On The Least Economical MapReduce Sets for Summarization Expressions

Ravi (Ravinder) Prakash G, PhD
Shri Dharmasthala Manjunatheswara College of
Engineering and Technology
Dhavalagiri, Dharwad 580 002

Kiran M
Data Analytics Research Laboratory
Bangalore 560 010

ABSTRACT

An intention of MapReduce Sets for Summarization expressions analysis has to suggest criteria how summarization expressions in summarization data can be defined in a meaningful way and how they should be compared. Similitude based MapReduce Sets for summarization Expression Analysis and MapReduce Sets for Assignment is expected to adhere to fundamental principles of the scientific summarization process that are expressiveness of summarization models and reproducibility of their summarization inference. Summarization expressions are assumed to be elements of a summarization expression space or Conjecture class and summarization data provide "information" which of these summarization expressions should be used to interpret the summarization data. An inference summarization algorithm constructs the mapping between summarization data and summarization expressions, in particular by a summarization cost minimization process. Fluctuations in the summarization data often limit the summarization precision, which we can achieve to uniquely identify a single summarization expression as interpretation of the summarization data. We advocate an information theoretic perspective on summarization expression analysis to resolve this dilemma where the tradeoff between summarization informativeness of statistical inference summarization and their summarization stability is mirrored in the information-theoretic summarization optimum of high summarization information rate and zero communication expression error. The inference summarization algorithm is considered as an outlier object summarization path, which naturally limits the resolution of the summarization expression space given the uncertainty of the summarization data.

Keywords

MapReduce, summarization expressions, kernel function.

1. INTRODUCTION

The field of general expression intention is to formalize the notion of expressions in precise mathematical terms [1]. Expressions are perceived as regular structures behind the summarization data sources, i.e., "the underlying deep regular structures are descriptions of the source, which are hidden via the sensing path" [3]. Expression Intention combines algebra, geometry and statistics to explain the nature of summarization data sources and, thereby, depicts a generative summarization modeling perspective on expression analysis. This philosophy argues for a distinct generative viewpoint to infer the probability distribution of the summarization data.

In many real-world situations the summarization data are represented or generated in a very high dimensional space and the information-processing task focuses on a low dimensional interpretation space. The analysis of visual summarization data like images or videos provides a very convincing example of this situation: intensity expressions that are sensed by a camera are mathematically represented as points in a space with $\#\{intensities\}^{\#\{pixels\}}$ dimensions. When segmenting an image in semantically distinct regions then the interpretation space contains $\#\{segments\}^{\#\{sites\}}$ elements where the number of sites is often much smaller than the number of pixels. The reader should note that the space of segmentations is still exponentially large in the number of sites. Consequently, we adopt a discriminative view of MapReduce Sets for Summarization Expression assignment: the summarization expressions which are inferred from the summarization data are elements of an interpretation space called Conjecture class and these summarization expressions are more or less closely related with the summarization data generating mechanism of the source. The Conjecture class often also reflects information about the aim of expression analysis, i.e., what the summarization expressions are used for in subsequent information processing.

The inference of summarization expressions in summarization data is formulated as an algorithmic search for a stable subset of the underlying Conjecture class. Stability is required to guarantee that the summarization expression analysis process would yield an equivalent outcome for the same structure of the summarization data source but a different realization of the outlier object process. A second, antagonistic requirement of the summarization expression analysis process is its specificity or informativeness: a small subset of the Conjecture class and in the limit a single Conjecture should be selected which poses a tradeoff to the stability requirement. Both summarization principles mirror the reproducibility and specificity requirement of the scientific reasoning [9].

Summarization expression analysis algorithms often follow an optimization principle. Desired summarization expressions are assigned a high score or low costs and undesirable summarization expressions are discarded by assigning a low score or high costs. In the following, we adopt the terminology of cost minimization rather than score maximization. A kernel function defines a partial order of Conjecture minimal kernels distinguish where the most preferred Conjecture. The outlier object in the summarization data, however, may introduce fluctuations in the kernels and the Conjecture with minimal kernels for one realization of the summarization data may no longer minimize kernels for a second realization of the summarization data. Therefore, we advocate stabilizing the set of kernel-minimal Conjecture by expanding it to a set of Conjecture with near-optimal kernels, also called approximation set. The size of such an approximation set is determined by information theoretic considerations. Conjecture in the approximation set is considered to be statistically indistinguishable.
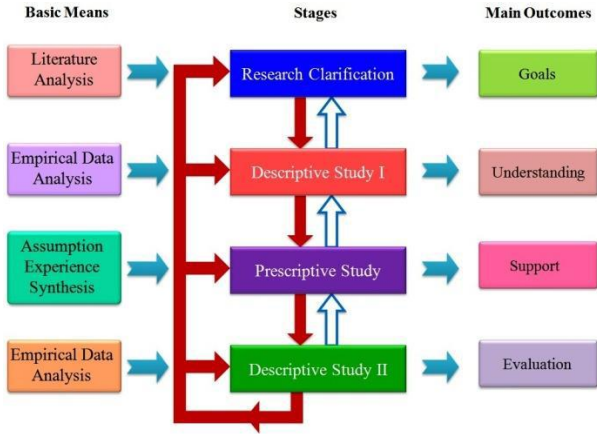
Fig 1: Research Plan: Basic Means, Stages, Main Outcomes

2. RESEARCH CLARIFICATION

2.1 Objects, measurements and Conjecture

Summarization expression analysis quantifies structures in summarization data, which usually relate to a set of objects. To mathematically characterize this problem domain we have to define what we mean by measurements and Conjecture. Given is a *set of objects* $\mathbf{O}^{(n)} = \{O_1,\ldots,O_n\} \subset \mathscr{O}, n \in \mathbb{N}$. Individual objects can be characterized by measurements either relative to an external reference frame, e.g., a coordinate system in a feature space, or by comparison to other objects. A measurement X is defined as a mapping of an object configuration in a measurement space, i.e.,

$$X : \mathscr{O}^1 \times \cdots \times \mathscr{O}^r \to \mathbb{K}, \qquad (O_1,\ldots,O_r) \mapsto X_{O_1,\ldots,O_r} . \quad (1)$$

The object configurations are often specified as collections of objects taken from the same object set $\mathscr{O}^1 = \cdots = \mathscr{O}^r$. For binary summarization data, the first and the second object set can differ $\mathscr{O}^1 \neq \mathscr{O}^2$. The most often used measurement types are feature vectors $X : \mathscr{O} \to \mathbb{R}^d$ denoted as $X_o \in \mathbb{R}^d$. Relational summarization data arise often in informatics applications and in network analysis problems. They are defined as $X : \mathscr{O} \times \mathscr{O} \to \mathbb{R}$, where $X_{o1,o2}$ denotes a proximity/Similitude value between object O_1 and O_2. More complicated summarization data structures than vectors or relations, e.g., three-way summarization data or (hyper) graphs, are employed in various applications.

