is worth exploring. Meanwhile, correlation discovery is a computationally intensive task in the sense that it requires a deep analysis of very large and growing repositories. This paper presents a novel system of discovering negative correlation for semantic concept mining and retrieval. It is designed to adapt to Hadoop MapReduce framework, which is further extended to utilize Spark, a more efficient and general cluster computing engine. The experimental results demonstrate the feasibility of utilizing big data technologies in negative correlation discovery.

Key words—Big Data, Negative Correlation, Hadoop, MapReduce, Spark, Multimedia Semantic Mining and Retrieval, Information Integration.

I. INTRODUCTION

In recent years, we have witnessed a deluge of big multimedia data such as texts, images, and videos. The data is big in terms of volume, variety, velocity, etc. For instance, Facebook ingests 500 terabytes of new data every day; while Walmart handles more than 1 million customer transactions every hour. Take Boeing 737 as another example. It generates 240 terabytes of flight data during a single flight across the US. During the flight, infrastructure and sensors generate massive log data in real-time. The big data are not just numbers, dates, and strings, but also the geospatial data, 3D data, audio data, video data, and unstructured texts (including log files and social media). Therefore, different approaches, techniques, tools and architectures are required for mining and retrieving knowledge, which is still in the development stage [1][2][3][4][5].

Despite the ubiquity of big multimedia data, effective management and retrieval of big multimedia data

are considered challenging research topics. For example, the conventional tag-based indexing and searching technologies suffer from the noisy and missing tag issue. As a result, more and more researchers turn to content-based approaches [6][7][8][9][10][11][12][13][14]. One main research task in the content-based big multimedia data retrieval field is multimedia concept mining and retrieval [15][16][17][18][19][20][21][22], which focus on mining semantic concepts such as face, car and airplane from raw videos directly.

Many approaches treat each concept as an individual class and convert one multi-concept detection problem into multiple binary classification problems [23]. Therefore, it ignores the correlations among different concepts. Nevertheless, the concepts are correlated in real-world multimedia data sets. For instance, some concepts co-occur more frequently, such as sky and cloud; while others rarely co-occur like road and fish. Such characteristics of correlations provide important context cues that can assist concept detection. Therefore, the calculation of correlations between concepts can help a lot in semantic concept mining and retrieval. There are many types of correlations, including Pearson correlation, Spearman correlation and Cross correlation, which detect the number of times two things occur together [24].

One major challenge in correlation discovery is the huge volume of related datasets. With the rapid development of multimedia, communication, and Web 2.0 technology, massive amounts of multimedia data have been increasingly available on desktops and smart mobile devices via the Internet. Statistics shows that 72 hours of videos with all sorts of tags are uploaded to YouTube every minute and about 1.54 million photos are uploaded to Flickr every day [25]. Accordingly, the annual TREC Video Retrieval (TRECVID) competition organized by National Institute of Standards and Technology (NIST) has the “Semantic Indexing” task for concept detection from a large amount of videos collected from Internet [26].

In this paper, we focus on a novel framework that

calculates negative concept correlations. In order to do it efficiently, two popular and widely used big data technologies, namely Hadoop and Spark, are adopted. The Apache HadoopTM software library is designed to scale up from a single server to thousands of machines, each offering local computation and storage. Rather than relying on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer. It delivers a highly-available service on top of a cluster of computers, each of which may be prone to failures. Apache SparkTM is a fast and general engine for large-scale data processing. Comparing to Hadoop, Spark is more advanced and contains many new features, and therefore has been deployed in many popular big data systems. Using these techniques, an efficient system is built to discover negative concept correlations from the multimedia big data.

In summary, the contribution of this paper are as follows. First, we propose a novel ICF (Integrated Correlation Factor) algorithm to estimate the negative correlation. Second, an efficient big multimedia data system using the ICF-based negative correlation discovery is built. Finally, ICF is utilized for multimedia concept mining and retrieval with the promising performance.

This paper is organized as follows. In Section II, we list and introduce several types of correlations and discuss their usages. In Section III, our proposed novel framework that includes negative correlation discovery and the ICF (Integrated Correlation Factor) algorithm are discussed. In Section IV, the proposed MapReduce algorithm on Hadoop is introduced. In Section V, an advanced system based on Spark is presented with its important modules explained in details. Section VI shows the comparison results of using the generated ICF values on the TRECVID 2015 dataset. Section VII draws the conclusion and identifies future research directions.

II. CORRELATION COEFFICIENT

Based on the category of the input data, correlations can be divided into the following four different types.

A. Nominal Scale

As the name implies, a nominal scale simply places the data into categories, without any order or structure. Nominal scales are used for labeling variables without any quantitative values. Nominal is like “name” while nominal scales could simply be called “labels”. A physical example of a nominal scale is the terms we use for colors. The underlying spectrum is ordered but the names are nominal. In research activities, a YES/NO scale is nominal. It has no order and there is no distance between YES and NO.

