Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE METHOD OF STOCHASTIC APPROACH ALGORITHM FOR PROBLEM SOLVING OF FEATURE SELECTION TECHNIQUE

MOCHAMAD WAHYUDI[1], MUHAMMAD ZARLIS[2], HERMAN MAWENGKANG[3], SYAHRIL EFENDI[4]

[1]Graduate Program of Computer Science, Universitas Sumatera Utara, Indonesia
[2,4]Department of Computer Science, Universitas Sumatera Utara, Indonesia
[3]Department of Mathematics, Universitas Sumatera Utara, Indonesia

E-mail: [1]mochamad.wahyudi@students.usu.ac.id, [2]m.zarlis@usu.ac.id, [3]mawengkang@usu.ac.id, [4]syahril1@usu.ac.id

ABSTRACT

The opinion from people can be adopted as important piece of information for most of the management during the decision-making process. The Internet and social media provide a major source of information about people's opinions. Due to the rapidly-growing number of online documents, it becomes both time-consuming and hard to obtain and analyze the desired opinionated information. The exploding growth in the Internet users is one of the main reasons that a method called sentiment analysis can help in extracting information about the opinion of people to classifies whether the opinion is positive or negative. One of the approaches in solving the sentiment analysis is feature selection method. However this technique contains a combinatorial behaviour and the analysis of the huge data can experience uncertainty parameter. This paper proposes a stochastic programming approach for solving the feature selection technique in order to obtain a decision from sentiment analysis

Keywords: *Sentiment Analysis, Feature Selection, Machine Learning, Stochastic Programming*

1. INTRODUCTION

Theoretically, sentiment analysis (SA) is an ordinary language processing task that categorizes the sentiments stated in review documents as "good" or "bad". Normally, SA is believed as a two-class classification problem. Though, some researchers advised a term called "neutral" for the third-class label. There are several studies about sentiment analysis that utilize diverse methods for data pre processing, feature selection, and sentiment grouping [1]–[3].The approach from statistics , namely, Chi square and information gain , can be used to remove unnecessary or irrelevant characteristics (which are always occur) such that the classification performance can be endorsed [4]. Supervised learning methods, such as, naïve Bayes, support vector machines, decision trees, and entropy modeling are exploited to categorize the sentiments of the reviews. In [5], suggest a new feature selection technique, called query expansion ranking.

Sentiment analysis belongs to natural language processing in pursuing the mood of the people about a product or topic. Sentiment analysis can be defined as a procedure to determine product reviews which turn up in the internet to conclude the overall opinion or feeling about a product [6], [7]. Users can express their opinion about the product online. Knowing consumer‘s opinion for a product is vital for the organization to take key choices. It is more valuable to the organization to foresee the consumer‘s view about the product. Duties in the sentiment analysis can be categorized into four levels according to the levels of granularity; document-level, aspect-level, phrase-level, and sentence-level sentiment analysis [8]–[10].

A model based on social media sentiment analysis and opinion mining is proposed by [11]. The study proposes a technique consisting of three parts of sentiment analysis, such as: pre-processing, feature extraction and classification. The pre-processing method is used to increase the accuracy of the system. The pre-processing technique used to obtain accurate results is an important factor in sentiment analysis. In this pre-processing technique, unigrams and bigrams are used for system personality skills.

Techniques based on a feature extraction and classification sentiment analysis were introduced by [12] who discussed techniques suggested by comparing existing systems in a study. In this study, in order to solve all the problems in the existing system, a machine learning classification is proposed to achieve a feature extraction model in order to make it more efficient and competent. The following categories are based on a feature-based sentiment analysis, such as: extraction, sentiment classification and sentiment evaluation.

Previously, sentiment analysis studies have been conducted using commonly used classification techniques, such as: Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) [13].

One of the problems in the classification of sentiment analysis in the form of a text is the number of attributes used in a dataset [14]. In general, the attributes of the text sentiment classification are very large, and if all of these attributes are used, the performance of the classifier will be reduced [15]. In order to be more accurate, the existing attributes must be selected with the correct algorithm [16].

Support vector machine is a supervised learning method that analyzes data and recognizes patterns used for classification [17]. The support vector machine has a deficiency in selecting appropriate parameters or features [17]. In other words, it can be said that the feature selection technique and the selection of the parameters in the support vector machine significantly affect the accuracy of the classification [18]. However, for some application problems, not all of these features are equally important and improved performance can be achieved by removing some features so that only features in the support vector machine have a significant impact on the accuracy of the classification [18].

Feature selection techniques are an important part of optimizing classifier performance [19]. Feature selection can be based on a large reduction in feature space, e.g. by eliminating attributes that are less relevant [20]. Feature selection techniques can make classifiers both more efficient and effective by reducing the amount of data analyzed and identifying appropriate features to be considered in the learning process [21]. There are two main types of feature selection techniques used in machine learning, namely: wrapper and filter [22]. The wrapper uses the accuracy of the classification of several algorithms as a function of its evaluation. Dataset that is not important, features that have many or very significant relationships will significantly reduce the level of accuracy of the classification by removing this feature, so that the level of efficiency and accuracy of the classification can be achieved [23].

There are several methods of optimization that can be used on the basis of a number of studies that have been performed to conduct sentiment analysis using feature selection techniques. [24] proposed a global optimization to solve the problem of the feature selection that they called the immune clonal genetic algorithm. [25] introduced a new evolutionary-incremental framework approach to addressing the challenge of feature selection. Recently, the optimization of the quadratic program was proposed by [26] to complete the feature selection with a non-linear pattern.