In the following, we use the generic notation $\mathbf{X}^{(n)} \in \mathscr{X}^{(n)}$ for a *set of measurements* to characterize these n objects $\mathbf{O}^{(n)}$. $\mathscr{X}^{(n)}$ denotes the corresponding measurement space of n objects. To simplify notation we omit the index $^{(n)}$ whenever the dependence on problem size is clear.

A Conjecture c (.) of a MapReduce Sets for Summarization Expression assignment problem is a function that assigns a set of objects or a set of object configurations to a summarization expression out of a summarization expression space $\mathscr{P}$, i.e.,

$$c : \mathscr{O}^1 \times \cdots \times \mathscr{O}^r \to \mathscr{P}, \qquad (O_1,\ldots,O_r) \mapsto c(O_1,\ldots,O_r) . \quad (2)$$

The intention of Conjecture does not depend on the measurements $X_{O_1,\ldots,O_r}$ but potential summarization expressions that are denoted by Conjecture are defined *prior* to any measurements. The reader should note that the notion of a "feasible solution" in applied mathematics and optimization often depends on constraints that are determined by measurements contrary to the intention in (2). Such situations can be modeled by unconstraint solution spaces with infinite kernels for those solutions that violate the constraints.

The Conjecture class for a MapReduce Sets for Summarization Expression assignment problem is defined as the set of functions assigning an object or an object configuration (In the following, we restrict Conjecture to map an object to a summarization expression. The more general situation of object configurations can be analyzed in an analogous way but involves a more complex notation) to an element of the summarization expression space, i.e.,

$$\mathscr{C}(\mathbf{O}) = \{c\,(\mathbf{O}) : \mathbf{O} \in \mathscr{O}\}. \quad (3)$$

A well-known example of a Conjecture class is the space of partitions or classification functions $c: \mathscr{O} \to \{1,\ldots,k\}$ which we use in classification or clustering. When clustering n objects into k clusters, then we restrict the space of all possible partition functions to $\mathscr{P}^{(n)} = \{1,\ldots,k\}^n$ for the object set $\mathbf{O}^{(n)}$. The corresponding Conjecture class is denoted by $\mathscr{C}^{(n)} = \mathscr{C}(\mathbf{O}^{(n)})$. For parameter estimation problems the summarization expressions are possible values of the matrices and the summarization expression space is a subset of the d-dimensional Euclidean rotations.

2.2 Posteriori loss function approximation

The Conjecture class is a set of functions that map objects or object configurations to expressions. Summarization expression analysis requires assessing the quality of Conjecture $c \in \mathscr{C}$. We adopt a kernel function viewpoint in this paper, which attributes a non-negative kernel value

$$R : \mathscr{C}^{(n)} \times \mathscr{X}^{(n)} \to \mathbb{R}_+, \qquad (c, \mathbf{X}^{(n)}) \mapsto R(c, \mathbf{X}^{(n)}) \quad (4)$$

to each Conjecture given the measurements ($\mathbb{R}_+ := [0, \infty)$). The non-negativity assumption does not restrict the choice of kernel functions since we can always replace $\tilde{R}(c, \mathbf{X}^{(n)}) := R(c, \mathbf{X}^{(n)}) - \inf_{c \in \mathscr{C}} R(c, \mathbf{X}^{(n)})$ for effectively computable minimal kernels.

The classical intention of statistical Inference [2] advocates to use the posteriori minimize as the solution of the inference problem. The best posteriori summarization expression denoted by $c^\perp(\mathbf{X}^{(n)})$ minimizes the posteriori loss function of the summarization expression analysis problem given the measurements $\mathbf{X}^{(n)}$, i.e.,

$$c^\perp(\mathbf{X}^{(n)}) \in \arg\ \min_{c \in \mathscr{C}(n)}\ R(c, \mathbf{X}^{(n)}) . \quad (5)$$

Although Conjecture map objects into a summarization expression space, the posteriori loss function minimizer $c^\perp(\mathbf{X}^{(n)})$ depends on measurements.

The intention requires for inferability of classifications that the Conjecture class is not "too complex" (i.e., finite computing) and, as a consequence, the solution $c^\perp(\mathbf{X}^{(n)})$ converges to the optimal solution which minimizes the expected loss function. A corresponding criterion has been derived for regression [4].

This classical inference intention is not applicable when the size of the Conjecture class grows with the number of objects like in clustering or other optimization problems of combinatorial nature. Without strong regularization we cannot hope to identify a single solution, which globally minimizes the expected loss function in the asymptotic limit n → ∞. Conjecture class of combinatorial problems often have an infinite computing and, therefore, are not inferable in the classical computing sense. Therefore, we replace the concept of a unique function as the solution of an inference problem

with a weighted set of kernel functions. The challenge of inference then amounts to determine a weight measure, which is concentrated on few solutions to achieve precision. The weights w are defined as functions which map triplets of a Conjecture, measurements and a resolution parameter to the unit interval, i.e.,

$$w: \mathscr{C}^{(n)} \times \mathscr{X}^{(n)} \times \mathbb{R}_+ \to [0,1], \quad (c, \mathbf{X}^{(n)}, \beta) \mapsto w_\beta(c, \mathbf{X}^{(n)}) . \quad (6)$$

The set of weights is denoted as

$$\mathscr{W}_\beta(\mathbf{X}^{(n)}) = \{w_\beta(c, \mathbf{X}^{(n)}) : c \in \mathscr{C}^{(n)}\} .$$

How should we choose the weights $w_\beta(c, \mathbf{X}^{(n)})$ that large weights are only assigned to kernel functions with low kernels? The partial ordering constraint

$$R(c, \mathbf{X}^{(n)}) \leq R(\tilde{c}, \mathbf{X}^{(n)}) \Leftrightarrow w_\beta(c, \mathbf{X}^{(n)}) \geq w_\beta(\tilde{c}, \mathbf{X}^{(n)}) , \quad (7)$$

ensures that kernel functions with minimal kernels $R(c^\perp, \mathbf{X}^{(n)})$ assume the maximal weight value. Weights are normalized to one i.e., $0 \leq w_\beta(c, \mathbf{X}^{(n)}) \leq 1$. The non-negativity constraint of weights allows us to write the weights as $w_\beta(c, \mathbf{X}^{(n)}) = \exp(-\beta f(R(c, \mathbf{X}^{(n)})))$ with the monotonic kernel function $f(x)$. Since $f(x)$ amounts to a monotone rescaling of the kernels $R(c, \mathbf{X}^{(n)})$ we resort to the common choice of Statistical weights with the inverse computational value β, i.e.,

$$w_\beta(c, \mathbf{X}^{(n)}) = \exp(-\beta R(c, \mathbf{X}^{(n)})) . \quad (8)$$

It is worth mentioning that standard approximation sets as introduced in the intention of approximation algorithms would correspond to binary weights

$$w_\beta^{bin}(c, \mathbf{X}^{(n)}) = \begin{cases} 1 & \text{if } R(c, \mathbf{X}^{(n)}) \leq R(c^\perp, \mathbf{X}^{(n)}) + 1/\beta \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

The weight $w_\beta(c, \mathbf{X}^{(n)})$ of a given Conjecture c is a random variable of the measurements $\mathbf{X}^{(n)}$. We consider the quantity

$$Z_\beta(\mathbf{X}^{(n)}) := \sum_{c \in \mathscr{C}(n)} w_\beta(c, \mathbf{X}^{(n)}), \quad (10)$$

which measures the total weight of Conjecture with low kernels. The weight sum is also known as the partition function in statistically when we use Statistical weights. In case of binary weights $Z_\beta(\mathbf{X}^{(n)})$ denotes the number of solutions that are $1/\beta$ close to the optimum.