B. Ordinal Scale

With ordinal scales, it is the order of the values with importance and significance, so smaller (<) and bigger (>) can be applied but the differences between them is not really known. The simplest ordinal scale is a ranking. There

is no objective distance between any two points on the subjective scale. For one case, the top may be far superior to the second; but to another case, the distance may be subjectively small. An ordinal scale only lets you interpret a gross order and not the relative positional distances. Spearman’s rank correlation coefficient is a nonparametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of -1 or +1 occurs when each of the variables is a perfect monotone function of the other.

C. Interval Scale

Interval scales are numeric scales in which not only the order but also the exact differences between the values are known. The classic example of an interval scale is Celsius temperature. An interval scale is nice because the realm of statistical analysis on these data sets opens up. It is for numeric variables, plus (+) and minus (−) can also be applied. Pearson product-moment correlation (ρ or r) coefficient is a measure of the linear correlation dependence between two variables X and Y , giving a value between -1 and +1 inclusive, where -1 is a total negative correlation, 0 is no correlation, and +1 is a total positive correlation. It is widely used in sciences as a measure of the degree of the linear dependence between two variables.

D. Ratio Scale

Ratio scales are the ultimate nirvana when it comes to measurement scales because they tell us about the order, they tell us the exact value between units, and they also have an absolute zero which allows for a wide range of both descriptive and inferential statistics to be applied. A ratio scale provides a wealth of possibilities when it comes to statistical analysis. It is for numeric variables with an absolute zero, like temperature, mass, etc. Multiply (\times) and divid (\div) can also be applied. This kind of scales is not often available in multimedia and social research; while most correlation coefficient methods for interval scales can also be applied for ratio scales.

III. INTEGRATED CORRELATION FACTOR (ICF)

While positive correlations are widely used in semantic concept mining and retrieval, very few research approaches explore negative correlations to improve the performance. Some studies speculated that negative correlations captured by the negative graph might slightly improve the performance. Jiang et al. [27] studied the effect of negative correlations and conducted an experiment on the TRECVID data set showing the performance gain is merely 1.3% when using the negative graph alone.

Although the co-occurrence of two concepts in one video shot/image increases the probability that they are positively correlated, the fact that one concept does not occur when the other appears does not indicate they are

negatively correlated. For example, given an image that depicts a cat with no human face in the picture does not mean the appearance of a cat will exclude the appearance of a face. This makes it more challenging to discover negative correlations. To address this challenge, a two-step hierarchical selection strategy is proposed in this study.

First, a conditional probability-based coarse filtering approach is applied. The purpose of this step is to eliminate the irrelevant correlations in an efficient way. Specifically, for a target concept C_T , C_T^+ and C_T^- represent the events that a data instance is positive or negative for C_T . Likewise, for a reference concept, C_R^+ and C_R^- represent the events that a data instance is positive or negative for C_R . If C_T and C_R are negatively correlated, the following conditions must hold.

$$\frac{P(C_T^+|C_R^-)}{P(C_T^+)} > 1 \quad (1)$$

$$\frac{P(C_T^-|C_R^+)}{P(C_T^-)} > 1 \quad (2)$$

Here, $P(E)$ indicates the probability of event E . The threshold of 1, appeared in the above inequalities, is not selected arbitrarily. It is the necessary condition for the negative correlation, i.e., $P(C_T^+|C_R^-) > P(C_T^+)$ and $P(C_T^-|C_R^+) > P(C_T^-)$. The first inequality indicates that the probability of C_T appearing decreases if C_R appears. On the other hand, the second inequality indicates that the probability of C_T appearing increases if C_R does not appear. From the association rule point of view, these two values are related to the conviction measurement introduced in [28]. Actually, for each concept pair, we just need to get the result of Equation (1) since:

$$\frac{P(C_T^-|C_R^+)}{P(C_T^-)} = \frac{P(C_T^-C_R^+)}{P(C_R^+)P(C_T^-)} = \frac{P(C_R^+|C_T^-)}{P(C_R^+)} \quad (3)$$

Therefore, once we have the result of Equation (1) for all concepts, it is easy to find the value of Equation (2) in the result table (simply switching the target concept and related concept) and save a half of the computation time. Then we define a Negative Independent Coefficient (NIC) to measure the negative correlations between concepts as follows.

$$NIC(T, R) = \frac{P(C_T^+|C_R^-)}{P(C_T^+)} + \frac{P(C_T^-|C_R^+)}{P(C_T^-)} \quad (4)$$

All concept pairs with the NIC value less than a certain threshold will be filtered. This step eliminates a large number of possible correlations and reduces the computational complexity significantly.

The second step is to filter the selected concepts more rigorously. The reason of adding this step is based on the observation that a huge number of data instances are not labeled and are given the inferred label 0. This is common in a large multimedia dataset as manual labeling is very expensive. For example, the concept pair ‘‘Indoor’’ and

‘‘Outdoor’’ should show perfect negative correlations. However, in our experiment, there are 104054 out of 115806 instances with negative labels for both concepts. Therefore, the conditional probability $P(C_{Outdoor}^+|C_{Indoor}^-)$ is only 0.0786, which severely deviates from 1. The problem could not be solved by discarding those data instances simply as it will introduce the bias that the two concepts are negatively correlated. These kinds of bias are clearly not what we want since we wish to use correlation information to improve concept detection for all kinds of testing instances.