If the dataset has an initial feature *p*, then the feature selection technique problem has possible solutions $2^p - 1$, so this is a non-polynomial (NP-hard) problem. Therefore, a combinatorial stochastic optimization is the appropriate optimization method for optimally completing a sentiment analysis based on a feature selection technique. The combinatorial form of optimization is consistent with the technique of feature selection, while the stochastic condition is due to uncertainty conditions in sentiment analysis based on the technique of feature selection

2. LITERATURE REVIEW

SA is a significant topic in Natural Language Processing and Artificial Intelligence. Also known as opinion mining, SA scoops people’s opinions, sentiments, evaluations, and emotions about entities such as products, services, organizations, individuals, issues, and events, as well as their related attributes. This kind of analysis has many beneficial applications. For instance, it concludes a product’s popularity agreeing to the user’s reviews. If the overall sentiments are bad, additional analysis may be executed to recognize which features contribute to the negative ratings so companies can redesign their businesses. Many studies have been done for sentiment analysis indifferent domains, languages, and methods [7], [27]–[31]. Among these studies, the machine learning methods are more admired since the models can be automatically trained and enhanced with the training datasets. In [32] utilize supervised machine learning techniques such as NB and SVM to sentiment classification. NB, SVM, MEM, and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DT are some of the commonly used machine learning methods [32]–[35].

Feature selection techniques are applied to rank characteristics so that non-informative characteristics can be eliminated to develop the classification performance [36]. Some researchers have investigated the effects of feature selection for sentiment analysis [2], [24], [34], [35], [37], [38]. Feature selection (FS) is roughly described as choosing a subset of available features in a dataset that is related with the response variable by ignoring irrelevant and unnecessary features [39]. Presume the dataset has p original features, the FS problem has $2p - 1$ possible solutions and so it is an NP-hard problem because of an exponential growth in computational complexity with p. FS approaches can be categorized into four categories: wrapper, filter, embedded and hybrid methods. In [40] presented a novel feature selection method based on

a model of content and syntax that splits the reviewed entities and the opinion context (i.e., sentiment modifiers). The test executed with these features and using the maximum entropy classifier displays conclusions comparable to that of state-of art approaches. [35] proposed a method for developing sentiment analysis performance is a hybrid method of feature selection that mixes the methods, information gain and rough sets. In [41] devised feature selection method intelligently explores and employs the syntactic and semantic information and shows how a heterogeneous feature set fused with a suitable technique of feature selection could improve the sentiment analysis performance. In [42] has focused on comparative study of varied approaches and resources that are utilized for mining opinions. In spite of the level of complexities included, the crux of the study is about the evaluation metrics altered in terms of using the annotated quotations from varied news that are offered by EMS news gathering engine. The study determines that the generic opinion mining systems needs the large lexicons and the specialized training or test data which could influence the accuracy levels of the models. In [1] has reviewed model of methods and approaches which shall directly aid opinion-oriented information-seeking systems. The study focused on approaches which seek new challenges that are raised by the sentiment-aware applications, when matched to the ones that already have traditional models of fact-based analysis .In the process of review, some of the discussions associated to available resources, datasets that are of benchmark and the evaluation campaigns were also selected. The obtainability and popularity of varied opinion-rich resources like the online review sites and the personal blogs, emerging opportunities and the challenges that are creeping up from the extensive adaptation of ICT trends are used for knowing the opinions. The level of flow in the opinion mining and sentiment analysis manages with computation treatment of opinion, text subjectivity and sentiments. In [43] suggested model towards refining aspect-level opinion mining towards conducting online customer reviews. The new generative topic model of JAS (Joint Aspect/Sentiment) model was suggested to obtain aspects and the aspect dependent sentiment lexicons that could be derived from the online customer reviews.

3. FEATURE SELECTION TECHNIQUE

One of the most important machine learning tools is the classification which considered as the main tool of supervised machine learning. Classification with high accuracy is an important task for researchers. Feature selection is one of the effective ways to increase classification accuracy.

Feature selection is the process of selecting features that are relevant to the response variable of the subject of research under study. In general, feature selection techniques are divided into three kinds: filter, wrapper and embedded methods [22], [36]. Recently, intelligent techniques have widely used in features selection; the reason is due to the existence of many simple and easy of intelligent techniques that give affecting solutions [44]–[46].

Due to the inefficiency of traditional search approaches in solving complex combinatorial optimization problems, various metaheuristic have been proposed, such as genetic algorithms (GAs) [47], differential evolution (DE) [48], and particle swarm optimization (PSO) [49], among many others [50]. Among various metaheuristic, PSO is well known for its algorithmic simplicity and computational efficiency. Some researchers have proposed to apply PSO to feature selection [51], [52]. However, canonical PSO has many limitations for feature selection [53], [54]. Firstly, PSO was originally proposed for continuous optimization problems, while feature selection is a combinatorial optimization problem. Secondly, although PSO shows promising performance on low-dimensional problems, it suffers from the *curse of big dimensionality* [55]. Cheng and Jin have proposed a PSO variant, known as competitive swarm optimizer (CSO), for large-scale optimization [56].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In the study [57], the approach was designed to select the characteristics of the difference in the categorical probability ratio for the selection of features on the basis of their relevance and class-discriminatory characteristics. The technique of choosing a categorical probability ratio difference feature combines two different methods, a categorical proportional difference and a probability proportion difference. The categorical proportional different method is based on the calculation of the contribution level of a term that distinguishes classes and only the terms that contribute the most are used in classification [58]. The probability ratio difference technique is based on the calculation of the level of ownership or association or the probability that the term belongs to a certain class, and the difference is a measure of the ability to distinguish that class. The term of classification with a higher level of ownership is chosen as a candidate feature. The advantage with the categorical proportional difference technique is that it measures the class-differentiating properties which are the main attributes of the salient features. It may remove terms that are insignificant for classification, such as terms that occur equally across both classes, and terms with a high document frequency, such as stop words (For example, A, the, an, etc.). The advantage with the probability ratio difference technique is that it can remove low document frequency terms such as rare and unimportant sentiment analysis terms. [59] have developed a new Fisher discriminant ratio based on a feature selection method for text data sentiment analysis