2.3 Generalization and the two instance scenario

To determine the optimal regularization of a MapReduce Sets for Summarization Expression assignment method we have to define and estimate the generalization performance of Conjecture. We adopt the two instance scenario with training and test summarization data described by respective object sets $\mathbf{O}'$, $\mathbf{O}''$ and corresponding measurements $\mathbf{X}'$, $\mathbf{X}'' \sim \mathbb{P}(\mathbf{X})$. Both sets of measurements are drawn Independent and identically distributed from the same probability distribution $\mathbb{P}(\mathbf{X})$. The training and test summarization data $\mathbf{X}'$, $\mathbf{X}''$ define two optimization problems $R(., \mathbf{X}')$, $R(., \mathbf{X}'')$. The two instance scenario or two sample set scenario is widely used in statistics and statistical Inference intention [5], i.e., to bound the deviation of posteriori loss function from expected loss function, but also for two-terminal systems in information intention [6].

Statistical expression analysis requires that inferred summarization expressions have to generalize from training summarization data to test summarization data since outlier object in the summarization data might render the solution

$c^\perp(\mathbf{X}') \neq c^\perp(\mathbf{X}'')$ unstable. How can we evaluate the generalization properties of solutions to a MapReduce Sets for Summarization Expression assignment problem? Before we can compute the kernels $R(., \mathbf{X}'')$ on test summarization data of approximate solutions $c(\mathbf{O}') \in \mathscr{C} (\mathbf{O}')$ on training summarization data we have to identify a summarization expression $c(\mathbf{O}'') \in \mathscr{C} (\mathbf{O}'')$ which corresponds to $c(\mathbf{O}')$. A priori, it is not clear how to compare summarization expressions $c(\mathbf{O}')$ for objects $\mathbf{O}'$ with summarization expressions $c(\mathbf{O}'')$ for objects $\mathbf{O}''$. Therefore, we define a bijective mapping

$$\psi : \mathscr{O}' \to \mathscr{O}'', \qquad \mathbf{O}' \mapsto \psi \circ \mathbf{O}' . \quad (11)$$

The mapping ψ allows us to identify a summarization expression Conjecture for training set of objects $c' \in \mathscr{C}(\mathbf{O}')$ with a summarization expression Conjecture for a test set of objects $c'' \in \mathscr{C}(\psi \circ \mathbf{O}')$. The reader should note that such a mapping ψ might change the object indices. In cases when the objects $\mathbf{O}'$, $\mathbf{O}''$ are elements of an underlying metric space, then a natural choice for ψ is the nearest neighbor mapping.

The mapping ψ enables us to evaluate summarization expression kernels on test summarization data $\mathbf{X}''$ for summarization expressions $c(\mathbf{O}')$ selected on the basis of training summarization data $\mathbf{X}'$. Consequently, we can determine how many training summarization expressions with large weights share also large weights on test summarization data, i.e.,

$$\Delta Z_\beta(\mathbf{X}', \mathbf{X}'') := \sum_{c \in \mathscr{C}(\mathbf{O}'')} w_\beta(c, \psi \circ \mathbf{X}') w_\beta(c, \mathbf{X}'') . \quad (12)$$

A large subset of Conjecture with jointly large weights indicates that low kernel Conjecture on training summarization data $\mathbf{X}'$ also perform with low kernels on test summarization data. The tradeoff between stability and informativeness for Statistical weights on (8) is controlled by maximizing β for given loss function $R(., \mathbf{X})$ under the constraint of large weight overlap $\Delta Z_\beta(\mathbf{X}', \mathbf{X}'') / Z_\beta(\mathbf{X}'') \approx 1$. A quantitative statement how close this ratio should approach unity requires a statistical decision intention as provided by information transmission.

2.4 Typicality of instances

A natural question in statistical inference arises from asymptotic considerations in the large n-limit. What is the asymptotic behavior of the log weight sum $\log Z_\beta(\mathbf{X}^{(n)})$ dependent on the problem/instance size n? As remarked above the measurements $\mathbf{X}^{(n)}$ of a particular MapReduce Sets for Summarization Expression assignment instance depend on the value n. In analogy to information intention [7] we assume that the log weight sums converge according to an asymptotic equipartition property, i.e.,

$$\mathscr{F}' := \lim_{n \to \infty} - \frac{\log Z_\beta(\mathbf{X}'^{(n)})}{\log |\mathscr{C}(\mathbf{O}'^{(n)})|} , \quad (13)$$

$$\mathscr{F}'' := \lim_{n \to \infty} - \frac{\log Z_\beta(\mathbf{X}''^{(n)})}{\log |\mathscr{C}(\mathbf{O}''^{(n)})|} , \quad (14)$$

$$\Delta \mathscr{F} := \lim_{n \to \infty} - \frac{\log \Delta Z_\beta(\mathbf{X}'^{(n)}, \mathbf{X}''^{(n)})}{\log |\mathscr{C}(\mathbf{O}''^{(n)})|} . \quad (15)$$

These assumptions (13, 14, 15) requires that the log weight sums normalized by the size of the Conjecture class converge towards deterministic limits. The quantities $\mathscr{F}'$, $\mathscr{F}''$ are known as the key values (up to a factor β^{-1}) for the instances $R(., \mathbf{X}')$, $R(., \mathbf{X}')$ in statistically. The factor $\log |\mathscr{C}^{(n)}|$ denotes

the problem size of the optimization problem, i.e., it is $O(n)$ for clustering problems with maximally k^n different partitions and $O(n\log n)$ for sorting problems with $\log |\mathscr{C}^{(n)}| = \log(n!)$.

