Optimizing online review inspired product attribute classification using self-learning particle swarm based Bayesian learning approach

Abstract

Bowing to the burgeoning needs of online consumers, exploitation of social media content for extrapolating buyer-centric information is gaining increasing attention of researchers and practitioners from service science, data analytics, machine learning and associated domains. The current paper aims to identify the structural relationship between product attributes and subsequently prioritize customer preferences with respect to these attributes while exploiting textual social media data derived from fashion blogs in Germany. A Bayesian Network Structure Learning (BNSL) model with K2-score maximization objective is formulated and solved. A self-tailored metaheuristic approach that combines Self-Learning Particle Swarm Optimization (SLPSO) with K2 algorithm (SLPSOK2) is employed to decipher the highest scored structures. The proposed approach is implemented on small, medium and large size instances consisting of nine fashion attributes and 18 problem sets. The results obtained by SLPSOK2 are compared with Particle Swarm Optimization/K2 score (PSOK2), Genetic Algorithm/K2 score (GAK2), and Ant Colony Optimization/K2 score (ACOK2). Results verify that SLPSOK2 outperforms its hybrid counterparts for the tested cases in terms of computational time and solution quality. Furthermore, the study reveals that psychological satisfaction, historical revival, seasonal information and facts and figure based reviews are major components of information in fashion blogs that influence the customers.

Keywords: Bayesian network structure learning; self-learning particle swarm optimization; machine learning; customer preference ordering; fashion products

1. Introduction

The emergence of several interactive online blogs, social media websites and online user interaction applications have motivated the computer scientist to exploit the social media data and extract useful information (Chong et al. 2013). In this context, many problems from distinct domains have been formulated and treated by big data analysis, sentiment analysis, data mining, qualitative and quantitative model building in the pursuit of transforming data to knowledge (Chong et al. 2015). The resulting knowledge is used as feedback to improve the functionality of a process or product. One such authentic form of data is the user comments that textually conceal the end-users' perspectives and preferences. Terabytes of textual data hosted by online selling platforms in the form of reviews serve as important data sources which when subjected to rigorous scientific analysis and sound interpretation are expected to reveal useful buyer-centric information. Identification of structural linkages between different context specific product attributes and establishing an improved understanding of dependencies and interdependencies between these attributes are essential for realizing customer oriented product designs. Thus, there is a compelling need for devising an efficient scientific methodology to glean unsupervised buyer-centric information from online textual social media data to foster customer oriented product designs. In this regard, this paper aims to develop a unique methodology for identifying customer preferences and their relationships amongst select attributes inspired from online blogs in the context of fashion products.

Fashion industry occupies a unique space in the online market inherently subjected to limited period sales with constantly changing and uncertain demands. The ability of online platforms in hosting a wide range of products encompassing variety of brands and simultaneously to gather diverse group of customers has enhanced the need for furthering the exiting potential of fashion blogging. Germany, France, Britain, USA and India are few countries who have witnessed wide-spread use of such online forums towards e-marketing in this domain. Sheridan et al. (2006) recognized the increased sales of fashion kits and cosmetic products through online forums. Studies that attempt to measure customer parameter affiliations of fashion products while identifying the potential criteria that can explain the future design of products in fashion industry are scarcely found in literature. The existing pricing and inventory models would not be directly applicable while dealing with fashion products and must witness significant modifications with particular emphasis towards design for demand fluctuations and product variety. Therefore, it is important to gain deeper understanding on the factors or parameters that influence customer purchasing decisions and product demands such as prices, discounts, competition, brands and loyalty (Huang et al., 2014). A system dynamic framework to assess risk in the supply chain of fast fashion apparel industry was developed by Mehrjoo et al. (2016). Yang et al. (2015) designed optimal reservation pricing strategy for a two-echelon fashion supply chain subject to partial forecast updates and asymmetric cost information. However, the existing quantitative approaches fail to capture real-time prediction of customer preferences. Qualitative prediction of customer approvals and disapprovals with respect to different products from blogging data is a challenging task. For example, interests of a customer for beauty (cosmetic) products significantly vary with respect to age, gender and place of origin. Ambiguity prevailing with personal choices of the customer or person posting the comment with respect to products further intensifies the difficulty in capturing people's sentiments in this domain.

In addition to the wide recognition of data mining and machine learning techniques for exploring and exploiting structured or unstructured data across multiple domains, its implementation is continuously gaining momentum for facilitating meaningful extraction of information from social media data. As identified by Injadat et al. (2016), the application of Bayesian Networks (BN) and Support Vector Machines (SVM) classification approaches significantly dominate the wide range of tools available while dealing with information extraction from various forms of social media data. According to them, 20% of the contributed papers in data mining research on social media employed BNs, and 22 % of the articles adopted support vector machines. Furthermore, majority of articles that focus on exploitation of text snippets from social media data through data mining consider BN based learning as a primary AI tool, whereas SVM is known to remain favorable for full-length texts. Other methods including Artificial Neural Networks (ANN), k-Nearest Neighbour (k-NN), Genetic Algorithms, Fuzzy, k-means etc. were spotted in less than 10 % of the articles in any domain.

Bayesian network structure learning (BNSL) models appertain to the broad class of probabilistic relational models (PRMs). PRMs foster the ability to handle uncertainty associated with multiple-valued attribute based relational structures with ease and precision (Le and He, 2014). In this study, the attribute-value relations pertaining to fashion industry are mapped using Bayesian network where the edges of the network establish the link between different attributes that define the customer preferences. The problem that evaluates the best structure for a given set of attribute-value database in BNs is NP-hard (Chickering, 2004). The need for developing efficient heuristics for excavating high quality structures was emphasized by Gheisari (2016). Therefore, this research attempts to address the aforementioned structure learning problem using the combination of a scoring algorithm (K2) and a superior swarm intelligence variant, Self-Learning Particle Swarm Optimization (SLPSO). While the new approach is called SLPSOK2, it integrates K2 scoring technique with self-learning particle swarm optimization to search for the best ordering of variables/attributes defined in the context of social media data originating from multiple fashion blogs. Ultimately, the superiority of the proposed approach is validated by comparing the results obtained by SLPSOK2 with self-tailored integrated versions of Genetic algorithm (GA), Ant colony optimization (ACO), and Particle swarm optimization (PSO) and K2 algorithm respectively called as GAK2, ACOK2, and PSOK2 in terms of solution quality and execution time. The next section gives a brief overview of the relevant literature.

2. Literature review

In recent years, market driven product design strategies have evolved rapidly exploiting the vast data available on account of emergence of online social networks (OSNs) (Jin et al. 2016a; Jin et al. 2016b). It is important to understand customer preferences for the effective functioning of an interactive blogging platform (Wang 2012). Several qualitative and quantitative approaches have been proposed to analyze and interpret data (Cao and Schniederjans, 2004; Alkahtani et al. 2018). Prime sources of online data arise from facebook, twitter, instagram and other social blogs. In this paper, user comment data from several frequently used fashion blogs in Germany is collected and translated for the purpose of studying the customer preferences and the inherent relationships. The data to knowledge conversion process is driven by the 4V's of big data: volume, variety, velocity and value. Li and Parlikad (2016) discussed deeper issues related to social media in the Internet of Industrial Things (SIoIT) perspective emphasizing on Industry 4.0 while identifying the social assets and building blocks for SIoIT.

Learning in Bayesian Networks (BNs) can be classified in to parameter and structure learning. Parameter learning corresponds to the learning approach where the structure of a given network is fixed and the conditional probability of each node is estimated, where as in structure learning, the parameters are known in advance and the best suitable structure is estimated. A parameter settings model using Bayesian networks was investigated by Pavon et al. (2008). Petitjean et al. 2018 proposed an accurate parameter estimation model using hierarchical Dirichlet processes. According to Aouay et al. (2013), structure learning approaches are further classified into Constraint based and Score based approaches. Zhou et al. (2016) studied the influence of parameters under monotonic constraints. Similarly, a constraint based technique using the theory of causation was proposed by Pearl and Verma (1995). Jiang et al. (2018) proposed an entropy based improved constraint based algorithm for Bayesian network learning by capturing conditional mutual information. Amongst score based approaches, Bayesian Dirichlet (BD) is a standard metric for estimating the structure quality as reported by Cano et al. (2013). Ojha et al. (2018) proposed Bayesian network model using K2 scoring technique for solving supply chain risk propagation problem. Score and search based approaches are being combined with evolutionary and swarm intelligence techniques for solving the Bayesian structure learning optimization problem (Wang and Liu 2018; Zhang et al. 2018).

Evolutionary algorithms and their variants have been deployed independently and in combination with other techniques to solve various combinatorial optimization problems (Chaudhry and Luo, 2005; Cho et al. 2017). Emulating, the biological evolution mechanism, GA is well-known solution approach to deal with many optimization problems (Lin 2013). In general, evolutionary algorithms have been successfully employed to address complex combinatorial problems such as the transportation, scheduling, and bin packing problems (Mogale et al. 2018). Evolutionary approaches have been successful problem solvers for Bayesian network learning and inference problems (Larrafiaga et al. 1996, Larranaga et al. 2013). Wong et al. (1999) published a short paper on using evolutionary programming approach to solve the BNSL problem. With focus to improve product acceptability Garces et al. (2016) demonstrated the application of Bayesian network and simulated annealing based optimization model on a medical stocking threading device.

Hybridization of evolutionary algorithms with Bayesian learning techniques has seen significant benefits (Pitiot et al. 2010; Askari and Ahsaee 2018). Ji et al. (2011) employed an Ant Colony Optimization (ACO) based hybrid method for structure learning and later conducted a comparative study on different swarm intelligence techniques (Ji et al. 2017). Yang et al. (2016) used bacterial foraging optimization and made a comparative analysis with other methods to prove the superiority of their approach. A review of articles focusing on application of evolutionary computation and various data mining techniques for effective customer relationship management was conducted by Krishna and Ravi (2016). However, the primitive versions of these algorithms face the problem of immature convergence. To overcome this difficulty, several advanced versions of the algorithms have been developed and implemented on wide-ranging class of problems (Wang et al., 2018), whilst the literature lacks significant attention of such approaches in this domain. In this paper, the efficiency of Self-learning particle swarm optimization (SLPSO) is investigated for searching appropriate ordering of customer preferences and their structures. SLPSO is a recently developed modified form of PSO proven to have better performance than previously developed metaheuristics by virtue of its four operators and a newly incorporated 'abest' principle.