3. STOCHASTIC PROGRAMMING

The stochastic nonlinear programming (SNLP) problems signify a significant class of the optimization problems due to their omnipresence in real life situations. Many systems in nature are inherently nonlinear, requiring nonlinear models for their depiction, and therefore, nonlinear programming approaches for optimization. Another important factor for attention is uncertainty. Very rarely are the system details accurately known. Frequently the parameters and variables are known only in terms of their ranges or, in some cases, in terms of their probability distributions. In such cases, the stochastic programming approaches need to be resorted to for optimization [60]

4. FRAMEWORK OF STOCHASTIC PROGRAMMING MODEL

In the following, the framework of two-stage stochastic programming model is concisely defined. The stochastic linear programming model is stated as follows:

$$\min \; c_x^T x + \sum_{s=1}^{S} p_s (q^T y^s) \quad (1)$$

$$s.t. \quad Ax = b \quad (2)$$

$$T^s x + W y^s = h^s \qquad s = 1,..,S \quad (3)$$

$$x, y^s \geq 0 \qquad s = 1,..,S \quad (4)$$

Equations (2) signify the first-stage model and Equations (3) signify the second-stage model. $\boldsymbol{x}$ is the vector of first-stage decision variables which is scenario-independent. The optimal value of x is not conditional on the realization of the indeterminate parameters. $\boldsymbol{c_x}$ is the vector of cost coefficient at the first stage. A is the first-stage coefficient matrix and $\boldsymbol{b}$ is the relating right-hand-side vector. $\boldsymbol{y}$ is the vector of second stage (recourse) decision variables. $\boldsymbol{q}$ is the vector of cost (recourse) coefficient matrix and $\boldsymbol{h}^s$ is the relating right-hand-side vector and T^s is the matrix that ties the two stages together where $s \in \Omega$ signifies scenarios in future and p_s is the probability that scenario s occurs. In the second-stage model, the random constraint expressed in (3), $\boldsymbol{h}^s - \boldsymbol{T}^s\boldsymbol{x}$, is the goal constraint: violations of this are permitted, but the related penalty cost, $\boldsymbol{q^T y}$, will affect the choice of $\boldsymbol{x}$. $\boldsymbol{q^T y}$ is the recourse penalty cost or second-stage value function and $\sum_{s=1}^{S} p_s (q^T y^s)$ represents the expected value of recourse penalty cost (second-stage value function).

5. THE FUNDAMENTALS OF THE TWO-STAGE STOCHASTIC PROGRAMMING MODEL

A two-stage stochastic programming is used to discuss and solve several planning and management problems that involve risks and uncertainties. Other programming models have fewer applications than stochastic problems with divergence compensation in constrained systems. A random vector and a deterministic vector are used in the solution of a two-stage stochastic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

programming problem. A deterministic initial plan problem solving will be performed in the first stage. Before determining the random conditions of the problem, a deterministic initial plan is established. If the stochastic program problem with the two-stage model can be solved, then selecting the deterministic initial plan ensures the existence of a random vector in the divergent system compensation.

Assume the following problems:

$$\min (C, X) \tag{5}$$

$$A^0 X = B^0 \tag{6}$$

$$AX = B \tag{7}$$

$$X \geq 0 \tag{8}$$

Where:

$$C = \{c_j\}, \quad j = 1,2,\dots,m \tag{9}$$

$$B = (b_i), \quad i = 1,2,\dots,m \tag{10}$$

$$B^0 = (b_k^0), \quad k = 1,2,\dots,m \tag{11}$$

$$A^0 = \|a_{kj}^0\|, \quad k = 1,2,\dots,m; \; j = 1,2,\dots,n \tag{12}$$

$$A = \|a_{ij}\|, \quad i = 1,2,\dots m; \; j = 1,2,\dots,n \tag{13}$$

Assume that the elements of matrix $A = A(\omega)$, vectors $B = B(\omega)$, and $C = C(\omega)$ are at random. So the problem solving process (5)-(8) will be divided into two stages, with the first stage consisting of the observation of random parameters in the conditions from the first stage, followed by the selection of a deterministic non-negative plan X^0 that meets the conditions (6). The Y specification of each random event that coincides (according to) the values $A(\omega^0)$ and $B(\omega^0)$ is determined in the second stage. Calculate the divergence of $B(\omega^0) - A(\omega^0)X^0$ that occurs after the realization of $\omega^0 \in \Omega$ in condition (7). Define the divergence compensation vector $Y \geq 0$ using the following formula:

$$D(\omega^0)Y(\omega^0) = B(\omega^0) - A(\omega^0)X^0 \tag{14}$$

Where

$D = \|d_{il}\|, \quad i = 1,2,\dots,m \; l = 1,2,\dots,n_1$ denotes a compensation matrix with random elements. As a result, it is assumed that the random realization ω observed in the second stage is unaffected by the preliminary plan X^0 selection.