Intention 1. The set $A_\varepsilon^{(n)}$ of jointly typical instances w.r.t. $p(\mathbf{X}'^{(n)},\mathbf{X}''^{(n)})$ is the set of instance pairs $(\mathbf{X}'^{(n)},\mathbf{X}'^{(n)}) \in \mathscr{X}^{(n)}\times\mathscr{X}^{(n)}$ with posteriori log partition functions close to the respective key values

$$A_\varepsilon^{(n)} = \{(\mathbf{X}'^{(n)},\mathbf{X}'^{(n)}) \in \mathscr{X}^{(n)}\times\mathscr{X}^{(n)} :$$

$$\left| - \frac{\log Z_\beta(\mathbf{X}'^{(n)})}{\log |\mathscr{C}(\mathbf{O}'^{(n)})|} - \mathscr{F}' \right| < \varepsilon , \qquad (16)$$

$$\left| - \frac{\log Z_\beta(\mathbf{X}''^{(n)})}{\log |\mathscr{C}(\mathbf{O}'^{(n)})|} - \mathscr{F}'' \right| < \varepsilon , \qquad (17)$$

$$\left| - \frac{\log \Delta Z_\beta(\mathbf{X}'^{(n)}, \mathbf{X}''^{(n)})}{\log |\mathscr{C}(\mathbf{O}'^{(n)})|} - \Delta\mathscr{F} \right| < \varepsilon\}. \qquad (18)$$

The reader should note that the weak law of large numbers guarantees convergence of posteriori entropies towards their expectation values in information intention. Due to the dependence of the weights $w_\beta(c, \mathbf{X}^{(n)})$ on the kernel function $R(.,\mathbf{X}^{(n)})$ convergence has to be required for a kernel function. We also conjecture that kernel functions that violate this convergence behavior cannot be used to define predictive summarization models.
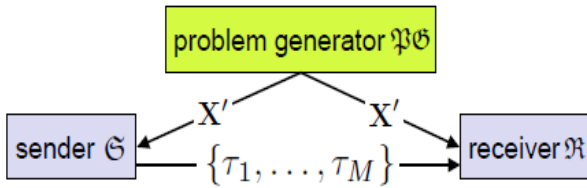

Fig 2: Generation of a set of *M* instruction problems by e.g. permuting the object indices.

3. DESCRIPTIVE STUDY I

In the following, we describe an information theoretic framework to determine which Conjecture are statistically indistinguishable due to outlier object in the measurements and how much we have to coarsen the Conjecture class. Random instruction concept suggests a summarization model intention to determine the maximal number of distinguishable n-key value strings in the Group of Hamiltonian space when the key value strings are exposed to outlier object in a communication path. We develop a generalization of this idea for solution spaces of optimization problems. The weight distribution $w_\beta(c, \mathbf{X}^{(n)})$, $c \in \mathscr{C}$ over the Conjecture class $\mathscr{C}$ corresponds to the subsets of key value strings assigned to a specific instruction vector in information intention. Outlier object perturbs the measurements and therefore, the weight distribution fluctuates. An algorithm to approximately minimize a kernel function and the measurements as input to this algorithm defines an outlier object path in an asymptotical communication scenario with a mapper M, a reducer R and a key generator KG. The key generator connects the mapper with the reducer by posing an optimization problem given a kernel function or an algorithm. Communication takes place by approximately optimizing a given kernel function, i.e., by calculating weight sets $Z_\beta(\mathbf{X}')$, $Z_\beta(\mathbf{X}'')$. This instruction concept will be referred to as approximation set instruction (ASI) since the weights are concentrated on approximate minimizers of the optimization problem. The outlier object path is characterized by a summarization expression kernel

function $R(c, \mathbf{X})$ that determines the path capacity of the ASI scenario. Selection and validation of MapReduce Sets for Summarization Expression assignment models are then achieved by maximizing the path capacity over a set of kernel functions $R_\theta (.,\mathbf{X})$, $\theta \in \Theta$ where θ indexes the various kernel functions or MapReduce Sets for Summarization Expression assignment objectives. In a more general setting an arbitrary algorithm which does not necessarily minimize a kernel function can be considered to define a weight distribution and thereby, to play the role of a outlier object path [8] due to fluctuations in the input or in the execution path.
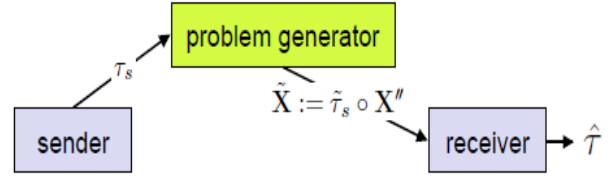

Fig 3: Communication process: (1) the mapper selects transformation τ_s, (2) the key generator draws $\mathbf{X}'' \sim \mathbb{P}(\mathbf{X})$ and applies $\tilde{\tau}_s = \psi \circ \tau_s \circ \psi^{-1}$ to it, and the reducer estimates $\hat{\tau}$ based on $\tilde{\mathbf{X}} = \tilde{\tau}_s \circ \mathbf{X}''$.

3.1 Instruction summarization by transformations

Before we describe the communication protocol we have to define the instruction for communication. We introduce random instruction intention to demonstrate the limits of asymptotically error free communication over an outlier object path. Random instruction refers to the fact that messages in random instruction summarization model are selected as a set of key value strings $\{\xi^{(j)} = (\xi_1^{(j)},\ldots,\xi_n^{(j)}), 1 \leq j \leq M\}$ with length $n = |\log \mathscr{C}(\mathbf{O}^{(n)})|$ that are drawn Independent and identically distributed according to a probability distribution $p(\xi)$. For sufficiently large n, the instructions all have mutual variations which are highly concentrated around the expected variation $2np(1 - p)$ with the probability $p = \mathbb{P}(\xi^{(1)} = 1)$. In the asymptotic limit $n \to \infty$ for $p = 1/2$, the random instructions uniformly partition the Group of Hamiltonian space of n-key value sequences into subsets of key value strings which can be reverse instruction without errors. In an analogous way, we cover the Conjecture class by weight distributions. To generate a uniform cover of the Conjecture class, we introduce a transformation

$$\tau : \mathscr{O} \to \mathscr{O}, \quad \mathbf{O} \mapsto \tau \circ \mathbf{O}. \qquad (19)$$

The set of all possible transformations is denoted as $\mathbb{T}$. Transformations that are restricted to object sets $\mathbf{O}^{(n)}$ of n objects are denoted by $\tau^{(n)} \in \mathbb{T}^{(n)}$. A random cover of the Conjecture class is then generated by selecting a set of transformations $\mathscr{T} = \{\tau_j^{(n)} \in \mathbb{T}^{(n)} : 1 \leq j \leq M, \tau_j^{(n)} \sim P(\tau^{(n)})\}$ with a rate $\rho := \log M / \log |\mathbb{T}^{(n)}|$. A natural choice of the probability distribution for transformations is the uniform distribution $P(\tau^{(n)}) = 1/|\mathbb{T}^{(n)}|$. The intuition behind the transformations is the following: When a transformation is applied to an object set $\mathbf{O}$ then the respective Conjecture $c(\mathbf{O})$ and the measurements $X_\mathbf{O}$ are transformed accordingly. Furthermore, the weights $w_\beta(c, \mathbf{X})$ are transformed by applying τ to c and $\mathbf{X}$, i.e., $\tau \circ w_\beta(c, \mathbf{X}) := w_\beta(\tau \circ c, \tau \circ \mathbf{X})$.