In order to tackle this difficulty, a novel strategy is proposed. The general assumption is that if two concepts are negatively correlated, their correlations would not be affected by the existence of the third concept, which is named as a control concept in this study. To formulate this problem, we define an integrated correlation factor (ICF) between the target concept and the reference concept, as formulated in Equation (5).

$$ICF(T, R) = \frac{1}{|\Omega| - 2} \sum_{D \in \Omega, D \neq T, D \neq R} \rho(C_T, C_R|C_D^+) \quad (5)$$

Here, Ω indicates the set of all concepts and $|\Omega|$ indicates the total number of concepts. C_D represents the control concept. C_D^+ is the condition that a data instance is positive for C_D , and $\rho(C_T, C_R|C_D^+)$ indicates the Pearson product-moment correlation coefficient [29] between the labels of C_T and C_R given C_D^+ . The reasons for adding the control concept C_D is that ICF represents an average quantitative metric of correlations under different conditions. For special cases where $\rho(C_T, C_R|C_D^+)$ is not defined, the default values are assigned as given in Table I. In this table, ‘‘All C_T^1 ’’ indicates that all data instances are positive for C_T and ‘‘All C_T^{-1} ’’ indicates that all data instances are negative for C_T . Correspondingly, ‘‘All C_R^1 ’’ and ‘‘All C_R^{-1} ’’ have similar meanings. All the special cases happen because the data instances have unique labels for either C_T or C_R , in which case the Pearson correlation coefficients are not defined. As shown in Table I, as long as C_T and C_R co-occur once, the value is set to a positive value, which imposes a relatively large penalty on that concept pair. Our empirical studies showed that when this value is 0.5, most of the significant negative concept pairs are captured.

After sorting all the ICF values in an ascending order, we observe that the combined correlation coefficients follow a quasi-normal distribution. Hence, a Gaussian probability density function is fit for all the ICF values. Different thresholds depending on the significance levels such as 95% and 67% could be set to select the significant negative correlations. As the Pearson product-moment correlation coefficients are symmetric, the concept pairs are selected. Therefore, the concept whose corresponding detector is less accurate is chosen as the target concept and the other one as the reference concept. The selection results are introduced in Section VI-B.

TABLE I
THE VALUES TO SET UNDER SPECIAL CONDITIONS

C_T	C_R	Value To Set
All C_T^1	All C_R^1	1
All C_T^{-1}	All C_R^{-1}	0
All C_T^1	All C_R^{-1}	Average value of negative Pearson correlation coefficients
All C_T^{-1}	All C_R^1	Same as above
Both C_T^{-1} and C_T^1 appear	All C_R^{-1}	Same as above
All C_T^1	Both C_R^1 and C_R^{-1} appear	Same as above
All C_T^1	Both C_R^1 and C_R^{-1} appear	0.5
Both C_T^{-1} and C_T^1	All C_R^1	0.5

IV. CORRELATION DISCOVERY BY HADOOP MAPREDUCE

MapReduce (MR) is a popular programming model introduced by Google [30], on which the distributed applications can be developed to efficiently process large amounts of data. Hadoop [31], its open-source implementation, provides a distributed file system called Hadoop Distributed File System (HDFS) and MR. Hadoop splits the files into large blocks and distributes them amongst the nodes in the cluster. To process the data, Hadoop MR transfers the packaged code for nodes to process in parallel, based on the data each node needs to process [32]. An MR program consists of two user-defined functions: a map function to process pieces of the input data called input splits, and a reduce function to aggregate the output of invocations of the map function. Both functions use user-defined key-value pairs as the input and output.

A collection of videos can be represented by Table II. Let S denote a shot in a video, C denote a concept, and v denote whether a concept appears in a shot. Value 1 represents the concept appears in the shot; while value -1 represents otherwise. Assume that there are totally M concepts and N shots in this video collection. To build the MR framework, we read the concept ID corresponding values from each video file and build the $\langle \text{Key}, \text{Value} \rangle$ pairs. Hence, each video file has several $\langle \text{Key}, \text{Value} \rangle$ pairs like:

$$\begin{aligned}
 & \text{Key} = (C_1, C_2), \text{ Value} = (v_1, v_2) \\
 & \text{Key} = (C_1, C_3), \text{ Value} = (v_1, v_3) \\
 & \dots \\
 & \text{Key} = (C_1, C_M), \text{ Value} = (v_1, v_M) \\
 & \text{Key} = (C_2, C_1), \text{ Value} = (v_2, v_1) \\
 & \dots \\
 & \text{Key} = (C_2, C_M), \text{ Value} = (v_2, v_M) \\
 & \dots \\
 & \text{Key} = (C_{M-1}, C_M), \text{ Value} = (v_{M-1}, v_M)
 \end{aligned}$$

The Map() function is used to generate the key-value pairs mentioned above; while the Reduce() function is used to calculate the conditional probability value in Equation (1) for each key. As discussed before, these two functions

TABLE II
A VIDEO DATASET

Shot ID	Concept ID	Value
S_1	C_1	v_{11}
...
S_1	C_M	v_{M1}
S_2	C_1	v_{12}
...
S_2	C_M	v_{M2}
...
S_N	C_1	v_{1N}
...
S_N	C_M	v_{MN}

use the defined key-value pairs as the input and output. Different from many symmetrical correlations like Pearson and Spearman correlations, conditional probability values are asymmetrical. Therefore, there are totally $N \times M \times M$ keys and $M \times M$ outputs as well.