Take a look at the following math program problems:

Determine the vectors X in n dimensions and $Y(\omega)$ in n_1, $\omega \in \Omega$ dimensions. Which results in:

$$\min_X E_\omega \left\{ (C(\omega), X) + \min_Y (H, Y(\omega)) \right\} \tag{15}$$

with constraints

$$A^0 X = B^0 \tag{16}$$

$$A(\omega)X + D(\omega)Y(\omega) = B(\omega), \qquad \omega \epsilon \Omega \tag{17}$$

$$X \geq 0, Y(\omega) \geq 0 \tag{18}$$

H is a penalty vector that is determined by the component value of the vector $Y(\omega)$, which is used to compensate for divergence. E is a mathematical notation for expectation. After determining the initial plan for X^0, we select the vector component $Y(\omega)$ by ensuring the smallest penalty for divergence compensation under the given conditions. In other words, once the vector X^0 has been determined, the following problem must be solved:

$$\left\{ \min_Y (H, Y(\omega)) D(\omega)Y(\omega) = B(\omega) - A(\omega)X^0, Y(\omega) \geq 0 \right\} \tag{19}$$

Problem (19) will have multiple plans, and the vector $Y(\omega)$ cannot be determined at each $\omega \in \Omega$, ensuring the discovery of conditions (17). Problems (15)-(18) are two-stage stochastic programming problems, and problem (19) is in the second stage.

Because there is always randomness affecting planned and management systems, models and approaches of problem solving in two-stage stochastic programming can be used from the perspective of planning and management operations (implementation). The two-stage model is also less sensitive to changes in the problem condition's parameters, making it more stable. As a result, the vector is suitable for the required first-stage plan because for each $\omega \in \Omega$, there exists a vector $Y \geq 0$ such that:

$$D(\omega)Y(\omega) = B(\omega) - A(\omega)X \tag{20}$$

Assume that the additional constraint mentioned implicitly in (20) occurs in the second stage of the resulting problem, and that the conditions assigned to the non-negative vector X of equation (16) have been determined.

Assume that $K_1 = \{X : A^0 = B^0, X \geq 0\}$ is defined by a defined constraint, but $K_2 = \{X : \forall\omega \in \Omega, \exists\, Y \geq 0, A(\omega)X = B(\omega) - D(\omega)Y(\omega)\}$ is defined by the resulting constraint. The set $K = K_1 \cap K_2$ is then the set of vectors X that solves the problem (15)-(18). If $X \in K$, then the vector X satisfies the constraints $A^0X = B$, $X \geq 0$, and the second stage problem (7) will have many plans until then.

The following results are required for further calculations:

Theorem 3.1. In a two-stage stochastic program problem, the set K with vector X is convex.

Proof: $K = K_1 \cap K_2$, but $K_1 = \{X : A^0 = B^0, X \geq 0\}$ is convex. Define $K_{2\omega} = \{X \mid \exists\, Y(\omega) \geq 0\}$ such that $A(\omega)X = B(\omega) - D(\omega)Y(\omega)$ is convex for a given (defined) $\omega \in \Omega$ set. As a result of convex sets, $K_2 = \cap_{\omega\in\Omega} K_{2\omega}$ and $K = K_1 \cap K_2$ are convex sets.

6. MULTI-STAGE STOCHASTIC PROGRAMMING

In a problem where time and uncertainty play a significant role, the decision model should be planned to permit the user to adopt a decision policy that can react to events as they unfold. The specific form of the decisions hinge on assumptions regarding the information that is available to the decision maker, when (in time) is it available and what adjustments (recourse) are available to the decision maker. In multi-stage stochastic programming (MSP) a lot of emphasis is set on the decision to be made today, given present resources, future uncertainties and possible recourse actions in the future. The uncertainty is signified through a scenario tree and an objective function is selected to characterize the risk related with the sequence of decisions to be made and the whole problem is then unraveled as a large scale linear or quadratic program ([33])

7. SCENARIO TREE

Scenario tree is a computationally viable way of discretizing the underlying dynamic stochastic data over time in a problem. A demonstration of scenario tree is given in Figure 1. In a scenario tree, each stage represents the stage of the time when new information is available to the decision maker. Therefore, the stages do not necessarily match up to time periods. They might include several periods in the planning horizon. Scenario tree be made up of several nodes and arcs at each stage. Each node n in the scenario tree characterizes a possible state of the world, linked with a set of data (stochastic demand, stochastic cost, etc.) in the related stage. The root node of the tree characterizes the current state of the world. The branches (arcs) in the scenario tree symbolize the scenarios for the next stage. A probability is linked to each arc of scenario tree which signifies the probability of the matching scenario to that arc. It should be noted that, the probability of each node in the scenario tree is calculated as the product of probability of the arcs from the root node to that node. Moreover, the sum of probabilities of nodes at each stage is equivalent to 1. A path from the root node to a node n defines one realization of the stochastic process from the present time to the period where node n appears. A full evolvement of the stochastic process over the entire planning horizon, i.e., the path from the root node to a leaf node, is called a scenario. In the scenario tree example of figure 1, we have 4 stages. Each node n in the tree has two branches to the next stage which signify two possible scenarios for the next stage, when we are at stage n. Therefore, we have 8 scenarios by the end of stage 4. A review of approaches for generating the scenario trees for multi-stage stochastic programs, based on the underlying random data process is given in ([34]).