3.2 Typicality of transformations

Analogous to random instruction strategy, we generate the transformations $\tau^{(n)} \sim P(\tau^{(n)})$ in a random way. The probability distribution $P(\tau^{(n)})$ is defined over the set of possible

transformations $\mathbb{T}^{(n)}$. An asymptotic equipartition property depends on the entropy density of the transformation set

$$\mathscr{H}(\tau) := \lim_{n \to \infty} -\frac{\log P(\tau^{(n)})}{\log |\mathbb{T}^{(n)}|} \; . \qquad (20)$$

For instruction, we choose ε-typical transformations $\tau^{(n)} \in \mathbb{T}_\varepsilon^{(n)}$ with the typical set $\mathbb{T}_\varepsilon^{(n)}$ being defined in the following way:

Intention 2. The set $T_\varepsilon^{(n)}$ of typical transformations w.r.t. $P(\tau^{(n)})$ is the set of transformations $\tau^{(n)} \in \mathbb{T}^{(n)}$ with the property

$$T_\varepsilon^{(n)} = \{ \tau^{(n)} \in \mathbb{T}^{(n)} : \left| -\frac{\log P(\tau^{(n)})}{\log |\mathbb{T}^{(n)}|} - \mathscr{H}(\tau) \right| < \varepsilon \} \; . \qquad (21)$$

Special cases of such transformations $\tilde{\tau}^{(n)}$ are random permutations when optimizing combinatorial optimization kernel functions like clustering summarization models or graph cut problems. In parametric statistics, the transformations are parameter grids of e.g. rotations when estimating the orthogonal transformations of PCA or SVD.

3.3 Communication protocol

Mapper $\mathfrak{S}$ and reducer $\mathfrak{R}$ agree on a kernel function for MapReduce Sets for Summarization Expression assignment $R(c, \mathbf{X}')$ and on a mapping function ψ. The following procedure is then employed to generate the instruction for the communication process:

1. Mapper $\mathfrak{S}$ and reducer $\mathfrak{R}$ obtain summarization data $\mathbf{X}'$ from the key generator $\mathfrak{PG}$.
2. $\mathfrak{S}$ and $\mathfrak{R}$ calculate the weight set $\mathscr{W}_\beta(\mathbf{X}')$.
3. $\mathfrak{S}$ generates a set of (random) transformations $\mathscr{T} := \{\tau_1, \dots, \tau_{2n\rho}\}$. The transformations define a set of optimization problems $R(c, \tau_j \circ \mathbf{X}')$, $1 \le j \le 2^{n\rho}$ to determine weight sets $\mathscr{W}_\beta(\tau_j \circ \mathbf{X}')$, $1 \le j \le 2^{n\rho}$.
4. $\mathfrak{S}$ sends the set of transformations $\mathscr{T}$ to $\mathfrak{R}$ who determines the set of weight sets $\{ \mathscr{W}_\beta(\tau_j \circ \mathbf{X}') \}_{j=1}^{2n\rho}$.

The reason behind this procedure is the following: Given the measurements $\mathbf{X}'$ the mapper has randomly covered the Conjecture class $\mathscr{C}(\mathbf{O}')$ by respective weight sets $\{ \mathscr{W}_\beta(\tau_j \circ \mathbf{X}') : 1 \le j \le 2^{n\rho} \}$. Communication can take place if the weight sets are stable under the stochastic fluctuations of the measurements. The criterion for reliable communication is defined by the ability of the reducer to identify the transformation which has been selected by the mapper. After this setup procedure, both mapper and reducer have a list of weight sets available.

How is the communication between mapper and reducer organized? During communication, the following steps take place as depicted in fig. 3:

1. The mapper $\mathfrak{S}$ selects a transformation τ_s as message and sends it to the key generator $\mathfrak{PG}$.
2. $\mathfrak{PG}$ generates a new summarization data set $\mathbf{X}''$ and establishes correspondence ψ between $\mathbf{X}'$ and $\mathbf{X}''$. $\mathfrak{PG}$ then applies the selected transformation τ_s, yielding $\widetilde{\mathbf{X}} = \psi \circ \tau_s \circ \psi^{-1} \circ \mathbf{X}''$.
3. $\mathfrak{PG}$ sends $\widetilde{\mathbf{X}}$ to the reducer $\mathfrak{R}$ without revealing τ_s.
4. $\mathfrak{R}$ calculates the weight set $\mathscr{W}_\beta(\widetilde{\mathbf{X}})$.
5. $\mathfrak{R}$ estimates the selected transformation τ_s, by using reverse instruction rule

$$\hat{\tau} = \arg\max_{\tau \in \mathscr{T}} \Sigma_{c \in \mathscr{C}(\mathbf{O}'')} w_\beta(c, \psi \circ \tau \circ \mathbf{X}') w_\beta(c, \widetilde{\mathbf{X}}) \; . \qquad (22)$$

In the case of discrete Conjecture class, then the communication path is bounded from above by the cardinality of $\mathscr{C}(\mathbf{X})$ if two conditions hold: (i) the path is outlier object free $\mathbf{X}' \equiv \mathbf{X}''$; (ii) the transformation set is sufficiently rich that every Conjecture can be selected as a global minimizer of the kernel function.

4. PRESCRIPTIVE STUDY I

To determine the optimal approximation precision for an optimization problem $R(., \mathbf{X})$ we have to derive necessary and sufficient conditions which have to hold in order to reliably identify the transformations $\tau_s \in \mathscr{T}$. The parameter β, which controls the concentration of weights and thereby the resolution of the Conjecture class, has to be adapted to the size of the transformation set $|\mathscr{T}|$. Therefore, we analyze the error probability of the reverse instruction rule (22) which is associated with a particular kernel function $R(., \mathbf{X})$ a rate ρ. The maximal value of β under the condition of zero error communication is defined as *approximation capacity* since it determines the approximation precision of the instruction scheme.

A communication error occurs if the mapper selects τ_s and the reducer reverse instructions $\hat{\tau} = \tau_j$, $j \ne s$. To estimate the probability of this event, we introduce the weight overlaps

$$\Delta Z_\beta^j := \Sigma_{c \in \mathscr{C}(\mathbf{O}'')} w_\beta(c, \psi \circ \tau_j \circ \mathbf{X}') w_\beta(c, \widetilde{\mathbf{X}}), \tau_j \in \mathscr{T}. \qquad (23)$$

The quantity ΔZ_β^j measures the number of Conjecture, which have jointly low kernels $R(c, \psi \circ \tau_j \circ \mathbf{X}')$ and $R(c, \widetilde{\mathbf{X}})$.

The probability of a communication error is given by a substantial overlap ΔZ_β^j induced by $\tau_j \in \mathscr{T} \backslash \{\tau_s\}$, $1 \le j \le M$, $j \ne s$, i.e.,

$$\mathbb{P}(\hat{\tau} \ne \tau_s \mid \tau_s) = \mathbb{P}(\max_{1 \le j \le M, j \ne s} \Delta Z_\beta^j \ge \Delta Z_\beta^s \mid \tau_s)$$

$$\overset{(a)}{\le} \quad \Sigma_{1 \le j \le M, j \ne s} \mathbb{P}(\Delta Z_\beta^j \ge \Delta Z_\beta^s \mid \tau_s)$$

$$\overset{(b)}{\le} \quad \Sigma_{1 \le j \le M, j \ne s} \mathbb{E}_{\mathbf{X}', \mathbf{X}''} \mathbb{E}_{\mathscr{T} \backslash \{\tau_s\}} \left[\left. \frac{\Delta Z_\beta^j}{\Delta Z_\beta^s} \right| \tau_s \right]$$

$$\overset{(c)}{=} \quad (M-1) \mathbb{E}_{\mathbf{X}', \mathbf{X}''} \left[\left. \frac{\mathbb{E}_{\tau_j : j \ne s} [\Delta Z_\beta^j | \mathbf{X}', \mathbf{X}'']}{\Delta Z_\beta^s} \right| \tau_s \right] \qquad (24)$$

The expectation $\mathbb{E}_{\mathscr{T} \backslash \{\tau_s\}}$ is calculated w.r.t., the set of random transformations τ_j, $1 \le j \le M$, $j \ne s$ where we have conditioned on the mapper selected transformation τ_s. The joint probability distribution of all transformations $\mathbb{P}(\mathscr{T}) = \prod_{j=1}^{M} \mathbb{P}(\tau_j)$ decomposes into product form since all transformations are randomly drawn from the set of all possible transformations $\{\tau_j\}$. It corresponds to the random instruction summarization in information intention.