The key contribution of this paper lies in formulating the online customer preference ordering problem as a BNSL problem and solving it using an integrated metaheuristic score based approach proposed to address the given problem. The novel approach combines K2 scoring technique and the PSO variant, SLPSO to arrive at optimal structure and the best customer preference ordering. A context specific reordering operator has been proposed to guide the stochastic optimization algorithm to arrive at quick and accurate solutions. In the end, the benefit of implementing Self–learning particle swarm optimization (SLPSO) over other state-of the art metaheuristics for the given problem is explored.

The remainder of this work is structured as follows. The next section outlines preliminary set of concepts required to perceive further development of this work. Section 4 outlines the problem environment and data collection procedure. Section 5 introduces the hybrid meta-heuristic with the help of a step-wise flow-diagram. Section 6 presents the scenario wise experimentations conducted and detailed discussion on the results obtained for each scenario. Finally, section 7 and section 8 discuss managerial insights, conclusion and future work.

3. Preliminaries

3.1 *Bayesian Network (BN)*

Bayesian network (BN) or belief network (BN) or causal network (CN) is a collection of directed arcs and nodes that is used to graphically represent the variables and their interdependencies. Each node represents a variable and the directed arcs represent the direction of parent-child relationship between variables. The weight of each arc represents the strength of relationship between the dependent variables. Independent variables or nodes are assigned zero weight to signify that the variables are not directly related. A BN is better visualized as a Directed Acyclic Graph (DAG). The key parameters on which BN is defined are (M, A, P), where $M = \{1, 2, 3, ..., n\}$ denotes set of nodes, A denotes set of arcs and P denotes the joint probability distribution that represents the network. The nodes are otherwise called as vertices and arcs as edges. The joint probability distribution P is defined by Equation 1.

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \tag{1}$$

where, $X = \{X_1, \ldots, X_n\}$ is a vector of n discrete variables, and $Pa(X_i)$, represents parents of variable, X_i, $P(X_i \mid Pa(X_i))$ represents the conditional occurrence probability of X_i given $Pa(X_i)$

3.2 *K2 algorithm*

The K2 algorithm classifies given set binary variables to a parent-child tree (Cooper and Herskovits, 1992). In this algorithm, the consistency of the resulting Directed Acyclic Graph (DAG) is determined using the K2 score, which is considered as a superior BN evaluation metric. Larger the value of the metric, better is the structural quality of the DAG**.** The procedure requires the following four inputs before start of execution: (1) the set of nodes (variables), M (2) the ordering of nodes (3) the upper bound, u and (4) the database, D. The step-wise procedure for the classification algorithm is given in (Cooper and Herskovits, 1992)**.** However, the fact that ordering of the variables must be provided prior to start of the algorithm still remains a striking disadvantage towards evaluating the exhaustive set of possible BNs for given set of variables, X_i. The probability, $f(i,\pi_i)$ of database D, given that π_i is the set of parents of X_i is calculated using Equation 2,

$$f(i,\pi_i) = \prod_{j=1}^{q_i} \frac{(r_i-1)!}{(N_{ij}+r_i-1)!} \prod_{k=1}^{r_i} \alpha_{ijk}!, \tag{2}$$

where, $q_i = |\phi_i|$ and ϕ_i is the set of possible instantiations of $\pi_i \mid \pi_i = \{p_1, p_2, ..., p_s\}$, and $\phi_i = \left\{v_1^{p_1}, ..., v_{r_{p_1}}^{p_1}\right\} \times ... \times \left\{v_1^{p_s}, ..., v_{r_{p_s}}^{p_s}\right\}$. r_i is cardinality of set V_i, where V_i is the set of possible values of attribute X_i. α_{ijk} is the number of occurrences in D where X_i is instantiated with k^{th} value, where $k \in \{1,2,3,...,r_i\}$ and ϕ_i is instantiated with j^{th} value, where $j \in \{1,2,3,...,q_i\}$. N_{ij} is the number of times X_i is instantiated with j^{th} value of ϕ_i in D and is given by $N_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$. The probability $f(i,\pi_i)$ is utilized to evaluate the K2 score for each node. The comprehensive K2 score, $K2_M$ for the whole network is the summation of the K2 scores obtained for individual nodes of the given network (Equation 3),

$$K2_M = \sum_{n=1}^{|M|} K2_n \tag{3}$$

where, $K2_n$ represents the K2 score of an individual node *n*.

```
Algorithm: K2
-------------------------------------------------------------
Input M, u, D
for i = 1:|N|
    π_i = ∅;
    P_old = f(i, π_i)
    proceed = 1;
    While proceed == 1 and |π_i| < u do
        let z be the node in pred(X_i) − π_i that maximizes f(i, π_i ∪ {z});
        P_new = f(i, π_i ∪ {z});
        if P_new > P_old then
            P_old = P_new;
            π_i = π_i ∪ {z};
        Else proceed == 0;
    end
    print ('Node:', X_i, 'Parent of X_i:', π_i);
end
```

Figure 1. Pseudo code for K2 algorithm

The pseudocode for determining the K2score using K2 algorithm is as shown in Figure 1. The following subsections give an overview of the metaheuristics used to determine the maximized K2 score in this paper.

3.3 *Genetic Algorithm (GA)*

Genetic algorithm is the most widely used naturally inspired evolutionary algorithm put forth by Goldberg (1989). GA populates its solution space by a parent selection process and subsequent offspring generation. The chromosome carries the solution information using a user specified encoding scheme. The exploration capability is defined by the power of crossover and mutation operators to reach for global optimums. Since its inception, the algorithm has evolved drastically and has been hybridized with newly devised operators and selection schemes. Nevertheless, it recognizes itself as the benchmark algorithm while dealing with complex optimization problems.

The implementation of GA consists the performing of following four steps. (1) Initial population generation (2) Fitness evaluation (3) Reproduction (4) Crossover and Mutation, and (5) Repair. In the first step N chromosomes are randomly generated, where, N is the size of population. Each chromosome is arranged as shown in Figure 2 (a), while the whole population is arranged in a matrix of size $N \times M$, where M is the number of nodes. In the second step, the fitness corresponding to each chromosome is evaluated using Equation (3). For the next step, this paper follows roulette wheel selection procedure for reproduction. In reproduction, the parents with highest fitness are frequently retained while the others are discarded. The fourth step performs crossover and mutation over the entire population according to probabilities p_c and p_m respectively. Single point crossover strategy is adopted for crossover. According to this strategy, a random position is chosen on the solution string of the parent individuals selected for cross over. Subsequently, the offspring individuals are created by exchanging all elements on the right side of the randomly chosen position. Integer swap mutation operator is chosen for performing the mutation. According to this, the mutated offspring or the mutant is created by swapping two randomly selected elements of the chromosome identified for mutation. The offspring individuals resulting from crossover and mutation are populated into the sample space for next iteration. In the last step, the sample space is repaired to not contain infeasible solutions. Steps 2-5 are repeated in a loop until the termination criteria (maximum number of iterations) is met.

3.4 *Ant Colony Optimization (ACO)*

Ant colony optimization (ACO) is a swarm intelligence optimization approach devised for solving combinatorial problems such as the knapsack, scheduling and travelling salesman problems (TSP) (Colorni et al. 1992; Neto and Filho 2013). In the literature, TSP is represented as a graph network (M, A), where M represents the set of nodes or cities and A represents the set of arcs or distance between the cities. TSP involves the obtaining of minimal tour that covers each city not more than exactly once. ACO is an efficient approach to evaluate optimal tours for any complex optimization problem that is reducible to the TSP structure. ACO mimics the pheromone deposit and evaporation principles adopted by ants to remember and forget their previous visits respectively. Since ACO is designed for minimization of total route length and the problem addressed in this paper is originally of maximization type, the objective function is rewritten as Equation (4) for the purpose of ACO implementation. The fitness calculated according to Equation (4) is analogous to the total distance travelled by the ants through the set of nodes, M.

$$\text{Minimize } f = -K2_M \tag{4}$$

The implementation of ACO consists of the following steps. First, the ant colony is initialized with sample feasible solutions generated by virtue of probability p_{ij}^k assigned to each arc (i, j) of ant k, where, $i, j \in M$ using Equation (5).

$$p_{ij}^k = \begin{cases} \dfrac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum\limits_{l \in N_i^k} \tau_{il}^{\alpha}\eta_{il}^{\beta}}, \forall j \in N_l^k \\ 0, \qquad \text{otherwise} \end{cases} \tag{5}$$

where, τ_{ij} represents the pheromone concentration on arc (i, j). N_i^k is the set of feasible routes from node i for ant k. α is the relative importance of pheromone level over heuristic information β. η_{ij} is the visibility of node j from node i. The next step performs pheromone updation through all the paths traversed by each ant according to Equation (6), where $\Delta\tau_{ij}^{k}$ is the quantity of pheromone deposited on each arc arc (i, j) of ant k. It is calculated using Equation (7), where f^{k} represents fitness of ant k. Given that, τ_{ij}^{k} and $\tau_{ij}^{'k}$ are the old and new pheromone levels respectively, the pheromone update equations for each arc (i, j) of ant k can be written as

$$\tau_{ij}^{'k} = \tau_{ij}^{k} + \Delta\tau_{ij}^{k} \tag{6}$$

$$\Delta\tau_{ij}^{k} = \frac{Q}{f^{k}} \tag{7}$$

The third step calculates the pheromone evaporation corresponding to each arc (i, j) of ant k and updates the pheromone content according to Equation (8), where ρ denotes the evaporation rate.

$$\tau_{ij}^{'k} = (1-\rho)\tau_{ij}^{k} \tag{8}$$

The inclusion of pheromone evaporation is essential to control premature convergence of the algorithm. Whilst extreme (too high or too low) values for ρ disturb the quality of solution obtained, values in the range [0.04 0.1] are commonly preferred. The aforementioned steps are repeated until the maximum iteration criteria is reached.

3.5 *Particle Swarm Optimization (PSO)*

Swarm intelligence based optimization techniques are well-known for their robustness in getting adapted to multiple domains with their parameters being nearly fixed (Aouay et al., 2013). Increasing number of researchers have used this as a robust solution approach across many disciplines. Kennedy and Eberhart (1995) pioneered the invention of particle swarm optimization (PSO) to solve multi-faceted NP-hard optimization problems. Since then, the algorithm has been used to solve variety of problems from multiple domains. The search process in PSO is directed towards cognitive and social dimensions. However the problem of immature convergence remains a striking challenge while dealing with complex problems with PSO.