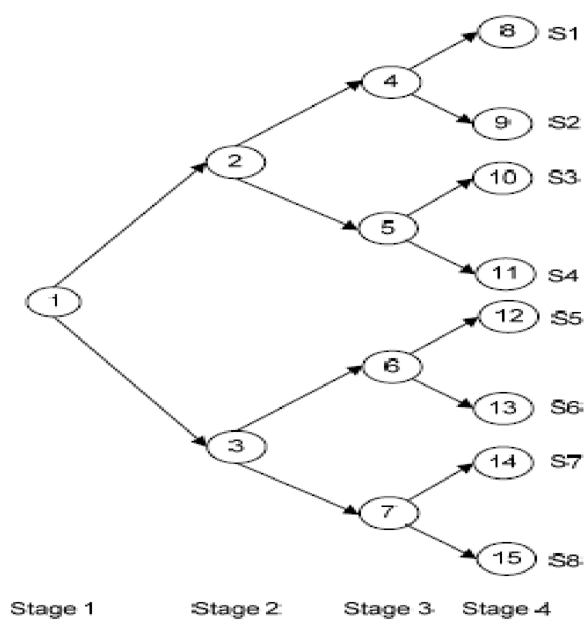

Figure 1. Scenario tree for multi-stage stochastic programming

Journal of Theoretical and Applied Information Technology
30th April 2021. Vol.99. No 8

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

8. STOCHASTIC MIXED INTEGER PROGRAMMING

The two-stage stochastic mixed-integer programming model requires integer values for subsets of the first and second stage variables. Assume $\overline{w}$ is a random variable used to model the data in a two-stage model to represent the problem. Because the stochastic program model is intended for decision making, a decision vector *x* must be chosen so that the consequences of the decision (as compared to some alternative $\overline{w}$ outcome) are accommodated in the optimal choice model. The consequences of the first stage decision are quantified using an optimization problem known as the recourse problem, which allows for observation (random variables). Assume that *w* represents an observation of $\overline{w}$. The consequences of choosing *x* versus the result *w* can then be modeled as follows:

$$h(x,w) = \min g(w)^T y \tag{21}$$

$$W(w)y \geq r(w) - T(w)x \tag{22}$$

$$y \geq 0, y_j \text{ integer } j \in J^2 \tag{23}$$

With J^2, the set of indexes can include some or all of the variables in $y \in R^{n_2}$. It is assumed here that all $W(w)$ realizations are matrices of size $m_2 \times n_2$.

9. EQUIVALENT DETERMINISTIC FORMULATION

Consider the linear stochastic program model shown below:

$$\left.\begin{array}{l} \min\ g_0(x,\tilde{\xi}) \\ \text{constraint } g_i(x,\tilde{\xi}) \leq 0, i = 1,\ldots,m, \\ \qquad x \in X \subset \mathbb{R}^n, \end{array}\right\} \tag{24}$$

where $\tilde{\xi}$ is a random vector that varies across the set $\Xi \subset \mathbb{R}^k$. To be more specific, it is assumed that the *F* family of "events" namely the subsets of Ξ, and the probability distribution *P* over *F* are known. As a result, the probability *P (A)* is known for each subset $A \subset \Xi$ that represents events, namely $A \in F$. Furthermore, the $g_i(x): \Xi \rightarrow \mathbb{R} \forall x, i$ function is assumed to be a random variable, and the probability distribution *P* is assumed to be independent. However, problem (24) is not "well defined" because the definition of "min" and the constraint are not clear, if the decision value *x* is taken into account before knowing the realization of $\tilde{\xi}$. As a result, the modeling process must be revised in order to produce an equivalent deterministic model (24).

10. THE FORMULATION PROCEDURE

The analogous model of the linear stochastic program with recourse is carried out in the following manner for problem (24). Consider the following:

$$g_i^+(x,\xi) = \begin{cases} 0 & \text{if } g_i(x,\xi) \leq 0 \\ g_i(x,\xi) & \text{others} \end{cases} \tag{25}$$

Constraint *i* of (24) is violated if and only if $g_i^+(x,\xi) > 0$ for a decision *x* and ξ realization of $\tilde{\xi}$. It can be stated here that for each constraint, a recourse or second-stage activity $y_i(\xi)$ is selected after observing the realization of ξ in order to anticipate the constraint violation - if any - by satisfying $g_i(x,\xi) - y_i(\xi) \leq 0$. This extra effort is assumed to result in additional costs or q_i penalties per unit, so these additional costs (referred to as the recourse function) are as follows:

$$Q(x,\xi) = \min_y \left\{ \sum_{i=1}^{m} q_i y_i(\xi) \,|\, y_i(\xi) \geq g_i^+(x,\xi), i = 1,\ldots,m \right\} \tag{26}$$

As a result, total costs - first stage and recourse costs - are incurred.

$$f_0(x,\xi) = g_0(x,\xi) + Q(x,\xi) \tag{27}$$

In addition to (26), consider a more general linear recourse program with a recourse vector $y(\xi) \in Y \subset \mathbb{R}^{\bar{n}}$, (*Y* a polyhedral set such as $\{y | y \geq 0\}$, an arbitrary fixed $m \times \bar{n}$ matrix *W* (recourse matrix), and a unit cost vector $q \in \mathbb{R}^{\bar{n}}$, which yields the below function recourse for (27).

$$Q(x,\xi) = \min_y \{q^T y | Wy \geq g^+(x,\xi), y \in Y\} \tag{28}$$

so with:

$$g^+(x,\xi) = (g_1^+(x,\xi),\ldots,g_m^+(x,\xi))^T \tag{29}$$

Consider a factory that produces m products; $g_i(x,\xi)$ is the difference between demand and output product i. Then $g_i^+(x,\xi) > 0$ indicates that there is a scarcity of product i in comparison to demand. Assuming that the factory is committed to meeting demand, problem (26), for example, can be interpreted as purchasing the market shortage of product i. Problem (28) can be generated by a second-stage or emergency production program with the input factor y and the technology represented by the matrix W. If the $W = I$, $m \times m$ identity matrix is used, (26) becomes a subset of (28).