The inequality (a) results from the union bound and (b) is due to Markov's inequality. The identity (c) exploits the fact that the transformations $\tau \in \mathscr{T}$ are Independent and identically distributed drawn according to the product measure $\mathbb{P}(\mathscr{T})$.

The expected overlap $\mathbb{E}_{\tau_j} \Delta Z_\beta^j$, $j \ne s$ with any other message τ_j, $j \ne s$ for given training summarization data $\mathbf{X}'$ and test summarization data $\mathbf{X}''$ conditioned on τ_s is defined by

$$\mathbb{E}_{\tau_j : j \ne s} [\Delta Z_\beta^j | \mathbf{X}', \mathbf{X}''] = \Sigma_{\tau_j \in T} P(\tau_j)$$

$$\Sigma_{c \in \mathscr{C}(\mathbf{O}'')} w_\beta(c, \psi \circ \tau_j \circ \mathbf{X}') w_\beta(c, \widetilde{\mathbf{X}})$$

$$^{(d)}\leq \ \ \Sigma_{c\epsilon\mathscr{C}(\mathbf{O}'')} \ w_\beta(c, \widetilde{\mathbf{X}}) \ \exp(-\log|\mathbb{T}|(\mathscr{H}(\tau)-\varepsilon))$$

$$\underbrace{\Sigma_{\tau j \epsilon T} \ w_\beta(\tau_j^{-1}\circ\psi^{-1}\circ c, \mathbf{X}')}_{\leq Z_\beta(\mathbf{X}')}$$

$$^{(e)}\leq \ \ \exp(-\log|\mathbb{T}|(\mathscr{H}(\tau)-\varepsilon)) \ Z_\beta(\mathbf{X}'') \ Z_\beta(\mathbf{X}') \qquad (25)$$

The inequality (d) results from the typically of $P(\tau_j)$. The last inequality (e) holds since the set $\{\tau_j^{-1}\circ\psi^{-1}\circ c : c \ \epsilon\mathscr{C}\,(\mathbf{O}''),$ $\tau_j \in \mathbb{T}\} \subset \mathscr{C}(\mathbf{O}')$, and extending the sum $\Sigma_{\tau j \epsilon T}$ to $\Sigma_{c\epsilon\mathscr{C}(\mathbf{O}')}$ only adds positive terms. Effectively, the sum over a random transformation τ_j decouples the two sums and yields a product of weight sums. Inserting result (25) into equation (24) yields

$$\mathbb{P}(\hat{\tau}\neq\tau_s \mid \tau_s)\leq (M-1) \exp(-\log|\mathbb{T}|(\mathscr{H}(\tau)-\varepsilon))$$

$$\mathbb{E}_{\mathbf{X}', \mathbf{X}''} \ Z_\beta(\mathbf{X}'') \ Z_\beta(\mathbf{X}') \,/\, \Delta Z_\beta{}^s \qquad (26)$$

$$^{(f)}\leq \exp(\log M - \log|\mathbb{T}|(\mathscr{H}(\tau)-\varepsilon) - (\mathscr{F}'-\varepsilon)\log|\mathscr{C}(\mathbf{O}')|$$

$$- (\mathscr{F}''-\varepsilon)\log|\mathscr{C}(\mathbf{O}'')| + (\Delta\mathscr{F}+\varepsilon)\log|\mathscr{C}(\mathbf{O}'')|)$$

$$= \exp(\log|\mathbb{T}|\,\rho - \log|\mathbb{T}|\mathscr{H}(\tau) - \log|\mathscr{C}'|\mathscr{F}' - \log|\mathscr{C}''|$$

$$(\mathscr{F}''-\Delta\mathscr{F}) + \varepsilon\,(\log|\mathbb{T}| + \log|\mathscr{C}'| + 2\log|\mathscr{C}''|)) , \qquad (27)$$

where we have introduced the rate intention $\rho = \log M/\log|\mathbb{T}|$. The term proportional to ε can be neglected since it becomes arbitrarily small in the limit $\lim_{n\to\infty}$ due to the assumed asymptotic equipartition property (13, 14, 15). Often, the assumption $|\mathscr{C}(\mathbf{O}')| = |\mathscr{C}(\mathbf{O}'')| = |\mathbb{T}|$ is justified and the bound (26) simplifies to

$$\mathbb{P}(\hat{\tau}\neq\tau_s \mid \tau_s)\leq \exp(-\log|\mathscr{C}|\,(\mathscr{I}-\rho-4\varepsilon)) \qquad (28)$$

$$\text{with } \mathscr{I}_\beta := \mathscr{H}(\tau) + \mathscr{F}' + \mathscr{F}'' - \Delta\mathscr{F} \qquad (29)$$

The quantity $\mathscr{I}_\beta$ plays the role of the mutual information in communication. Error free communication requires $\rho < \mathscr{I}_\beta$, i.e., the rate ρ should not exceed $\mathscr{H}(\tau) + \mathscr{F}' + \mathscr{F}'' - \Delta\mathscr{F}$.

How can this upper bound (28) with quantity $\mathscr{I}_\beta$ be interpreted? A close look at equation (26) reveals that the bound depends on the term

$$\mathbb{E}_{\mathbf{X}', \mathbf{X}''} \ \frac{Z_\beta(\mathbf{X}'')\,Z_\beta(\mathbf{X}')}{|\mathbb{T}|\Delta Z_\beta{}^s} \ = \ \mathbb{E}_{\mathbf{X}', \mathbf{X}''} \ \exp(-\log|\mathbb{T}|/Z'_\beta$$

$$- \log|\mathscr{C}''|/Z''_\beta + \log|\mathscr{C}''|/Z''_\beta) \qquad (30)$$

with the abbreviation $Z'_\beta = Z_\beta(\mathbf{X}')$, $Z''_\beta = Z_\beta(\mathbf{X}'')$. The term $\log(|\mathbb{T}|/Z'_\beta)$ counts the number of ways you can form statistically distinguishable subset of the complete transformation class $\mathbb{T}$, the second term $\log(|\mathscr{C}''|/Z''_\beta)$ measure the same property on the reducer side and the last term $\log(|\mathscr{C}''|/Z''_\beta)$ accounts for double counting of the overlap. The three terms together define mutual information between the selected message τ_s and the reconstructed message $\hat{\tau}$.