The mathematical procedure of PSO is explained as follows. The first step involves generation of initial population of size $N \times M$, where N and M represent population size and number of nodes respectively. In this step, position of each particle is initialized to a non-repetitive integer array, whereas velocity of each particle is randomly initialized. The position of a particle in the population is as shown in Figure 2 (a). The next step involves calculating the fitness of each particle in the population using Equation (3). In the subsequent step, the velocity and position of particle k in dimension d are updated according to Equations (9) and (10) respectively, where ω represents inertia weight, c_1 and c_2 are the acceleration constants that represent the cognitive and social evolution of the particle swarm respectively. r_1 and r_2 are random numbers drawn from uniform distribution in between [0 1]. $pbest_{kt}^d$ and $gbest_{kt}^d$ denote the local (particle) and global best solutions respectively. In the final step, the updated particles are repaired by eliminating infeasible points which are eventually replaced by appropriate numbers such that the resulting solution strings contain only non-repetetive integer arrays in between [1 M], where, M represents number of nodes in the given instance.

$$v_{k(t+1)}^d = \omega v_{kt}^d + c_1 r_1 (pbest_{kt}^d - x_{kt}^d) + c_2 r_2 (gbest_{kt}^d - x_{kt}^d) \tag{9}$$

$$x_{k(t+1)}^d = x_{kt}^d + v_{k(t+1)}^d \tag{10}$$

3.6 ***Self-learning particle swarm optimization (SLPSO)***

Self-learning particle swarm optimization (SLPSO) is a recently developed variant of PSO devised to reach global optimal solutions using an adaptive learning framework (Li et al., 2012). It is characterized by virtue of its four operators: the exploitation operator (Equation 11), the jumping operator (Equation 12), the exploration operator (Equation 13) and the convergence operator (Equation 14), by which it distinguishes itself from the previously developed versions of PSO. Equations 11-14 represent the position and velocity update rules followed by this algorithm. For a given swarm of dimension *d* with *k* particles,

$$v_{k(t+1)}^{d} = \omega v_{kt}^{d} + \eta \cdot r_{kt}^{d} (pbest_{kt}^{d} - x_{kt}^{d}) \tag{11}$$

where v_{kt}^{d} represents the velocity of d^{th} dimension for k^{th} particle in iteration *t*. ω and η are the inertia and acceleration constants respectively, r_{kt}^{d} is picked from uniform distribution over [0 1], $pbest_{kt}^{d}$ is the particle best obtained for the particle *k* in iteration *t*.

$$x_{k(t+1)}^{d} = x_{kt}^{d} + v_{(avg)t}^{d} \cdot N(0,1) \tag{12}$$

where x_{kt}^{d} is the position vector for d^{th} dimension of particle *k* in iteration *t*, $v_{(avg)t}^{d}$ represents the average velocity of particles in a population of size *N*, and $N(0,1)$ is a random number picked from normal distribution with mean 0 and variance 1. In Equations (13) and (14), $pbest_{rand}^{d}$ and $abest_{rand}^{d}$ denote the particle best of a random particle and best position reached by SLPSO respectively.

$$v_{k}^{d} = \omega v_{k}^{d} + \eta \cdot r_{k}^{d} (pbest_{rand}^{d} - x_{k}^{d}) \tag{13}$$

$$v_{k}^{d} = \omega v_{k}^{d} + \eta \cdot r_{k}^{d} (abest_{rand}^{d} - x_{k}^{d}) \tag{14}$$

The pseudo code for SLPSO and the algorithm flow diagram can be found in Li et al. (2012).

4. Problem environment and data collection

The problem addressed in the paper is inspired from a practical case of fashion marketing in Germany. The social media data collected from multiple online sources included user comments from selected fashion blogs. The comments summarized the likes/dislikes of the person posting the comment about a product in general or in specific about one or more attributes. The comments were posted by three kinds of persons: one who is reviewing but not buying the product, one who is marketing the product, and the one who has bought the product. The text in the user comment data comprises hidden information about sex, age and culture of the person posting the comment. This information is captured in the form of attributes by referring to relevant literature (Shoham, 2003). Later on, the attribute text were further refined by three industry and two academic experts and select variables that represent fashion attributes have been shortlisted (Table 1). It is hard to understand the overall perspective of the market from the existing information with respect to these attributes manually. Owing to the unstructured and randomly distributed data, we propose to model the problem as a BNSLP with aim to draw out the underlying structure between the variables (attributes).

Raw data collected from various fashion blogs was originally in German and consisted of more than 3000 entries. Subsequently the comments were translated to English for ease of processing. The efficiency of translation was found to be appreciable at a rate of one error for every 50 comments. Data cleansing was meticulously carried out to eliminate duplicate, missing, junk and empty values from the raw form of data. Finally, the processed data consisted 2368 rows of user comments in English.

The user comments are subsequently annotated against each attribute using binary numbers to capture the relevance of an attribute with respect a comment. For every combination of the set of attributes and the set of user comments, following question is answered: "*Does the discussion in a comment significantly expresses positive information about an attribute?*". If the answer for the aforementioned question is "Yes" for a particular comment, then the corresponding column of the attribute (variable) is given a marking of 1 in the database, otherwise 0. The presence of the attribute in each comment is identified by the presence of relevant text from Table 1 in the comment. For example, consider the following sample comment taken from real data set:

"*Short hairstyles are for summer 2016. That trend Short Cuts are by no means boring, showed the models who walked in fashion shows for spring / summer 2016 with a variety of stylings on the catwalk. So if you have not yet taken the radical step, should no longer be contend with and married. Become a shag, pixie cut or shaved short hair trendsetter! The Shag is a great way to familiarize yourself with shorter hair. The perfect length for Frans section is from about chin length hair. Most of the bangs and the top coat is rather short cut, as with Edie Campbell at Chloé.*".

A closer observation of the above comment would reveal the presence of information about season, market drive, aesthetic presence, positive opinion, facts and figures based reviews in the comment. For instance, the text "summer/winter" and "2016" signify the presence of seasonal information and facts & figures respectively. The corresponding variables for this comment in database D are annotated to unity. Similarly, all the comments are numerically annotated with respect to all attributes. The list of all attributes and the corresponding texts used to distinguish the appearance of an attribute in a comment are displayed in table 1. Mathematically, we define the binary variable p_{ik} to capture the presence of attribute k in comment i, where, $k \in K$, set of all attributes and $i \in I$, set of all comments in database D. The binary set of annotations are provided to the SLPSOK2 algorithm as part of the problem input.

Table 1. Text used to identify the presence of an attribute while classifying the comments.

Node number	Variable (Attribute) name	Attribute text used for comment classification
1	Social attitude	Social, life, culture, cloth, Christmas, school, family, party, glamourous, livelihood, festival, tradition, society, friends, colleagues, community, value, emotion, happy, together, relationship, belief, desirable, worth, opinion, trust, confidence, peers, lifestyle, life event, religion
2	Psychological satisfaction	Desires, opinions, faithful, trust, belief, satisfaction, happy, enjoy, willing, want, desire, convenient, comfortable, useful, user-friendly, brand, value, like, perform, need
3	Aesthetic presence	Aesthetic, appearance, sense, designer, model, beauty, conscious, participate, communicate, convey, experience, appreciate, dream, transform, reality, survive
4	Market drive	Market, need, demand, sale, sell, trend, advertisement, promotion, discount, customer desire, segment, value, price, influence, government, policy, program, purchase, buy, target market, new customer, strategy, tactic
5	Historical revival	Society, archives, museum, costumes, historical, reference, trends, tradition, god, generation, lively, passion, church, lord, visitation, blessing, forefather, heaven
6	Positive Opinion	Enhancement, improvement, increase, degrade, destroy, destructive, motivate, delight, happy, unhappy, satisfy, unsatisfied, pleased, unpleased, pleasure, gorgeous, beautiful, luxury, development, secure, safety, acceptable, secure, safety, acceptable, enrich, excellent, admirable, superb
7	Season information	Sweater, sunglass, sunscreen, moisturize, hot-coat, heater, weather, sunlight, cold, hand glow, wool, snowfall, raincoat, rain, umbrella,
8	Young fashion trend	Women, men, ladies, girl, boy, lipstick, nail polish, makeup, parent, child, baby, toy, gorgeous, handsome, stud, earning, heels, mother, father, grandmother, nail polish, hairstyle, hair color, T-shirt, skirt,
9	Facts and figures	Fact, figure, date, year, month, graph, 00,200, 10, 19

5. Solution approach

In this section we present the proposed integrated approach to solve the aforementioned problem. As discussed earlier the K2 algorithm proposed by Cooper and Herskovits (1992) efficiently evaluates the structural relationship based on the K2 score for a given set of ordering on product attributes. It is intended to determine the tree-structure of interdependent relationships between variables for a given set of binary data. The procedure needs the pre-ordering of variables to evaluate the exhaustive parent child relationships between the nodes (variables). K2 algorithm uses the K2 score to determine the superiority of one structure over the other. The approach inevitably suffers from poor exploration capability for different sizes and permutations of ordering. With addition of each node (variable) the number of combinations to be evaluated would increase by $\underline{\lfloor M+1} - \underline{\lfloor M}$ where, M is the number of nodes in the original problem. The exploding number of possibilities to search for maximized K2 score intensifies the problem complexity. For large samples, the BNSL problem poses to be NP hard (Chickering et al. 2004). Self Learning Particle Swarm Optimization (SLPSO) is marked to exhibit high robustness with fixed range of parameters for a wide range of problems (Yang and Nguyen, 2012). On the other hand, it individually lacks the efficacy to arrive at accurate network structure, nevertheless having an immense search process capability in discrete spaces. The advantage of integrating K2 algorithm with evolutionary and swarm intelligence metaheuristics was realized by Chen et al. (2007) and revisited by Yang et al. (2016). However, the implementation of the procedure with advanced metaheuristics and its validation on practically relevant problems is much less explored. In this context, K2 and SLPSO algorithms are integrated to overcome the individual discrepancies and exploit the synergy to derive the best inherent structure of relationships between the annotated binary variables and to extract useful information for the aforementioned problem. The hybrid swarm intelligence technique reaches for global maxima in the vast set of discrete solutions. The resulting optimal structure gives the most suitable combination of directional dependencies between the extracted variables (attributes). A similar effort is carried out to integrate the state of the art metaheuristics PSO, GA and ACO with K2 algorithm and the hybrid counterparts are respectively named as

PSOK2, GAK2, and ACOK2. Later, the computational ability of the proposed approach is compared against other three hybrid metaheuristics. The following sub-sections provide detail explanation to further understand the metaheuristic approach and their adaptation to current context.