Finally, a nonlinear recourse program can be used to define the recourse function against (8); for example, $Q(x,\xi)$ can be chosen as:

$$Q(x,\xi) = \min\{q(y) | H_i(y) \geq g_i^+(x,\xi), i = 1,\dots,m; y \in Y \subset \mathbb{R}^n\} \quad (30)$$

where $q: \mathbb{R}^n \to \mathbb{R}$ and $H_i: \mathbb{R}^n \to \mathbb{R}$ are presumptively known.

In the applied case, the decision maker simply looks at the deterministic equivalent formulation, a two-stage stochastic program with recourse, to minimize the expected value of total costs (i.e., first stage and recourse costs).

$$\min_{x \in X} E_\xi f_0(x,\xi) = \min_{x \in X} E_\xi \{g_0(x,\xi) + Q(x,\xi)\} \quad (31)$$

The above two-stage problem can be expanded to a two-stage recourse program as follows: In addition to the two decisions x and y that must be made in stages 1 and 2, the problem now faces $K+1$ of the sequential decision $x_0, x_1, \dots, x_K (x_\tau \in \mathbb{R}^{n_\tau})$ that must be made at stage $\tau = 0, 1, \dots, K$. The term "stage" can, but does not have to, be interpreted as "period of time".

For the sake of simplicity, assume that (24)'s objectivity is deterministic, that is, $g_0(x,\xi) = g_0(x)$. At the $\tau(\tau \geq 1)$ stage, it is known that the realization of $\xi_1, \dots, \xi_\tau$ from the random vector $\xi_1, \dots, \xi_\tau$ and the previous decision $x_0, \dots, x_{\tau-1}$ must be decided against x_τ in order for the constraint to be satisfied (with the constraint function g_τ)

$g_\tau(x_0, \dots, x_\tau, \xi_1, \dots, \xi_\tau \leq 0)$ is met, which at this point can only be accomplished by a precise selection of x_τ, which is based on prior knowledge of decisions and realization. So, assuming the cost function $q_\tau(x_\tau)$, the recourse function is obtained at stage $\tau \geq 1$.

$$Q_\tau = (x_0, x_1, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau) = \min_{x_\tau} \{q_\tau(x_\tau) \mid g_\tau(x_0, x_1, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau) \leq 0\}$$

Which identifies the best recourse $\hat{x}_\tau$ action to take at time τ based on the previous decision and the observed realization up to stage τ, that is,

$$\hat{x}_\tau = \hat{x}_\tau(x_0, \dots, x_{\tau-1}, \xi_1, \dots, \xi_\tau), \tau \geq 1$$

As a result, we can calculate the total cost of the double-stage problem for multiple phases.

$$f_0(x_0,\xi_1,\cdots,\xi_K) = g_0(x_0) + \sum_{\tau=1}^{K} E_{\tilde{\xi}_1,\cdots,\tilde{\xi}_\tau} Q_\tau(x_0,\hat{x}_1,\cdots,\hat{x}_{\tau-1},\xi_1,\cdots,\xi_\tau) \quad (32)$$

The deterministic equivalent of the double stage stochastic program problem with recourse ABC is obtained.

$$\min_{x_0 \in X} \left[g_0(x_0) + \sum_{\tau=1}^{K} E_{\tilde{\xi}_1,\cdots,\tilde{\xi}_\tau} Q_\tau(x_0,\hat{x}_1,\cdots,\hat{x}_{\tau-1},\tilde{\xi}_1,\cdots,\tilde{\xi}_\tau) \right] \quad (33)$$

It is clearly a direct generalization of the two-stage stochastic program by recourse (31).

11. STOCHASTIC PROGRAMMING MODEL FRAMEWORK

In mathematical programming problems where some data are unknown, that is, they are not known for uncertainty, random impact, or statistical variation, the stochastic programming problem is called. Stochastic programming provides a general framework for modeling the path dependency of stochastic processes in the optimization model. The stochastic programming model considered in this study may be stated as follows:

$$\min_x f^1(x)+Q(x)$$
$$g^1(x)=0,$$
$$h^1(x)\le 0,$$
$$g^1: R^{nZ_1}\to R^{m_e},$$
$$h^1: R^{n_1}\to R^{m_i},$$
$$x\in R_+^{n_1}.$$

Where to:

$$Q(x)=E_\xi Q(x,\xi(w))$$
$$Q(x,\xi(w))=\min_y f^2(y(w),w)$$
$$g^2(x,y(w),w)=0$$
$$h^2(x,y(w),w)\le 0$$
$$g^2: R^{n_1+n_2}\times\Omega\to R^{y_e}$$
$$h^2: R^{n_1+n_2}\times\Omega\to R^{y_i}$$
$$y\in Y$$

Ω is a σ-equipped probability space. F is a measure of probability, ξ is a random variable with a measure of probability, and f^1, f^2, g^1, g^2, h^1, h^2 are

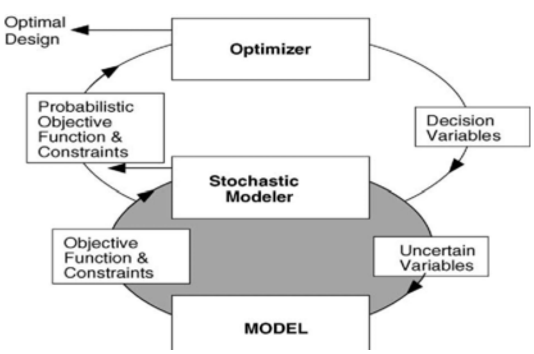

Figure 2. Approach framework with stochastic programming

nonlinear, distinguishable, but not convex. x is the first stage variable, y *(w)* is the second stage variable. The *y* set is defined as a set of *Y* subsets.
The main feature of the second stage of stochastic programming is the act of assistance. The decision set shall be divided into two parts. The decision before the parameters of the problem are known shall be taken as the first stage; and a decision after the uncertainty is revealed.
The framework for the use of stochastic modeling in feature selection techniques for sentiment analysis is described in the Figure 2 above.