5. PRESCRIPTIVE STUDY II
The analysis of the error probability suggests the following inference principle for controlling the appropriate strengths which implements a form of summarization model selection: the approximation precision is controlled by β which has to be

maximized to derive more precise solutions or summarization expressions. For small β the rate ρ will be low since we resolve the space of solutions only in a coarse grained fashion. For too large β the error probability does not vanish which indicates confusions between $\tau_j, j \neq s$ and τ_s. The optimal β-value is given by the largest β or, equivalently the highest approximation precision

$$\beta^* = \arg \ \max_{\beta\epsilon[0,\infty)} \ \mathscr{I}_\beta(\tau_s, \hat{\tau}) . \qquad (31)$$

Another choice to be made in modeling is to select a suitable kernel function $R(.,\mathbf{X})$ for the MapReduce Sets for Summarization Expression assignment problems at hand. Let us assume that a number of kernel functions $\{R_\theta(.,\mathbf{X}), \theta \in \Theta\}$ are considered as candidates. The approximation capacity $\mathscr{I}_\beta(\tau_s, \hat{\tau})$ depends on the kernel function through the weights. Therefore, we can rank the different summarization models according to their $\mathscr{I}_\beta(\tau_s, \hat{\tau})$ values. Robust and informative kernel functions yield a higher approximation capacity than simpler or more brittle summarization models. A quantifiable choice is to select the kernel function

$$R^*(c, \mathbf{X}) = \arg \ \max_{\theta\epsilon\Theta} \ \mathscr{I}_\beta(\tau_s, \hat{\tau}|R_\theta) . \qquad (32)$$

Where both the random variables τ_s and $\hat{\tau}$ depend on $R_\theta(c, \mathbf{X})$, $\theta \in \Theta$. The selection rule (32) prefers the summarization model which is "expressive" enough to exhibit high information content (e.g., many clusters in clustering) and, at the same time robustly resists to outlier object in the summarization data set. The key values, which are measured in the communication setting, are context sensitive since they refer to a Conjecture class $\mathscr{C}(\mathbf{X})$, i.e., how finely or coarsely functions can be resolved in $\mathscr{C}$.

6. DESCRIPTIVE STUDY II
To demonstrate the approach to regularized optimization we will apply it to an almost trivial optimization problem, i.e., minimizing the Group of Hamiltonian variation to a reference key value string $\xi' = (\xi'_1, \xi'_2,\ldots, \xi'_n) \in \{-1, 1\}^n$ of n key values. This optimization problem describes the reverse instruction step in classical communication intention. The kernel function for communication measures the difference between a key value string $s \in \{-1, 1\}^n$ and a reference instruction ξ', i.e.,

$$R(s, \xi') = \sum_{i=1}^n \mathbb{I}\{s_i\neq\xi'_1\} = \frac{1}{2}(n - \sum_{i=1}^n s_i\xi'_i) . \qquad (33)$$

The variable s has to be optimized and the posteriori minimum is $s = \xi'$. However, ξ' is exposed to path outlier object and, in the spirit of approximation set instruction, we should only approximate it. The weights of approximate solutions are defined by

$$\mathscr{W}_\beta(\xi') = \{ w_\beta(s, \xi') = \exp(-\frac{\beta}{2}(n - \Sigma_{1\leq i\leq n} s_i\xi'_i)) \} . \qquad (34)$$

The mapper uses this measurement ξ' and permutes the key values according to one of the randomly selected transformations $\mathscr{T} := \{\tau_1,\ldots, \tau_{2M}\}$. Permutations which leave ξ' invariant are excluded. This set of randomly selected transformations generates a instruction with instruction vectors $\{\tau_1\circ\xi',\ldots, \tau_{2M}\circ\xi'\}$.

During communication, a second key value string ξ'' is generated by the key generator. The reducer then receives the message $\tilde{\xi} = \tau_s\circ\xi''$ when the mapper decides to communicate with transformation τ_s. This process defines the approximation problem $R(s, \tilde{\xi}) = \frac{1}{2}(n - s\cdot\tilde{\xi})$ on the reducer side. Based on the summarization data $\tilde{\xi}$ reducer has to estimate the

transformation τ_s which has been communicated by the mapper.

Let us assume that the probability $\delta := \mathbb{P}(\xi'_i \neq \xi''_i)$ characterizes the communication path. Therefore, a fraction $\hat{\delta}$ n key values are different between the first key value sequence ξ' and the second key value sequence ξ''_i, i.e., $\hat{\delta} = \frac{1}{n} | \{i : \xi'_i \neq \xi''_i\}|$.

The weight sums $Z_\beta(\xi)$, $\xi \in \{\xi', \xi''\}$ are given by

$$Z_\beta(\xi) = \Sigma_{s\epsilon\mathscr{C}(\xi)} \exp(-\frac{\beta}{2}(n - \Sigma_{i\leq n} s_i\xi_i))$$

$$= \exp(-\beta\frac{n}{2}) \Pi_{i\leq n} \Sigma_{si \epsilon \{-1,1\}} \exp(\frac{\beta}{2}s_i\xi_i)$$

$$= \exp(-\frac{n\beta}{2}) 2^n (\cosh\frac{\beta}{2})^n. \qquad (35)$$

The number of jointly approximating key value strings is determined by

$$\Delta Z_\beta = \Sigma_{s\epsilon\mathscr{C}(\xi'')} \exp(-\beta(n - \frac{1}{2}\Sigma_{i\leq n} s_i(\xi'_i + \xi''_i)))$$

$$= \exp(-\beta n) \Pi_{i\leq n} (\exp(\frac{\beta}{2}(\xi' + \xi'') + \exp(-\frac{\beta}{2}(\xi' + \xi''))$$

$$= \exp(-\beta n) 2^n (\cosh\beta)^{n(1-\delta^\wedge)}. \qquad (36)$$

The mutual information (30) for the special case of minimizing Group of Hamiltonian variations is determined by

$$\mathscr{I}_\beta = \mathscr{H}(\tau) + \mathscr{F}' + \mathscr{F}'' - \Delta\mathscr{F}$$

$$= \ln 2 - \lim_{n\to\infty}\frac{1}{n}(\ln Z_\beta(\xi') + (\ln Z_\beta(\xi'') - \ln\Delta Z_\beta$$

$$= (1-\delta)\ln\cosh\beta - 2\ln\cosh\frac{\beta}{2}$$

$$= \ln 2 + (1-\delta)\ln\cosh\beta - \ln(\cosh\beta + 1). \qquad (37)$$

where we have estimated the size of the set of possible random transformations as $|\mathbb{T}| = 2^n$. In the case of a biased sequence with $\pi := \mathbb{P}(\xi'_i = 1) \neq \frac{1}{2}$ the cardinality of the transformation set is $|\mathbb{T}| = 2^{\mathscr{H}(n)}$ with the binary entropy $\mathscr{H}(\pi) = -\pi\log_2\pi - (1-\pi)\log_2(1-\pi)$.