5.1 *Encoding scheme*

This sub-section presents the encoding scheme followed in this paper for PSO, GA, ACO and SLPSO algorithm. Figure 2(a) shows the array structure or encoding scheme for a 5-node problem.

1	3	5	4	2

(a) Ideal solution string

1	1	5	4	2

(b) Solution string with repeated value error

1	-2	5	0.6	2

(c) Solution string with negative value and fractional value errors

Figure 2. Encoding scheme for 5-node problem.

It is worth recollecting that whether it is the chromosome in GA, or the tour in ACO, or the particle in PSO and SLPSO, the encoding scheme retains the same structure for all the cases. Since the variables are node numbers which represent the ordering on the given set of nodes (variables), they are bound to be integers. However, there is high chance of occurrence of a non-integer while generating initial solutions and in the post solution update phase of GA, PSO and SLPSO. In such cases, the integrality of variables is ensured by the xmod operator. The definition and functions for the xmod operator are presented in the following sub-section.

5.2 *The xmod operator*

The xmod operator is actually defined to operate on the randomly generated initial solution and updated solutions in such a way that undesired solution is eliminated and the solution string contains only positive integer non-repetitive arrangements. It serves importantly two functions. (1) To generate an initial feasible solution. (2) Modulate the solution string to a non-repetitive positive integer array. The randomly generated initial solutions in GA, PSO and SLPSO, may or may not be feasible solutions in the sense that they might contain non-positive or fractional values. These infeasibilities are removed by virtue of the first function of the xmod operator. As each particle represents the ordering of nodes it is bound to be a non-repetitive integer array. After initial particle generation there is a possibility that the array might contain repeated integer numbers. These are rectified by the “replcacemr” module embedded in the xmod operator. The “replacemr” is an exclusive user defined function that includes logic to identify repeated integers and replace them with missing integers in the particle array such that the array contains non-repetitive integer arrangement. Figure 2 give a visual illustration of the three types of infeasibilities discussed.

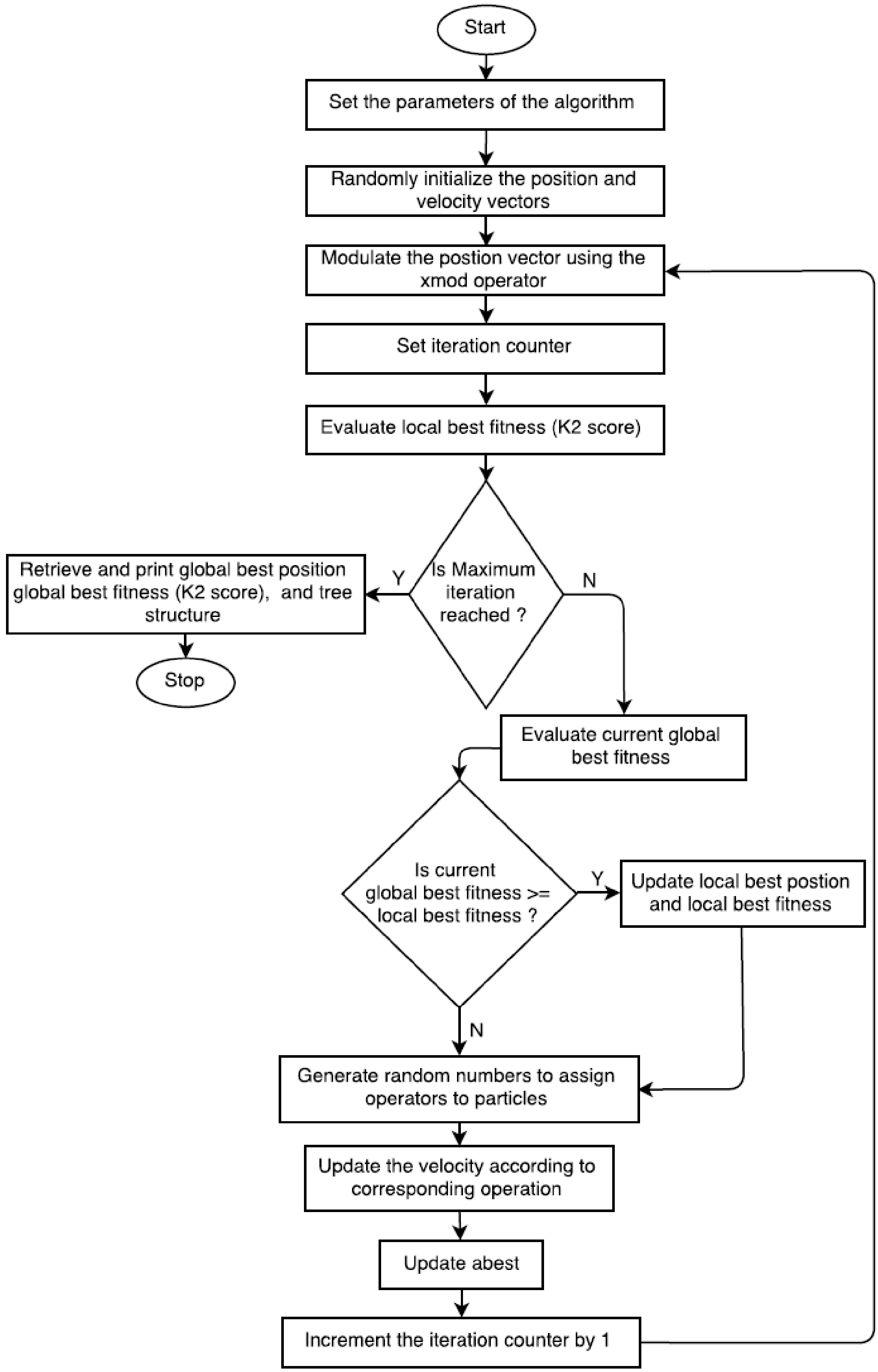


Figure 3. Step-wise flow of execution in SLPSOK2

5.3 *Generation of initial feasible solution*

Generation of initial basic feasible solution has been a standard practice in stochastic search processes. Adopting an efficient initial start strategy is known to greatly influence the execution times while searching for global optimum while dealing with NP hard problems. In this paper, to ensure a decent start of exploration, instead of randomly generating the particle array, the initial particle array is carefully chosen in such a way that there are no repeated integers, float or non-positive values in the array. Furthermore, the population is initiated with multiple starting points to maintain solution diversity.

5.4 *Algorithm steps*

The flow diagram for step-wise execution of the proposed SLPSOK2 algorithm is explained in Figure 3. The descriptions of other hybrid versions GAK2, ACOK2 and PSOK2 are self-explanatory and can be easily deduced from the detail explanation provided for SLPSOK2.

6. Computational study

6.1 *Experimental setup*

Experimentations have been conducted on 5 node, 7 node and 9 node problem scenarios with different combinations of maximum parents allowed for each node. In all, 18 problem sets are solved using GAK2, ACOK2, PSOK2 and SLPSOK2 algorithms, out of which first four problem sets correspond to 5 node scenario, the next six problem sets belong to 7 node scenario and final 8 sets correspond to 9 node scenario. Coding and implementation is carried out in MATLAB 2013a installed on a high end computing machine. The ordering of variables is provided as input to both the algorithms prior to start of each iteration. Initially the particle is randomly initialized and modulated using the xmod operator to generate an initial feasible solution. The fitness function is maximized and the best K2score, corresponding order of variables, execution times and parent child graphs are recorded for different scenarios. The population size is set to 50 and all the experiments are subjected to 50 iterations and 5 trial runs. The parameters for GA and ACO are adopted from standard literature. Experimental parameters for the swarm optimization techniques are taken from Kennedy and Eberhart (1995) for PSO and Li et al. (2012) for SLPSO. In the next subsection we discuss the results obtained for different scenarios considered.

6.2 *Result and discussion*

This section summarizes the results obtained while solving the BNSL problem for fashion marketing domain under three broad problem scenarios: 5 node, 7 node and 9 node. In addition to, PSO and SLPSO, the K2 algorithm is also integrated with state of the art algorithms, Genetic Algorithm (GA) and Ant Colony Optimization (ACO). The resulting GAK2 and ACOK2 are implemented on the proposed BNSL problem and the results are compared. Franceschetti et al. (2017) used lowest and average global fitness measures to evaluate the relative quality of solutions obtained using metaheuristics. Tables 2-4 encapsulate the Lowest K2score (LKS), Highest K2score (HKS), Average K2score (AKS), converged K2score (CKS), Standard Deviation K2score (SDKS), execution times and best preference orders obtained for all the problems while solving them using each of the four algorithms. It can be noticed that as the number of nodes in the problem increases the time and energy required for convergence is higher and increases exponentially. Hence, we chose to limit our experimentation to problems with less than 10 nodes. It is evident from Tales 2-4 that the average execution time required to converge for SLPSOK2 is less than the time required for other three metaheuristics. The percentage difference in computational times (average execution times) between SLPSOK2 and GAK2 for problem set 1 is calculated from Table 2 as (16.25-10.11)*100/16.25 which is equal to 37.78%. By this procedure, the average of percentage difference between SLPSOK2 and GAK2 over all the 18 instances is computed, which, when rounded off amounts to 31%. Similar procedure is adopted to calculate the percentage advantage of SLPSOK2 over ACOK2 and PSOK2. It is found that on an average, the novel integrated approach, SLPSOK2 outperforms GAK2, ACOK2, and PSOK2 by 31%, 17%, and 49% respectively with respect to computational time. Convergence graphs for 18 problem sets solved using the four algorithms are displayed in Figs. 4 and 5.