12. CONCLUSIONS

It is no doubt that, the Internet and social media would be able to provide a major source of information about people's opinions. This paper concerns about the exploding growth in the Internet users which creates the main reasons that a method called sentiment analysis comes up. This analysis can help in extracting information about the opinion of people to classifies whether the opinion is positive or negative. One of the approaches in solving the sentiment analysis is feature selection method. This paper proposes the idea of using stochastic programming to tackle the feature selection optimization.

REFERENCES:

[1] B. Pang L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends® Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.
[2] D.-H. Yang and G. Yu, "A method of feature selection and sentiment similarity for Chinese micro-blogs," *J. Inf. Sci.*, vol. 39, no. 4, pp. 429–441, 2013.
[3] M. Kaya, G. Fidan, and I. H. Toroslu, "Sentiment analysis of turkish political news," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012, vol. 1, pp. 174–180.
[4] M. A. Fattah, "A novel statistical feature selection approach for text categorization.," *J. Inf. Process. Syst.*, vol. 13, no. 5, 2017.
[5] T. Parlar and S. A. Özel, "A new feature selection method for sentiment analysis of Turkish reviews," in *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 2016, pp. 1–6.
[6] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining.," in *LREc*, 2010, vol. 10, no. 2010, pp. 1320–1326.
[7] M. Sadegh, R. Ibrahim, and Z. A. Othman, "Opinion mining and sentiment analysis: A survey," *Int. J. Comput. Technol.*, vol. 2, no. 3, pp. 171–178, 2012.
[8] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi, and T. Li, "Dual sentiment analysis: Considering two sides of one review," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2120–2133, 2015.
[9] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," *World Wide Web*, vol. 17, no. 4, pp. 723–742, 2014.

[10] Z. Madhoushi, A. R. Hamdan, and S. Zainudin, "Sentiment analysis techniques in recent works," in *2015 Science and Information Conference (SAI)*, 2015, pp. 288–291.
[11] S. Mathapati and S. H. Manjula, "Sentiment analysis and opinion mining from social media: A review," *Glob. J. Comput. Sci. Technol.*, 2017.
[12] M. Trupthi, S. Pabboju, and G. Narasimha, "Improved feature extraction and classification—Sentiment analysis," in *2016 International Conference on Advances in Human Machine Interaction (HMI)*, 2016, pp. 1–6.
[13] R. Dehkharghani, H. Mercan, A. Javeed, and Y. Saygin, "Sentimental causal rule discovery from Twitter," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4950–4958, 2014.
[14] H. Wang, L. Liu, W. Song, and J. Lu, "Feature-based sentiment analysis approach for product reviews," *J. Softw.*, vol. 9, no. 2, pp. 274–279, 2014.
[15] J. Wang *et al.*, "Aspect Sentiment Classification with both Word-level and Clause-level Attention Networks.," in *IJCAI*, 2018, vol. 2018, pp. 4439–4445.
[16] T. Xu, Q. Peng, and Y. Cheng, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis," *Knowledge-Based Syst.*, vol. 35, pp. 279–289, 2012.
[17] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
[18] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5197–5204, 2011.
[19] L. Wang and Y. Wan, "Sentiment classification of documents based on latent semantic analysis," in *International Conference on Computer Education, Simulation and Modeling*, 2011, pp. 356–361.
[20] P. Koncz and J. Paralic, "An approach to feature selection for sentiment analysis," in *2011 15th IEEE International Conference on Intelligent Engineering Systems*, 2011, pp. 357–362.
[21] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, 2013.
[22] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Machine Learning Proceedings 1994*, Elsevier, 1994, pp. 121–129.
[23] S.-W. Lin, Y.-R. Shiue, S.-C. Chen, and H.-M. Cheng, "Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11543–11551, 2009.
[24] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Syst. Appl.*, vol. 80, pp. 323–339, 2017.
[25] M.-H. Sigari, M.-R. Pourshahabi, and H.-R. Pourreza, "A new evolutionary-incremental framework for feature selection," *Adv. Artif. Intell.*, vol. 2014, 2014.
[26] R. V Isachenko and V. V Strijov, "Quadratic programming optimization with feature selection for nonlinear models," *Lobachevskii J. Math.*, vol. 39, no. 9, pp. 1179–1187, 2018.
[27] A. Kaur and V. Gupta, "A survey on sentiment analysis and opinion mining techniques," *J. Emerg. Technol. Web Intell.*, vol. 5, no. 4, pp. 367–371, 2013.
[28] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," in *2015 Science and Information Conference (SAI)*, 2015, pp. 352–361.
[29] Z. Hailong, G. Wenyan, and J. Bo, "Machine learning and lexicon based methods for sentiment classification: A survey," in *2014 11th web information system and application conference*, 2014, pp. 262–265.
[30] H. P. Patil and M. Atique, "Sentiment analysis for social media: a survey," in *2015 2nd International Conference on Information Science and Security (ICISS)*, 2015, pp. 1–4.
[31] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2015.
[32] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," *arXiv Prepr. cs/0409058*, 2004.
[33] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," *arXiv Prepr. cs/0205070*, 2002.