However, since too many keys are generated, the computation gain is limited. It can be seen that the Mappers running time complexity is $O(NM^2)$ and the Reducers is $O(M^2)$. If multiple Reducers can be run in parallel, we should definitely see the speed improvement based on the scale. The time wasted on the mapping phase is much more than the time saved on the reducing part. In the next step, we can filter the outputs as mentioned in Section III to generate the ICF values for each concept pair.

V. CORRELATION DISCOVERY ON SPARK

In contrast to Hadoop's two-stage disk-based MapReduce paradigm, Spark's multi-stage in-memory primitives provide performance up to 100 times faster for certain applications [33][34][35]. Spark provides many operations called transformations like map, filter, flatMap, sample, groupByKey, reduceByKey, union, join, cogroup, mapValues, sort, and partitionBy; while Hadoop only provides two operations, namely map and reduce. Spark Core is the foundation of the overall project. It provides distributed task dispatching, scheduling, and basic I/O functionalities. The fundamental programming abstraction is called Resilient Distributed Datasets (RDDs), a logical collection of data partitioned across machines. RDDs can be created by referencing datasets in external storage systems or by applying coarse-grained transformations (e.g., map, filter,

TABLE III
CORRELATION MATRIX

	S_1	S_2	...	S_N
C_1	v_{11}	v_{12}	...	v_{1N}
C_2	v_{21}	v_{22}	...	v_{2N}
C_3	v_{31}	v_{32}	...	v_{3N}
...
C_M	v_{M1}	v_{M2}	...	v_{MN}

reduce, join) on the existing RDDs. The RDD abstraction is exposed through a language-integrated API in Java, Python, Scala, and R similar to local, in-process collections. This simplifies the programming complexity because the way that the applications manipulate RDDs is similar to how to manipulate the local collections of data.

Our Spark solution is presented as a single Java driver class that uses the rich Spark API to find the correlations of all versus all concepts. In order to build a better correlation discovery system using Spark, we first read the keys (shot ID) and values in Table II. After reading the data from all video files, we can group the values as an array or an iterator in Java using `combineByKey`. Table III shows the data representation.

$$Key = (C_1), Value = (v_{11}, v_{12}, \dots, v_{1N})$$

$$Key = (C_2), Value = (v_{21}, v_{22}, \dots, v_{2N})$$

$$Key = (C_3), Value = (v_{31}, v_{32}, \dots, v_{3N})$$

...

$$Key = (C_M), Value = (v_{M1}, v_{M2}, \dots, v_{MN})$$

To correlate all versus all concepts, we have to create the Cartesian product of groups by groups. This will generate all possible combinations of (C_i, C_j) . These combinations are the output values of the Cartesian products, which are the input keys for computing the correlations:

$$Key = (C_1, C_2), Value = (v_1[], v_2[])$$

$$Key = (C_1, C_3), Value = (v_1[], v_3[])$$

...

$$Key = (C_1, C_M), Value = (v_1[], v_M[])$$

$$Key = (C_2, C_1), Value = (v_2[], v_1[])$$

...

$$Key = (C_2, C_M), Value = (v_2[], v_M[])$$

...

$$Key = (C_{M-1}, C_M), Value = (v_{M-1}[], v_M[])$$

Now, we are ready to calculate the conditional probability values for all possible combinations of concepts. Afterward, we can filter the half of the key pairs when $i > j$ in (C_i, C_j) and generate the NIC values using Equation (4). The key pairs with the NIC values less than a threshold will be discarded. In the next step, the ICF values are generated. We create new key value pairs as follows. The keys have been filtered based on the NIC values between the concept pairs. The values here are the same. However, for different

keys, the target concept and related concept are different, as mentioned in Equation (5). For each key pair, the rest concepts would be the control concepts. The ICF values are then generated for each key pair and the final output is shown as follows.

$$Key = (C_1, C_2), Value = (ICF_{1,2})$$

...

$$Key = (C_1, C_M), Value = (ICF_{1,M})$$

$$Key = (C_2, C_3), Value = (ICF_{2,3})$$

...

$$Key = (C_2, C_M), Value = (ICF_{2,M})$$

...

$$Key = (C_{M-1}, C_M), Value = (ICF_{M-1,M})$$

VI. EXPERIMENTS AND RESULTS

A. Dataset

In the experiment, the IACC.1.A dataset is chosen from the TRECVID 2015 benchmark [36], whose semantic indexing (SIN) task aims to recognize the semantic concept contained within a video shot, which is an essential technology for retrieval, categorization, and other video exploitations. Here, the concepts refer to the high-level semantic objects such as a car, road, and tree. It has several challenges such as data imbalance, scalability, and semantic gap [37][38].