Table 5 presents a comprehensive summary of the results derived from Tables 2-4. Psychological satisfaction was found to be the most preferred criteria according to all the algorithms for the given instances. However, the least preferred criteria was found to be varying with respect to different algorithms, especially with increasing problem size. The best choice of solution is made based on highest convergence score (CKS) and minimum standard deviation score (SDKS) obtained for each dataset (Table 5). The CKS, SDKS, and customer attribute preference orders for first four problem sets were found to be identical with respect to all the algorithms. In some cases (problem sets 3, 4, 6, 8, and 9), ACOK2 and SLPSOK2 performed equally with respect to converged K2 score, whereas in majority of the cases SLPSOK2 provided best results with respect to CKS, excluding few anomalies (problem sets 10, 12, 17) where ACOK2 and PSOK2 individually outperformed other algorithms. The scores and execution time of ACOK2 are found to be closer to that obtained by SLPSOK2 for all the instances. This is because the search process in ACO directly includes discrete integer solution space unlike for GA and PSO where the updated solution is first continuous and is later approximated to an integer solution. However, the algorithm fails to reap the benefits of this advantage for larger instances. With respect to consistency, SLPSOK2 provided consistent results with lesser standard deviation score compared to other algorithms in most cases except two (problem sets 6 and 11), thus signifying the superiority of solution over its hybrid counterparts with respect to both solution accuracy and precision. Based on the above

Table 2. K2 scores, Execution times and Best preference orders obtained using GAK2, ACOK2, PSOK2 and SLPSOK2 for problem sets 1 to 6.

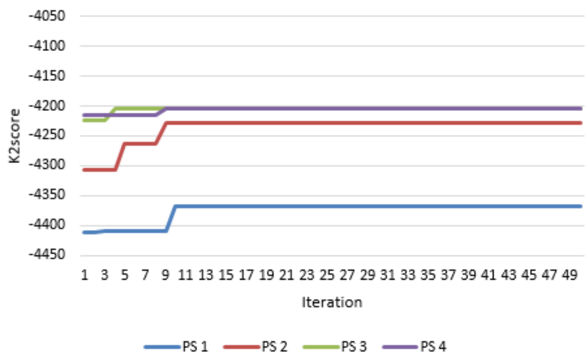
Problem set	Algorithm	Highest K2score (HKS)	Lowest K2score (LKS)	Average K2score (AKS)	Converged K2score (CKS)	Std. Dev. K2score (SDKS)	CPU Time (sec)	Clock Time (sec)	Average Execution Time (sec)	Best Preference Order
1	GAK2	-4367.05	-4367.05	-4367.05	-4367.05	0	17.50	15.00	16.25	25143
	ACOK2	-4367.05	-4367.05	-4367.05	-4367.05	0	15.11	13.95	14.53	25143
	PSOK2	-4367.05	-4367.05	-4367.05	-4367.05	0	19.61	18.13	18.87	25143
	SLPSOK2	-4367.05	-4367.05	-4367.05	-4367.05	0	11.16	9.06	10.11	25143
2	GAK2	-4228.16	-4228.16	-4228.16	-4228.16	0	247.30	223.16	235.23	25413
	ACOK2	-4228.16	-4228.16	-4228.16	-4228.16	0	189.45	182.80	186.13	25413
	PSOK2	-4228.16	-4228.16	-4228.16	-4228.16	0	250.2	228.78	239.51	25413
	SLPSOK2	-4228.16	-4228.16	-4228.16	-4228.16	0	157.69	154.81	156.25	25413
3	GAK2	-4203.75	-4203.75	-4203.75	-4203.75	0	425.13	388.77	406.95	25413
	ACOK2	-4203.45	-4203.45	-4203.45	-4203.45	0	292.00	279.19	285.60	25413
	PSOK2	-4203.45	-4203.45	-4203.45	-4203.45	0	430.08	393.95	412.01	25413
	SLPSOK2	-4203.45	-4203.45	-4203.45	-4203.45	0	252.16	248.72	250.44	25413
4	GAK2	-4203.75	-4203.75	-4203.75	-4203.75	0	503.92	462.96	483.44	25143
	ACOK2	-4203.45	-4203.45	-4203.45	-4203.45	0	377.41	363.88	370.65	25143
	PSOK2	-4203.75	-4203.75	-4203.75	-4203.75	0	547.13	500.05	523.59	25143
	SLPSOK2	-4203.45	-4203.45	-4203.45	-4203.45	0	203.06	184.09	193.58	25143
5	GAK2	-7019.56	-7019.56	-7019.56	-7019.56	0	32.70	28.56	30.63	2713654
	ACOK2	-7019.56	-7019.56	-7019.56	-7019.56	0	18.72	16.83	17.77	2746153
	PSOK2	-7019.56	-7019.56	-7019.56	-7019.56	0	39.27	37.28	38.28	2761354
	SLPSOK2	-7019.56	-7019.56	-7019.56	-7019.56	0	18.08	16.08	17.08	2743651
6	GAK2	-6723.79	-6741.51	-6732.66	-6723.79	6.5100	422.77	392.71	407.74	2763154
	ACOK2	-6712.54	-6715.54	-6714.5	-6714.95	1.1543	298.29	282.94	290.62	2136754
	PSOK2	-6718.75	-6722.75	-6721	-6721.17	1.4328	582.53	540.72	561.62	2367451
	SLPSOK2	-6714.95	-6718.4	-6716.18	-6714.95	1.3051	277.04	254.18	265.61	2136745

Table 3. K2 scores, Execution times and Best preference orders obtained using GAK2, ACOK2, PSOK2 and SLPSOK2 for problem sets 7 to 12.

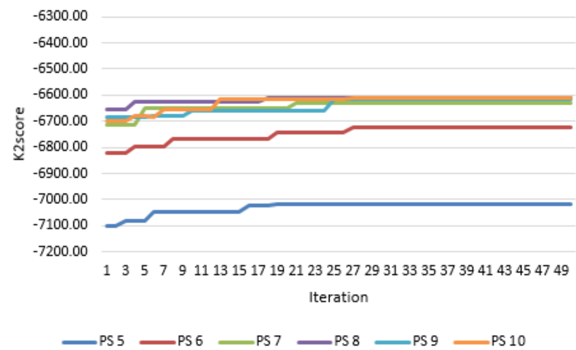
Problem set	Algorithm used	Highest K2score (HKS)	Lowest K2score (LKS)	Average K2score (AKS)	Converged K2score (CKS)	Std. Dev. K2score (SDKS)	CPU Time (sec)	Clock time (sec)	Average Execution Time (sec)	Best Preference Order
7	GAK2	-6627.53	-6631.28	-6629.26	-6628.43	1.5027	769.87	717.18	743.52	2475613
	ACOK2	-6627.39	-6631.39	-6630.03	-6627.39	1.5323	672.02	601.13	636.58	2457136
	PSOK2	-6628.42	-6631.66	-6629.96	-6628.92	1.5711	984.79	925.92	955.36	2741563
	SLPSOK2	-6625.54	-6626.18	-6625.94	-6625.54	0.3251	569.29	526.16	547.73	2741563
8	GAK2	-6604.59	-6612.63	-6609.55	-6608.59	3.3378	1245.31	1150.70	1198.01	2547136
	ACOK2	-6608.12	-6616.36	-6612.69	-6608.12	2.9969	1013.96	971.16	992.56	2571346
	PSOK2	-6617.89	-6625.15	-6623.05	-6622.89	3.0282	2461.98	2314.80	2388.39	2571346
	SLPSOK2	-6608.12	-6611.71	-6610.44	-6608.12	1.4162	953.54	886.97	920.26	2571346
9	GAK2	-6614.32	-6657	-6635.06	-6614.32	15.3840	1271.86	1194.39	1233.13	2713456
	ACOK2	-6607.41	-6611.12	-6609.09	-6608.12	1.4221	1298.01	1254.68	1276.34	2571346
	PSOK2	-6608.43	-6611.43	-6609.67	-6608.53	1.6072	2047.59	1905.73	1976.66	2517436
	SLPSOK2	-6605.92	-6608.92	-6608.02	-6608.12	1.2329	1237.50	1198.08	1217.79	2571346
10	GAK2	-6610.65	-6615.22	-6612.41	-6609.65	2.2125	1541.03	1430.07	1485.55	2157346
	ACOK2	-6608.12	-6610.99	-6609.55	-6608.12	1.0180	1389.43	1341.09	1365.26	2571436
	PSOK2	-6610.35	-6612.63	-6611.26	-6612.63	1.1312	1913.99	1899.00	1906.49	2547613
	SLPSOK2	-6609.24	-6609.51	-6609.33	-6609.41	0.1256	1092.23	1102.67	1097.45	2547613
11	GAK2	-9646.28	-9646.28	-9646.28	-9646.28	0	49.70	45.36	47.53	234976518
	ACOK2	-9646.28	-9646.28	-9646.28	-9646.28	0	45.13	42.28	43.70	273495168
	PSOK2	-9646.28	-9647.91	-9647.08	-9646.28	0.7963	50.68	52.05	51.37	273891654
	SLPSOK2	-9646.28	-9647.13	-9646.81	-9646.28	0.4396	34.54	30.99	32.77	273891654
12	GAK2	-8982.35	-8998.42	-8993.19	-8994.35	6.2879	766.88	712.66	739.77	297435681
	ACOK2	-8980.02	-8992.61	-8983.74	-8980.02	5.3723	627.68	597.02	612.35	275913864
	PSOK2	-8968.29	-8980.92	-8976.25	-8978.19	5.5189	1000.71	1071.42	1036.06	236987541
	SLPSOK2	-8991.32	-8992.64	-8992.13	-8991.32	0.6965	538.60	510.65	524.62	236987541

Table 4. K2 scores, Execution times and Best preference orders obtained using GAK2, ACOK2, PSOK2 and SLPSOK2 for problem sets 13 to 18.