[34] C. Nicholls and F. Song, “Comparison of feature selection methods for sentiment analysis,” in *Canadian Conference on Artificial Intelligence*, 2010, pp. 286–289.
[35] B. Agarwal and N. Mittal, “Prominent feature extraction for review analysis: an empirical study,” *J. Exp. Theor. Artif. Intell.*, vol. 28, no. 3, pp. 485–498, 2016.
[36] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, no. Mar, pp. 1157–1182, 2003.
[37] F. Akba, A. Uçan, E. A. Sezer, and H. Sever, “Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews,” in *8th European Conference on Data Mining*, 2014, vol. 191, no. 2002, pp. 180–184.
[38] K. Sagar and A. Saha, “Qualitative usability feature selection with ranking: a novel approach for ranking the identified usability problematic attributes for academic websites using data-mining techniques,” *Human-centric Comput. Inf. Sci.*, vol. 7, no. 1, pp. 1–24, 2017.
[39] V. Aksakalli and M. Malekipirbazari, “Feature selection via binary simultaneous perturbation stochastic approximation,” *Pattern Recognit. Lett.*, vol. 75, pp. 41–47, 2016.
[40] A. Duric and F. Song, “Feature selection for sentiment analysis based on content and syntax models,” *Decis. Support Syst.*, vol. 53, no. 4, pp. 704–711, 2012.
[41] A. Abbasi, “Intelligent feature selection for opinion classification,” *Technology*, vol. 54, no. 14, pp. 1269–1277, 2003.
[42] A. Balahur, R. Steinberger, E. Van Der Goot, B. Pouliquen, and M. Kabadjov, “Opinion mining on newspaper quotations,” in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, 2009, vol. 3, pp. 523–526.
[43] X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen, “Aspect-level opinion mining of online customer reviews,” *China Commun.*, vol. 10, no. 3, pp. 25–41, 2013.
[44] O. S. Qasim and Z. Y. Algamal, “Feature selection using particle swarm optimization-based logistic regression model,” *Chemom. Intell. Lab. Syst.*, vol. 182, pp. 41–46, 2018.
[45] D. Rodrigues *et al.*, “A wrapper approach for feature selection based on bat algorithm and optimum-path forest,” *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2250–2258, 2014.
[46] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, “Feature selection based on rough sets and particle swarm optimization,” *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 459–471, 2007.
[47] K.-H. Han and J.-H. Kim, “Quantum-inspired evolutionary algorithm for a class of combinatorial optimization,” *IEEE Trans. Evol. Comput.*, vol. 6, no. 6, pp. 580–593, 2002.
[48] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.
[49] A. Banks, J. Vincent, and C. Anyakoha, “A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications,” *Nat. Comput.*, vol. 7, no. 1, pp. 109–124, 2008.
[50] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, “A survey on metaheuristics for stochastic combinatorial optimization,” *Nat. Comput.*, vol. 8, no. 2, pp. 239–287, 2009.
[51] A. Unler and A. Murat, “A discrete particle swarm optimization method for feature selection in binary classification problems,” *Eur. J. Oper. Res.*, vol. 206, no. 3, pp. 528–539, 2010.
[52] L.-Y. Chuang, S.-W. Tsai, and C.-H. Yang, “Improved binary particle swarm optimization using catfish effect for feature selection,” *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12699–12707, 2011.
[53] Y.-J. Zheng, H.-F. Ling, J.-Y. Xue, and S.-Y. Chen, “Population classification in fire evacuation: A multiobjective particle swarm optimization approach,” *IEEE Trans. Evol. Comput.*, vol. 18, no. 1, pp. 70–81, 2013.
[54] B. Xue, M. Zhang, and W. N. Browne, “Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms,” *Appl. Soft Comput.*, vol. 18, pp. 261–276, 2014.
[55] Y. Zhai, Y.-S. Ong, and I. W. Tsang, “The emerging" big dimensionality",” *IEEE Comput. Intell. Mag.*, vol. 9, no. 3, pp. 14–26, 2014.
[56] R. Cheng and Y. Jin, “A competitive swarm optimizer for large scale optimization,” *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 191–204, 2014.
[57] B. Agarwal and N. Mittal, “Categorical probability proportion difference (CPPD): a feature selection method for sentiment classification,” in *Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology*, 2012, pp. 17–26.
[58] M. Simeon and R. Hilderman, “Categorical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proportional difference: A feature selection method for text categorization," in *Proceedings of the 7th Australasian Data Mining Conference-Volume 87*, 2008, pp. 201–208.

[59] S. Wang, D. Li, Y. Wei, and H. Li, "A feature selection method based on fisher's discriminant ratio for text sentiment classification," in *International Conference on Web Information Systems and Mining*, 2009, pp. 88–97.

[60] V. Gupta and N. Kallus, "Data-Pooling in Stochastic Optimization," *Available SSRN 3397336*, 2019.

[61] F. Y. Al-Aboosi and M. M. El-Halwagi, "A stochastic optimization approach to the design of shale gas/oil wastewater treatment systems with multiple energy sources under uncertainty," *Sustainability*, vol. 11, no. 18, p. 4865, 2019.

[62] J. Li and W. Yuan, "Stochastic combinatorial optimization via poisson approximation," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 2013, pp. 971–980.