This data set contains 200 hours of videos with the durations between 10 seconds and 3.5 minutes. The labels of 130 concepts were given by the collaborative annotation organized by NIST. After data pre-processing, there are 144774 shots and one keyframe was extracted from each shot. The detection scores of all shots were downloaded from the DVMM Lab of Columbia University [39]. The list of concepts and detailed explanations can be found in [26]. In order to increase the number of ground truth in the negative association selection module, the TRECVID 2015 training labels are also utilized.

B. Negative Correlation Selection Results

For 130 concepts, there are totally 8385 pair-wise concept correlations. The top 10 negative correlations using the NIC-based selection and the ICF-based selection are shown in Table IV. For the NIC-based selection, the concept pairs are selected by adding the two probability ratios on the left side of Equations (1) and (2).

It should be pointed out that some negative correlations are caused by the definitions of the concepts. For example, the ‘‘Two people’’ concept indicates that there must be exactly two people in the video shot so the ‘‘Single_Person’’ concept does not occur. The ‘‘Building’’ concept means the shots of an exterior of a building so that it has a negative correlation with the ‘‘Indoor’’ concept. The full explanations of all concepts can be found in [26]. As

TABLE IV
COMPARISON OF NEGATIVE CORRELATION SELECTION

Rank	NIC-based	ICF-based
1	Entertainment, Building	Road, Waterscape_Waterfront
2	Infants, Industrial	Indoor, Plant
3	Person, Helicopter_Hovering	Indoor, Vegetation
4	Person, Natural-Disaster	Daytime_Outdoor, Indoor
5	Person, Airplane_Flying	Indoor, Outdoor
6	Canoe, Bus	Suburban, Indoor
7	Telephones, Swimming	Indoor, Building
8	Cats, Person	Trees, Indoor
9	Canoe, Car_Racing	Male_Person, Female_Human_Face_Closeup
10	Person, Birds	Two_People, Single_Person

introduced in Section III, we used the average value minus one standard deviation as the threshold, which corresponds to 0.67 significance level, to select the negative concept pairs. The selected correlations are shown in the right column of Table IV. There are totally seven target concepts in the top 10 ICF-based negative correlations, which are “Road”, “Indoor”, “Daytime_Outdoor”, “Suburban”, “Trees”, “Male_Person”, and “Two_People”. These concepts are common and appear in many papers [40][41][42].

We obtain the shot IDs and the corresponding values from all the training video files and use the Map() function to build the key-value pairs (shot ID, value). Next, we use the combine function in Spark to combine their pairs by the shot IDs and get a table similar to Table III. This table is considered as the input for the next layer. In the next layer, the output key is the key pairs of the Cartesian products of all shots; while the output values are the pairwise vectors that contain values from each shot. The conditional probability values are then calculated to generate the NIC values by the vector pairs. As mentioned in Sections III and V, some key pairs are discarded due to the low NIC values and the rest are kept to generate the ICF values. These steps are shown in Fig. 1.

A framework of semantic concept mining and retrieval was proposed in [43]. Its training section consists of the “Multimedia Concept Mining” subcomponent and the “Concept Mining Enhancement” subcomponent. The basic idea of the “Multimedia Concept Mining” subcomponent is as follows. For N data instances (e.g., images/video shots) and M concepts, M concept detection models are trained such that for each instance, the model m outputs a score measuring the likelihood that concept m exists in that data instance. A more detailed discussion of this subcomponent can be found in [43]. In this paper, an enhanced “Concept Mining Enhancement” subcomponent within the multimedia big data infrastructure is proposed, which improves the overall semantic concept mining and retrieval framework. First, all the class labels are organized into a label matrix so that each row contains the labels of different concepts for one data instance. Next, significant negative correlations are selected using this label matrix in the negative correlation selection module. A set of features are extracted from the original training data set to train a

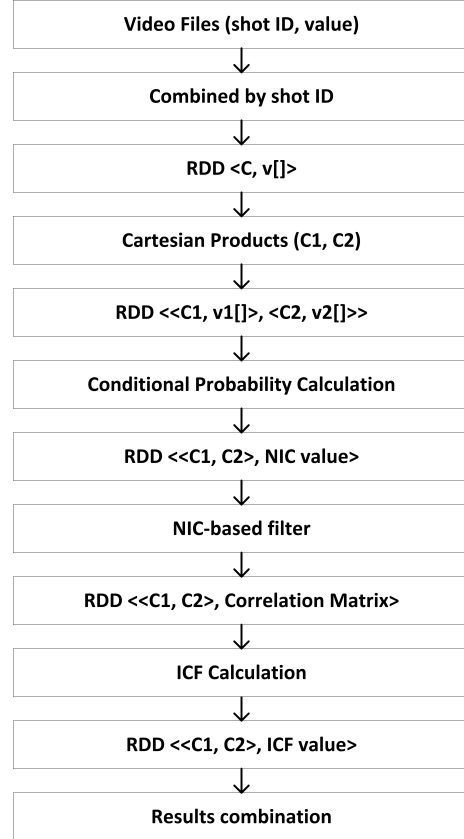


Fig. 1. Spark implantation of the ICF calculation

Multiple Correspondence Analysis (MCA) based negative weight estimation model [44][45][46][47]. The weights generated from this model together with the output scores from the “Multimedia Concept Mining” subcomponent are normalized and then used to train a regression-based score integration model. The selected negative correlations, MCA models, and regression models are stored for the testing section.