The optimal value for β determined by the maximum of $\mathscr{I}_\beta$, i.e.,

$$\frac{dI_\beta}{d\beta} = (1-\delta)\frac{\sinh\beta}{\cosh\beta} - \frac{\sinh\beta}{\cosh\beta+1} = 0. \qquad (38)$$

$$\Rightarrow \quad \cosh\beta = \frac{1-\delta}{\delta}, \quad \cosh\beta + 1 = \delta^{-1}. \qquad (39)$$

Inserting these values into equation (37) yields

$$\mathscr{I}_\beta = \ln 2 + (1-\delta)\ln\frac{1-\delta}{\delta} - \ln\frac{1-\delta}{\delta} - \ln\frac{1}{\delta}$$

$$= \ln 2 + (1-\delta)\ln(1-\delta) - \delta\ln\delta$$

$$= \ln 2 - \mathscr{H}(\delta). \qquad (40)$$

Equation (40) shows that optimally approximating the Group of Hamiltonian variation of key value strings by approximation set instruction yields the path capacity of the binary symmetric path with key value error probability δ.

7. CONCLUSION

MapReduce Sets for Summarization expression analysis explores the questions how similar different summarization expressions are and how we should compare them. The underlying topology and metric of a Conjecture class are often chosen ad hoc in applications and usually do not derive from properties of the summarization data source. Approximation

set instruction as a summarization model validation principle establishes a notion of *summarization expression equivalence* by considering them as statistically indistinguishable when the summarization expression differences cannot be exploited for instruction. Summarization expressions with the same or similar weights are considered to be equally acceptable solutions and these weights directly depend on the objective or kernel function. To justify a natural topology and metric, we have to validate the underlying kernel function for the summarization expression analysis problem. The reader should realize that the assumption of a kernel function assumes a lot of information about the Conjecture class; it essentially establishes a partial order of Conjecture.

Summarization model selection and validation requires estimating the generalization ability of summarization models from training to test summarization data. "Good" summarization models show a high expressiveness and they are robust w.r.t., outlier object in the summarization data. This tradeoff between *informativeness* and *robustness* ranks different summarization models when they are tested on new summarization data and it quantitatively describes the underfitting/overfitting dilemma. In this paper we have explored the idea to use approximation sets of clustering solutions as a communication instruction. The *approximation capacity* of a kernel function provides a selection criterion, which renders various summarization models comparable in terms of their respective key value rates. The number of reliably extractable key values of a summarization expression analysis kernel function $R(.,\mathbf{X})$ defines a "task sensitive information measure" since it only accounts for the fluctuations in the summarization data $\mathbf{X}$ which actually have an influence on identifying an individual summarization expression or a set of summarization expressions.

The maximum entropy inference principle suggests that we should average over the statistically indistinguishable solutions in the optimal approximation set. Such a summarization model averaging strategy replaces the original kernel function with the value and, thereby, it defines a continuation method with maximal robustness. Algorithmically, maximum entropy inference can be implemented by annealing methods. The entropy naturally answers the question in many summarization data analysis applications, which regularization term should be used without introducing an unwanted bias. The second question, how the regularization parameter should be selected, is also answered by: Choose the parameter value that maximizes the approximation capacity! The link to robust optimization is analyzed from a theoretical computer science viewpoint.

For summarization model selection can be applied to all combinatorial or continuous optimization problems, which depend on outlier object summarization data. The outlier object level is characterized by two sample sets $\mathbf{X'}$, $\mathbf{X''}$. It has been posteriorly explored by summarization model validation problems for summarization model based clustering of high dimensional Gaussian distributed summarization data and of Boolean summarization data. The well-known spin glass phase of maximum likelihood estimations for Gaussian sources is identified as structure with zero information content for instruction. IT can also be used to select summarization models for spectral clustering. Furthermore, denoising of Boolean matrices guided by the generalization capacity of SVD suggests a cutoff rank for the SVD spectrum.

The reader should realize that we only require an objective or kernel $R(.,\mathbf{X})$ to define a weight distribution. Any other mechanism to arrive at such a concept of approximate

solutions will serve the same purpose. In principle, this concept of measuring the generalization performance of algorithms can be applied to algorithm evaluation and also to robust algorithm summarization. It endows the space of algorithm with a topology since two algorithms are neighbors if their approximation sets for the same input distributions share a high overlap. Such methods to measure the robustness of algorithms to errors in the computation or in the input will be in high demand to program novel hardware that trades consumption against precision of computation. So far we are completely lacking summarization principles for algorithm engineering, which consider this tradeoff between usage and correctness. We are also convinced that the information theoretic analysis of algorithms will shed new light on the relation between computational complexity and statistical complexity – the two faces of complexity science whose relation is far from being understood.

8. REFERENCES

[1] Ravi Prakash G, Kiran M, and Saikat Mukherjee, "Asymmetric Key-Value Split Pattern Assumption over MapReduce Behavioral Model" International Journal of Computer Applications (0975 – 8887) Volume 86 – No 10, Page 30-34, January 2014.

[2] Kiran M., Saikat Mukherjee and Ravi Prakash G., Characterization of Randomized Shuffle and Sort Quantifiability in MapReduce Model, International Journal of Computer Applications, 51-58, Volume 79, No. 5, October 2013.

[3] Amresh Kumar, Kiran M., Saikat Mukherjee and Ravi Prakash G., Verification and Validation of MapReduce Program model for Parallel K-Means algorithm on Hadoop Cluster, International Journal of Computer Applications, 48-55, Volume 72, No. 8, June 2013.

[4] Kiran M., Amresh Kumar, Saikat Mukherjee and Ravi Prakash G., Verification and Validation of MapReduce Program Model for Parallel Support Vector Machine Algorithm on Hadoop Cluster, International Journal of Computer Science Issues, 317-325, Vol. 10, Issue 3, No. 1, May 2013.

[5] Aniruddha Basak, Irina Brinster and Ole J. Mengshoel. MapReduce for Bayesian Network Parameter Learning using the EM Algorithm, Proc. of Big Learning: Algorithms, Systems and Tools, 1-6, December 2012.

[6] Berli'nska, J., Drozdowski, M.: Scheduling divisible MapReduce computations. J. Parallel Distrib. Comput 71(3), 450-459 (2011).

[7] Emanuel Vianna, Giovanni Comarela, Tatiana Pontes, Jussara Almeida, Virgilio Almeida, Kevin Wilkinson, Harumi Kuno, Umeshwar Dayal. Analytical Performance Models for MapReduce Workloads, Int J Parallel Prog 41:495-525 (2013).

[8] Erik B. Reed and Ole J. Mengshoel. Scaling Bayesian Network Parameter Learning with Expectation Maximization using MapReduce, Proc. of Big Learning: Algorithms, Systems and Tools, 1-5, December 2012.

[9] Ravi Prakash G, Kiran M and Saikat Mukherjee, On Randomized Preference Limitation Protocol for Quantifiable Shuffle and Sort Behavioral Implications in MapReduce Programming Model, Parallel and Cloud Computing, Vol. 3, Issue 1, Pages 1-14, January 2014.