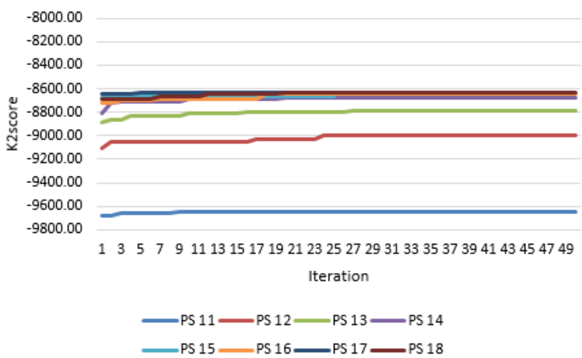
Problem set	Algorithm used	Highest K2score (HKS)	Lowest K2score (LKS)	Average K2score (AKS)	Converged K2score (CKS)	Std. Dev. K2score (SDKS)	CPU Time (sec)	Clock time (sec)	Average Execution Time (sec)	Best Preference Order
13	GAK2	-8783.19	-8785.31	-8784.02	-8785.21	1.13434	1440.64	1349.48	1395.06	295748136
	ACOK2	-8784.21	-8785.21	-8784.77	-8784.57	0.43272	1240.99	1180.75	1210.87	217493685
	PSOK2	-8784.77	-8786.1	-8785.59	-8784.77	0.70198	2642.70	2598.01	2620.36	247915368
	SLPSOK2	-8784.45	-8785.28	-8784.97	-8784.45	0.42868	1103.88	1006.53	1055.20	247915368
14	GAK2	-8672.39	-8676.07	-8674.78	-8675.29	1.56561	2645.84	2452.48	2549.16	274859316
	ACOK2	-8667.8	-8671.29	-8669.9	-8667.8	1.33904	1929.48	1850.51	1889.99	294783651
	PSOK2	-8663.94	-8667.55	-8666.59	-8667.34	1.53539	4182.68	4283.91	4233.29	271839456
	SLPSOK2	-8654.42	-8654.96	-8654.76	-8654.42	0.27070	1610.54	1500.19	1555.36	254978316
15	GAK2	-8659.47	-8663.47	-8662.15	-8662.11	1.63343	3461.48	3192.18	3326.83	284593761
	ACOK2	-8634.62	-8637.6	-8636.25	-8636.72	1.48860	2950.79	2815.52	2883.16	259347186
	PSOK2	-8632.72	-8635.97	-8633.99	-8632.72	1.59162	5136.48	4783.60	4960.04	215794836
	SLPSOK2	-8621.76	-8624.76	-8622.42	-8621.85	1.31268	2453.55	2270.73	2362.14	254971836
16	GAK2	-8620.72	-8640.78	-8636.34	-8638.62	8.77957	4029.24	3775.10	3902.17	294837165
	ACOK2	-8631.22	-8651.57	-8640.69	-8631.22	7.46187	3595.01	3399.00	3497.01	241573986
	PSOK2	-8615.62	-8632.15	-8628.52	-8630.52	7.24466	6010.64	5832.56	5921.60	295473816
	SLPSOK2	-8621.02	-8629.98	-8627.98	-8628.92	3.91552	2949.43	2728.45	2838.94	251473986
17	GAK2	-8628.18	-8634.51	-8632.76	-8632.08	2.76705	5433.39	5017.97	5225.68	275934816
	ACOK2	-8629.05	-8631.89	-8630.77	-8629.05	1.52855	4701.64	4508.06	4604.85	259714836
	PSOK2	-8628.92	-8632.96	-8631.36	-8628.92	2.18570	7408.30	6878.90	7143.60	251473986
	SLPSOK2	-8632.86	-8633.49	-8633.26	-8632.86	0.31964	4019.56	3892.45	3956.01	273491586
18	GAK2	-8628.26	-8633.42	-8630.88	-8632.86	2.42794	6480.96	6179.63	6330.30	273491865
	ACOK2	-8628.28	-8634.29	-8631.08	-8629.28	2.39103	5271.01	5075.50	5173.25	254971368
	PSOK2	-8634.47	-8640.47	-8636.74	-8634.47	2.41850	7673.09	7650.19	7661.64	251798364
	SLPSOK2	-8627.28	-8629.68	-8628.84	-8627.28	1.04307	4108.37	3936.22	4022.29	257913486



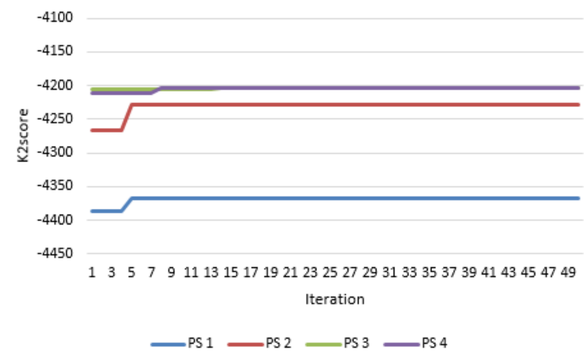
(a) 5 node problem with GAK2



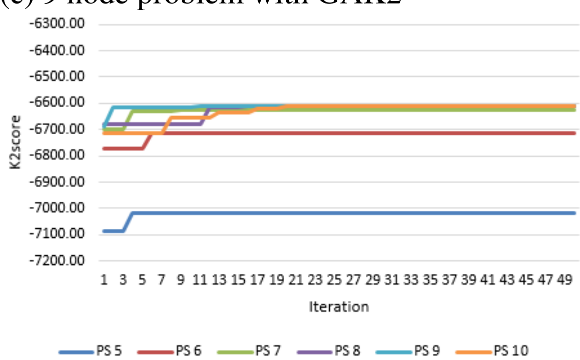
(b) 7 node problem with GAK2



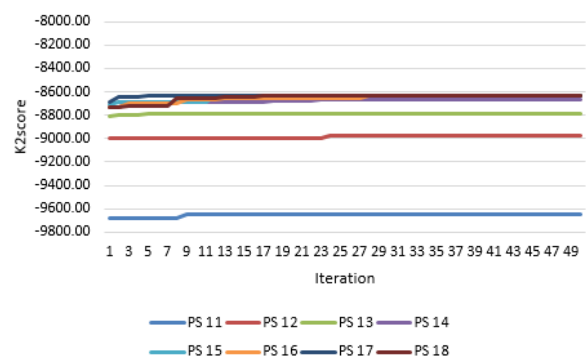
(c) 9 node problem with GAK2



(d) 5 node problem with ACOK2



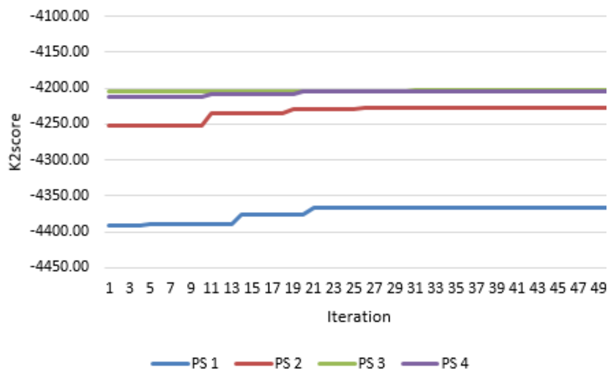
(e) 7 node problem with ACOK2



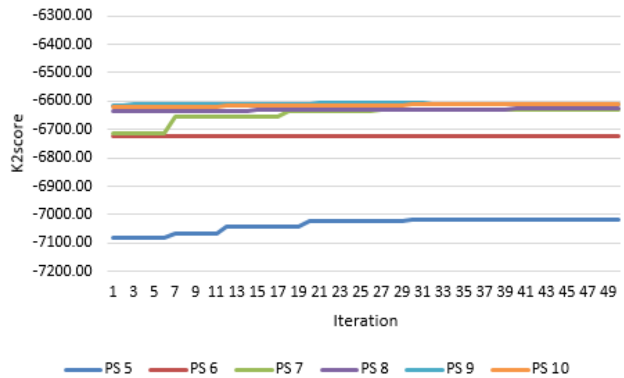
(f) 9 node problem with ACOK2

*PS- Problem Set

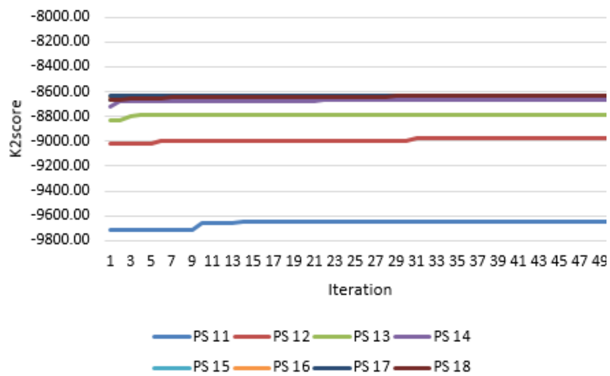
Figure 4. Convergence of K2score for GAK2 and ACOK2.