In the testing section, the testing data instances are plugged into all the concept detection models to generate the testing scores. The same features are extracted from that data instance to get the MCA-based weights. After the scores and weights are normalized, they are input to the

TABLE V
MAP VALUES AT DIFFERENT NUMBER OF INSTANCES RETRIEVED

Framework	MAP@10	MAP@20	MAP@30	MAP@40	MAP@50	MAP@100	MAP@200	MAP@500	MAP@2000
Base	0.45077	0.40841	0.37285	0.35755	0.33353	0.24407	0.15928	0.13049	0.16538
Subtraction	0.47292	0.39973	0.36313	0.33913	0.31965	0.22274	0.15056	0.11547	0.12814
Random	0.36005	0.31563	0.25743	0.23170	0.20680	0.13966	0.09827	0.08455	0.09293
DASD	0.48268	0.40204	0.36449	0.33404	0.32726	0.24311	0.15950	0.12220	0.13391
ICF-based	0.86262	0.73554	0.65211	0.60537	0.57122	0.47294	0.36569	0.33970	0.40617

regression-based score integration model to generate a new set of re-ranked scores. Finally, the new output scores are evaluated. The score integration module could be viewed as a two-layer neural network.

C. Performance of the Proposed Framework

In order to evaluate the proposed negative correlation discovery method, it is compared with the following four approaches. The first one, denoted as “Base”, has no modification on the raw scores. The second one, denoted as “Subtraction”, is an intuitive approach that subtracts the scores of a reference concept from those of a target concept. The third one, denoted as “Random”, randomly selects a reference concept. In the experiment, the randomly selected target concepts are: “Running” for “Road”; “Hand”, “Old_People”, and “Computer” for “Indoor”; “Beards” for “Daytime_Outdoor”; “Greeting” for “Suburban”; “Doorway” for “Trees”; “Chair” for “Male_Person”; and “Infants” for “Two_People”. The fourth one, denoted as “DASD”, applies the domain adaptive semantic diffusion (DASD) framework [48]. All negative affinities are kept as described in [48] and the number of iterations was set to 20. We compare our results with DASD since it also focuses on correlation discovery and achieves the best performance among many similar technologies.

The Mean Average Precision (MAP) values at different numbers of the retrieved data instances are reported. All results are the average of three-fold cross-validation over the seven ICF-based target concepts, appeared in Table IV. Table V shows the MAP comparison results. As can be seen, our proposed ICF-based method outperforms the four comparison methods by a large margin across all different MAP measures. “Random”, as expected, shows the worst performance. “DASD” obtains a slightly higher MAP@10 value compared to “Base” and “Subtraction”, but quickly drops after MAP@20. In addition, we found that a further increasing number of iterations did not improve the performance. ‘Subtraction’ shows a small gain at MAP@10 compared to “Raw”, but the values also decline when more retrieval results are considered. Moreover, the proposed ICF-based negative correlation discovery adapts to Hadoop and further to Spark, which makes it ready to scale-up to larger datasets, in terms of both the number of concepts and the number of data instances.

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel system of learning the negative correlations to provide valuable context cue

for concept mining and retrieval. The system adopts both Hadoop and Spark for efficient big data processing. Our proposed framework is compared to four approaches with the performance of semantic concept mining and retrieval. The experimental results demonstrate that negative correlations between the reference concepts and the target concepts discovered by the ICF-based method can effectively help retrieve the target concepts.

The proposed correlation discovery technique can be extended and used in many kinds of correlations and coefficients. The only difference is in the last steps of Fig. 1, meaning that different correlation algorithms can be applied to get the desired correlations in different circumstances. Furthermore, some correlations can be ternary or even quaternary. In such cases, the Cartesian part can be replaced to generate the corresponding groups for correlation discovery.

REFERENCES

- [1] D. Liu, Y. Yan, M.-L. Shyu, G. Zhao, and M. Chen, “Spatio-temporal analysis for human action detection and recognition in uncontrolled environments,” *Int. J. Multimed. Data Eng. Manag.*, vol. 6, no. 1, pp. 1–18, Jan. 2015.
- [2] Y. Yan, Y. Liu, M.-L. Shyu, and M. Chen, “Utilizing concept correlations for effective imbalanced data classification,” in *Proceedings of the IEEE 15th International Conference on Information Reuse and Integration*, Aug 2014, pp. 561–568.
- [3] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, “An effective content-based visual image retrieval system,” in *Proceedings of the Computer Software and Applications Conference*, 2002, pp. 914–919.
- [4] S.-C. Chen, S. Sista, M.-L. Shyu, and R. Kashyap, “Augmented transition networks as video browsing models for multimedia databases and multimedia information systems,” in *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, 1999, pp. 175–182.
- [5] S.-C. Chen, M.-L. Shyu, and R. Kashyap, “Augmented transition network as a semantic model for video data,” *International Journal of Networking and Information Systems*, vol. 3, no. 1, pp. 9–25, 2000.
- [6] L.-C. Chen, J.-W. Hsieh, Y. Yan, and D.-Y. Chen, “Vehicle make and model recognition using sparse representation and symmetrical {SURFs},” *Pattern Recognition*, vol. 48, no. 6, pp. 1979 – 1998, 2015.
- [7] C. Chen, Q. Zhu, L. Lin, and M.-L. Shyu, “Web media semantic concept retrieval via tag removal and model fusion,” *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 4, pp. 61:1–61:22, October 2013.
- [8] K.-T. Chuang, J.-W. Hsieh, and Y. Yan, “Modeling and recognizing action contexts in persons using sparse representation,” in *Advances in Intelligent Systems and Applications - Volume 2*, ser. Smart Innovation, Systems and Technologies, J.-S. Pan, C.-N. Yang, and C.-C. Lin, Eds. Springer Berlin Heidelberg, 2013, vol. 21, pp. 531–541.
- [9] L. Lin and M.-L. Shyu, “Weighted association rule mining for video semantic detection,” *International Journal of Multimedia Data Engineering and Management*, vol. 1, no. 1, pp. 37–54, 2010.