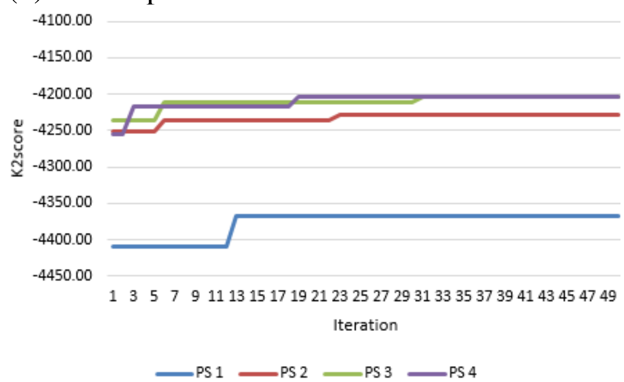
(a) 5 node problem with PSOK2



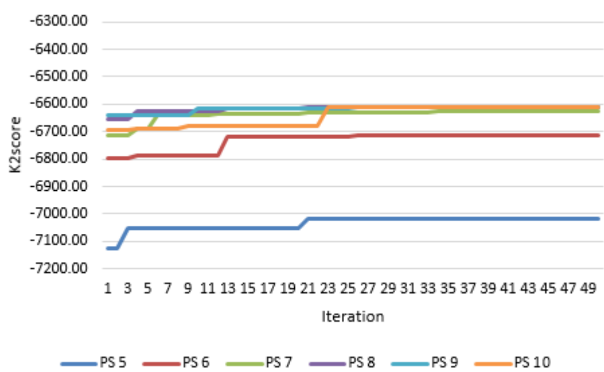
(b) 7 node problem with PSOK2



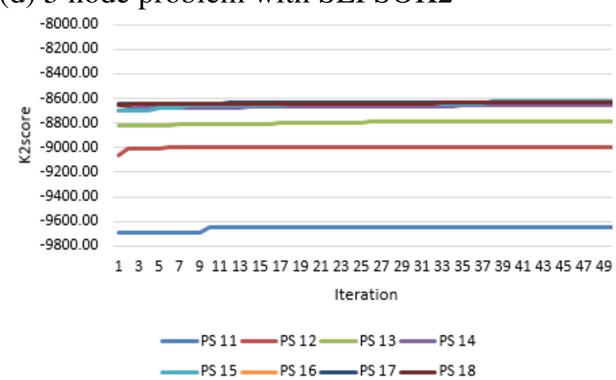
(c) 9 node problem with PSOK2



(d) 5 node problem with SLPSOK2



(e) 7 node problem with SLPSOK2



(f) 9 node problem with SLPSOK2

*PS- Problem Set

Figure 5. Convergence of K2score for PSOK2 and SLPSOK2

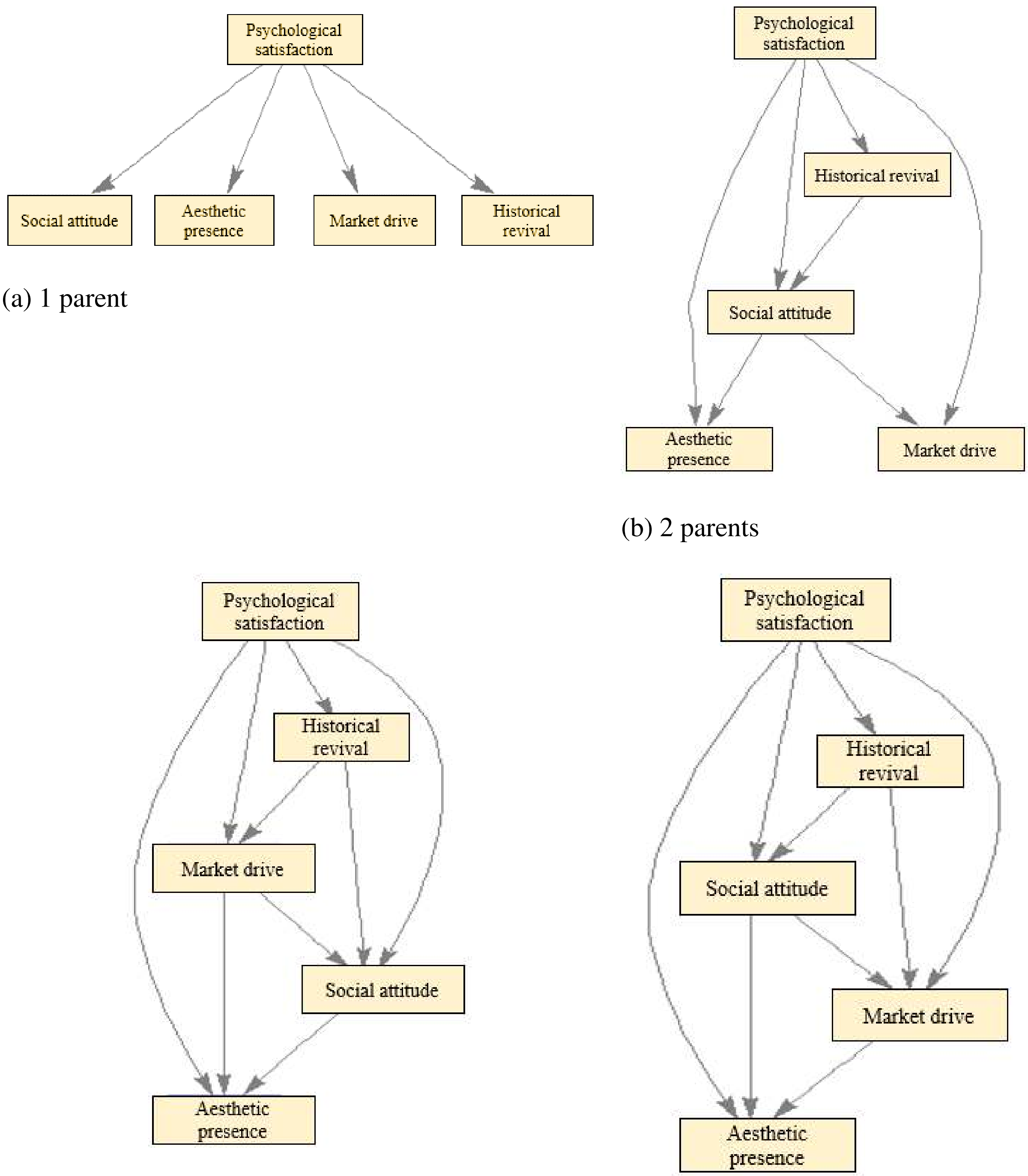
Table 5. Summary of results for problem sets 1-18.

Problem set	GAK2		ACOK2		PSOK2		SLPSOK2		Highest Converged K2score (CKS)	Minimum Std. Dev. K2score (SDKS)
	Highest preference	Lowest preference	Highest preference	Lowest preference	Highest preference	Lowest preference	Highest preference	Lowest preference		
1	2	3	2	3	2	3	2	3	All	All
2	2	3	2	3	2	3	2	3	All	All
3	2	3	2	3	2	3	2	3	ACOK2, PSOK2 & SLPSOK2	All
4	2	3	2	3	2	3	2	3	ACOK2 & SLPSOK2	All
5	2	4	2	3	2	4	2	1	All	All
6	2	4	2	4	2	1	2	5	ACOK2 & SLPSOK2	ACOK2
7	2	3	2	6	2	3	2	3	SLPSOK2	SLPSOK2
8	2	6	2	6	2	6	2	6	ACOK2 & SLPSOK2	SLPSOK2
9	2	6	2	6	2	6	2	6	ACOK2 & SLPSOK2	SLPSOK2
10	2	6	2	6	2	3	2	3	ACOK2	SLPSOK2
11	2	8	2	8	2	4	2	4	All	GAK2 & ACOK2
12	2	1	2	4	2	1	2	1	PSOK2	SLPSOK2
13	2	6	2	5	2	8	2	8	SLPSOK2	SLPSOK2
14	2	6	2	1	2	6	2	6	SLPSOK2	SLPSOK2
15	2	1	2	6	2	6	2	6	SLPSOK2	SLPSOK2
16	2	5	2	6	2	6	2	6	SLPSOK2	SLPSOK2
17	2	6	2	6	2	6	2	6	PSOK2	SLPSOK2
18	2	5	2	8	2	4	2	6	SLPSOK2	SLPSOK2

discussion, the least preferred criteria useful for the design of fashion products are considered according to results obtained by SLPSOK2. For smaller problems aesthetic presence was found to be the least preferred criteria, whereas for problems larger in size positive opinion of other customers mattered the least. The most frequently observed resulting structural relationships between the various attributes of the German fashion media context can be seen in Figures 6-9 shown for different problem sets.

6.2.1 5 node scenario

In this scenario the social media fashion data is annotated against the first five variables (attributes) of Table 1. Each variable or attribute is considered as a node in the Bayesian network. From Figs. 6(a)-(d) it can be seen that in majority of the cases node 2 (psychological satisfaction) acts as the most important attribute that drives the marketing strategy for fashion blogging. The first four problem sets correspond to this scenario. For all the cases, it is observed that the attributes are directly related to psychological satisfaction. However, remaining attributes behave differently in each case. Due to the restriction on maximum parent size, some of the dependencies are ignored and are assumed to be absent in the market. However, they appear in further problem sets. For the first case, remaining four attributes are found to participate independently. For the second case, aesthetic presence and market drive depend up on social attitude. Social attitude in turn directly depends up on the historical revival. Here, aesthetic presence and market drive are found to behave independently. Problem sets 3 and 4 exhibit similar structural relationships. In addition to the dominance of psychological satisfaction over other attributes, market drive is given higher priority over social attitude for problem set 3. The remaining relationships are as presented in the Figs. 6(a)-(d). It can be noticed that aesthetic presence has least preference in all the cases.



(a) 1 parent

(b) 2 parents

(c) 3 parents

(d) 4 parents

Figure 6. Attribute structural relationship for 5 node scenario.

6.2.2 7 node scenario

Seasonal variation may induce significant changes in customer preference over other attributes in the market. Also, positive opinioned reviews sometimes effect the decision made by a customer. The seven node scenario investigates the behavior of variables in the presence of the two additional attributes. Problem sets (5-10) deal with the current scenario. One can interpret from Figs. 7(a)-(e) that psychological satisfaction continues to dominate other attributes for this case. Historical and seasonal information are the next preferred attributes amongst remaining so much so that reviews with seasonal information are given higher preference than positive opinioned reviews. Problem sets 8 and 9 observe to have identical structural relationship amongst variables with the same preference order (Fig. 7(d)). Market drive, historical revival and positive opinion are found to be highly dependent on other attributes and independent with respect to each other. Figs. 7(a)-(e) give a visual illustration of dependencies observed for this scenario.

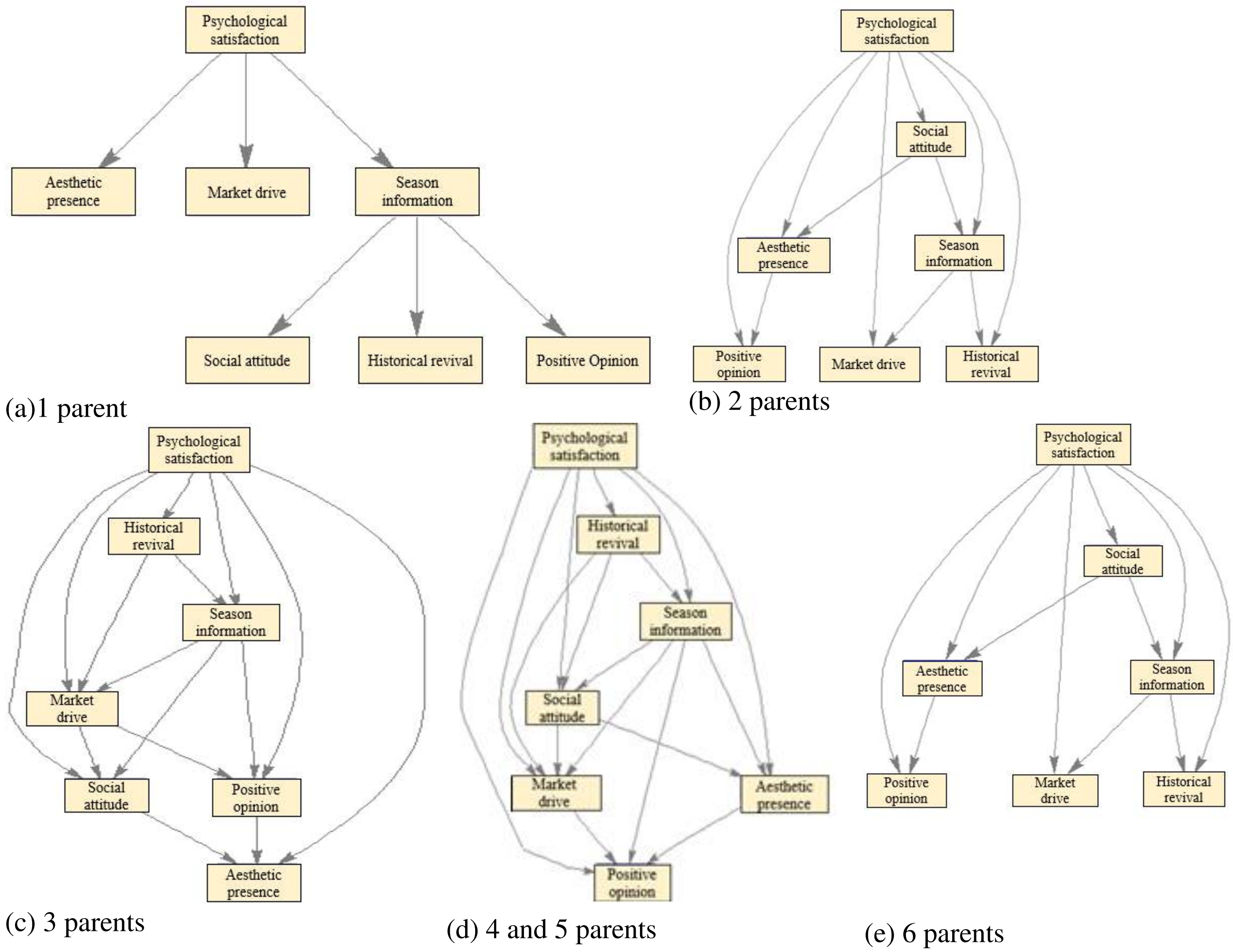
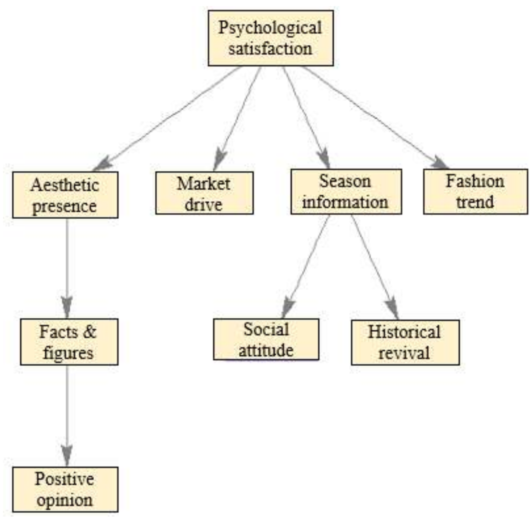