- [10] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, "Feature selection using correlation and reliability based scoring metric for video semantic detection," in *Proceedings of the Fourth IEEE International Conference on Semantic Computing*, 2010, pp. 462–469.
- [11] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen, "Video semantic concept discovery using multimodal-based association classification," in *Proceedings of the IEEE International Conference on Multimedia & Expo*, July 2007, pp. 859–862.
- [12] X. Li, S.-C. Chen, M.-L. Shyu, and B. Furht, "Image retrieval by color, texture, and spatial information," in *Proceedings of the 8th International Conference on Distributed Multimedia Systems*, September 2002, pp. 152–159.
- [13] X. Huang, S.-C. Chen, M.-L. Shyu, and C. Zhang, "User concept pattern discovery using relevance feedback and multiple instance learning for content-based image retrieval," in *Proceedings of the Third International Workshop on Multimedia Data Mining, in conjunction with the 8th ACM International Conference on Knowledge Discovery & Data Mining*, July 2002, pp. 100–108.
- [14] M.-L. Shyu, S.-C. Chen, and R. Kashyap, "Generalized affinity-based association rule mining for multimedia database queries," *Knowledge and Information Systems (KAIS): An International Journal*, vol. 3, no. 3, pp. 319–337, August 2001.
- [15] D. Liu and M.-L. Shyu, "Semantic motion concept retrieval in non-static background utilizing spatial and temporal visual information," *International Journal of Semantic Computing*, vol. 7, p. 43:67, 2013.
- [16] Y. Yan, J.-W. Hsieh, H.-F. Chiang, S.-C. Cheng, and D.-Y. Chen, "Plsa-based sparse representation for object classification," in *Proceedings of the 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 1295–1300.
- [17] S.-C. Chen, S. Rubin, M.-L. Shyu, and C. Zhang, "A dynamic user concept pattern learning framework for content-based image retrieval," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 36, no. 6, pp. 772–783, Nov 2006.
- [18] M.-L. Shyu, C. Haruechaiyasak, S.-C. Chen, and N. Zhao, "Collaborative filtering by mining association rules from user access sequences," in *Proceedings of the International Workshop on Challenges in Web Information Retrieval and Integration*, April 2005, pp. 128–135.
- [19] T. Meng and M.-L. Shyu, "Leveraging concept association network for multimedia rare concept mining and retrieval," in *Proceedings of the IEEE International Conference on Multimedia & Expo*, July 2012, pp. 860–865.
- [20] S.-C. Chen, M.-L. Shyu, and C. Zhang, "Innovative shot boundary detection for video indexing," in *Video Data Management and Information Retrieval*, S. Deb, Ed. Idea Group Publishing, 2005, pp. 217–236.
- [21] S.-C. Chen, M.-L. Shyu, C. Zhang, and R. L. Kashyap, "Identifying overlapped objects for video indexing and modeling in multimedia database systems," *International Journal on Artificial Intelligence Tools*, vol. 10, no. 4, pp. 715–734, 2001.
- [22] S.-C. Chen, M.-L. Shyu, and C. Zhang, "An intelligent framework for spatio-temporal vehicle tracking," in *Proceedings of the 4th IEEE International Conference on Intelligent Transportation Systems*, August 2001, pp. 213–218.
- [23] E. A. Cherman, J. Metz, and M. C. Monard, "Incorporating label dependency into the binary relevance framework for multi-label classification," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1647–1655, February 2011.
- [24] S. Perera, *Instant MapReduce Patterns - Hadoop Essentials How-to*. Packt Publishing, May 2005.
- [25] E. Gabarron, L. Fernandez-Luque, M. Armayones, and A. Y. Lau, "Identifying measures used for assessing quality of youtube videos with patient health information: A review of current literature," *Interactive Journal of Medical Research*, vol. 2, no. 1, March 2013.
- [26] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVID," in *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, October 2006, pp. 321–330.
- [27] Y.-G. Jiang, J. Wang, S.-F. Chang, and C.-W. Ngo, "Domain adaptive semantic diffusion for large scale context-based video annotation," in *Proceedings of the 12th IEEE International Conference on Computer Vision*, Sept 2009, pp. 1420–1427.
- [28] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," in *Proceedings of the ACM SIGMOD international conference on management of data*, vol. 26, 1997, pp. 255–264.
- [29] K. Pearson, "Notes on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.
- [30] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, Jan. 2008. [Online]. Available: <http://doi.acm.org/10.1145/1327452.1327492>
- [31] Apache, "Hadoop," <http://hadoop.apache.org>, accessed Oct, 2015.
- [32] F. Fleites, S. Cocke, S.-C. Chen, and S. Hamid, "Efficiently integrating mapreduce-based computing into a hurricane loss projection model," in *Proceedings of the 14th IEEE International Conference on Information Reuse and Integration*, Aug 2013, pp. 402–407.
- [33] Apache, "Spark," <https://spark.apache.org>, accessed Oct, 2015.
- [34] U. Berkeley, "amplab," <https://amplab.cs.berkeley.edu>, accessed Oct, 2015.
- [35] R. S. Xin, J. Rosen, M. Zaharia, M. J. Franklin, S. Shenker, and I. Stoica, "Shark: Sql and rich analytics at scale," in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '13. New York, NY, USA: ACM, 2013, pp. 13–24. [Online]. Available: <http://doi.acm.org/10.1145/2463676.2465288>
- [36] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, G. Quenot, and R. Ordelman, "TRECVID 2015 – an overview of the goals, tasks, data, evaluation mechanisms and metrics," in *Proceedings of TRECVID 2015*. NIST, USA, 2015.
- [37] Q. Zhu, L. Lin, M.-L. Shyu, and D. Liu, "Utilizing context information to enhance content-based image classification," *International Journal of Multimedia Data Engineering and Management*, vol. 2, no. 3, pp. 34–51, 2011.
- [38] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE Multimedia*, vol. 18, no. 3, pp. 32–43, March 2011.
- [39] Y.-G. Jiang, "Prediction scores on TRECVID 2010 data set," <http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/>, 2010, last accessed on September 8, 2011. [Online]. Available: <http://www.ee.columbia.edu/ln/dvmm/CU-VIREO374/>
- [40] L.-C. Chen, J.-W. Hsieh, Y. Yan, and D.-Y. Chen, "Vehicle make and model recognition using sparse representation and symmetrical surfs," in *Proceedings of the 16th IEEE International Conference on Intelligent Transportation Systems*, Oct 2013, pp. 1143–1148.
- [41] J.-W. Hsieh, K.-T. Chuang, Y. Yan, and L.-C. Chen, "Sparse representation for recognizing object-to-object actions under occlusions," in *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service*, ser. ICIMCS '13. New York, NY, USA: ACM, 2013, pp. 117–120.
- [42] L.-C. Chen, J.-W. Hsieh, Y. Yan, and B.-Y. Wong, "Real-time vehicle make and model recognition from roads," in *Proceedings of the 12th International Conference on Information Technology and Applications in Outlying Islands*, May 2013, pp. 1033–1040.
- [43] T. Meng, Y. Liu, M.-L. Shyu, Y. Yan, and C.-M. Shu, "Enhancing multimedia semantic concept mining and retrieval by incorporating negative correlations," in *Proceedings of the IEEE International Conference on Semantic Computing*, June 2014, pp. 28–35.
- [44] Q. Zhu, Z. Li, H. Wang, Y. Yang, and M.-L. Shyu, "Multimodal sparse linear integration for content-based item recommendation," in *Proceedings of the IEEE International Symposium on Multimedia*, 2013, pp. 187–194.
- [45] Q. Zhu, M.-L. Shyu, and H. Wang, "Videotopic: Content-based video recommendation using a topic model," in *Proceedings of the IEEE International Symposium on Multimedia*, 2013, pp. 219–222.
- [46] Q. Zhu and M.-L. Shyu, "Sparse linear integration of content and context modalities for semantic concept retrieval," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 2, pp. 152–160, June 2015.
- [47] Q. Zhu, M.-L. Shyu, and H. Wang, "Videotopic: Modeling user interests for content-based video recommendation," *International Journal of Multimedia Data Engineering and Management (IJM-DEM)*, vol. 5, no. 4, pp. 1–21, October 2014.
- [48] Y.-G. Jiang, J. Wang, S.-F. Chang, and C.-W. Ngo, "Domain adaptive semantic diffusion for large scale context-based video annotation," in *Proceedings of the International Conference on Computer Vision*, Kyoto, Japan, September 2009, pp. 1420–1427.