Figure 7. Attribute structural relationship for 7 node scenario.

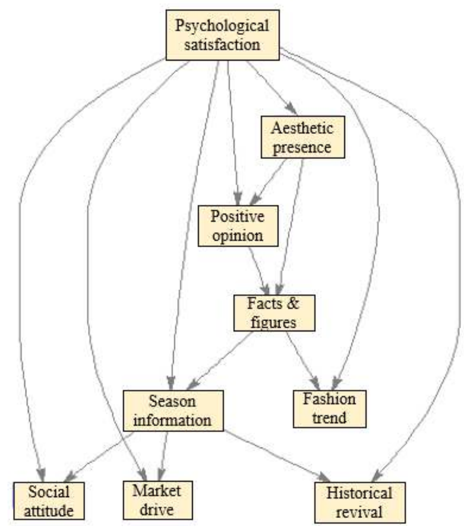
6.2.3 9 node scenario

The third and final scenario aims to capture the effect of reviews that contain information about younger fashion trends and facts and figures on the underlying relationship structure. In other words, this scenario captures the behavior of cognitive satisfaction and emotional satisfaction simultaneously. Problem sets (11-18) correspond to this case. Here again, physiological satisfaction and positive opinion are the highest and lowest preferred attributes respectively. Historical revival is found to be the second highest preferred attribute in majority of the cases.

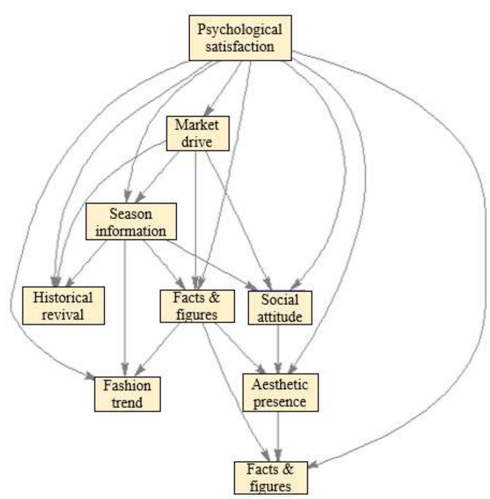
Further, group of variables in problem sets 16 and 17 are made up of identical tree structure (Fig. 7(f)). As the number of maximum parent size is allowed to increase, a lesser change in underlying tree structure is observed. The number of independent attributes are found be less in this case for any problem set. This scenario poses a dense structure of relationships relative to previous scenarios. The visual illustration is shown in Figs. 8 and 9.



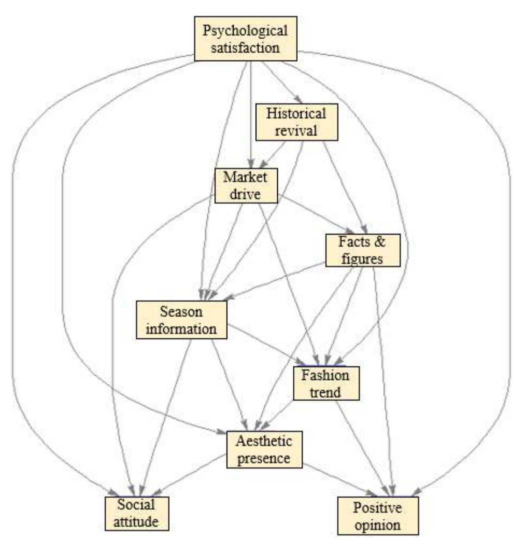
(a) 1 parent



(b) 2 parents

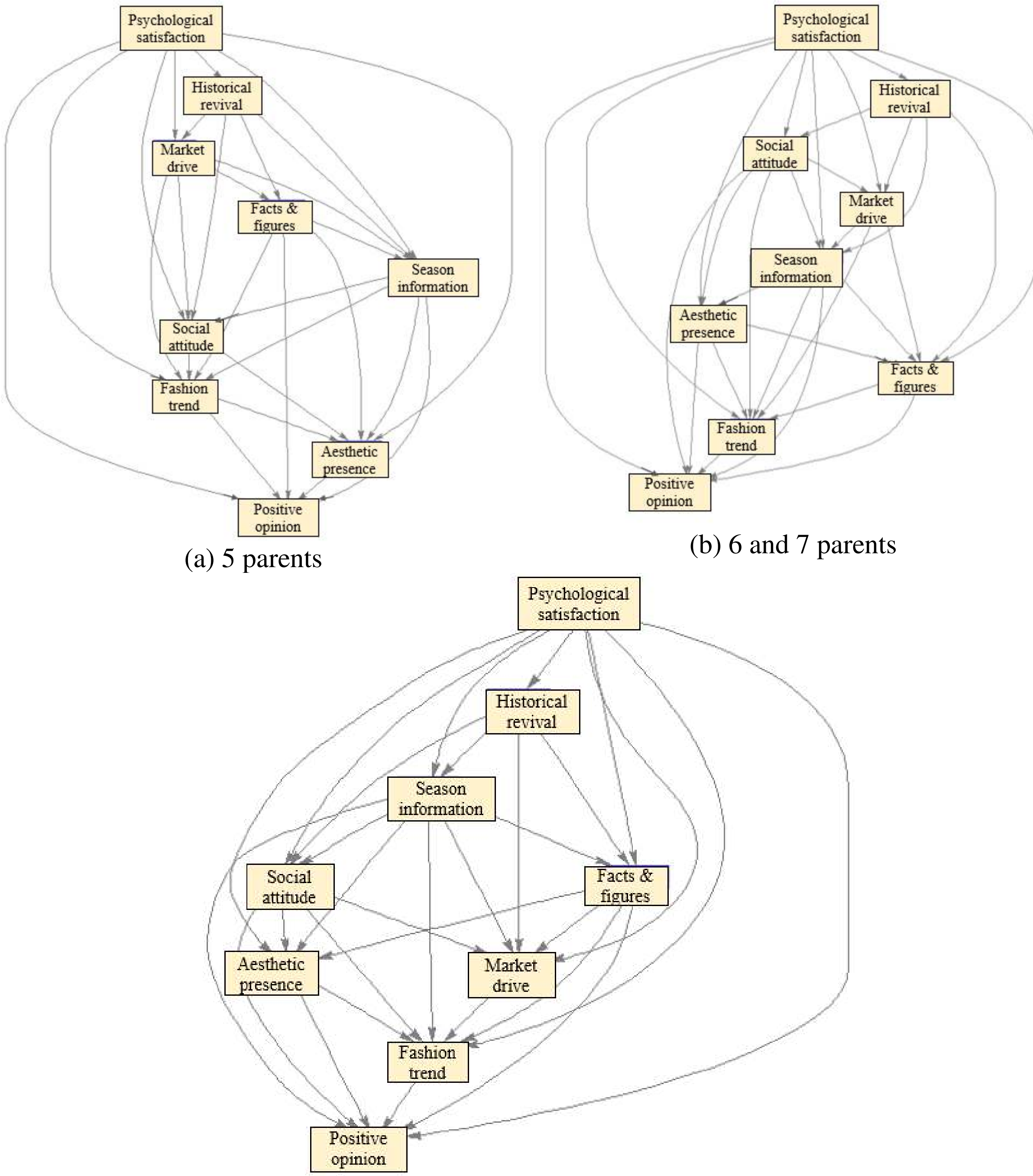


(c) 3 parents



(d) 4 parents

Figure 8. Attribute structural relationship for 9 node scenario.



(a) 5 parents

(b) 6 and 7 parents

(c) 8 parents

Figure 9. Attribute structural relationship for 9 node scenario.

Previous works in the domain present methodologies that used derived knowledge from search attribute information to gain advantage over competitors in online and traditional supermarkets (Degeratu et al. 2000). Further, the customer preference ordering information can be constructively used for selection of attributes in the planning of customer driven product design (Jin et al. 2016b; Liu et al. 2011) which would enable the product designers to relate to customer choice at an early stage.

7. Managerial implications

In globalization era blogger based marketing plays a bigger role in forestalling development of the organization moving towards strategic development and market development. The increasing use of ICT based technology like Ipads, IPhones, tablets, notepads and other digital devices has multiplied the need for online blogging as an effective media to get connected and attract new customers and paves way to penetrate the global market. Henceforth, supporting the wide spread use of online blogs as a marketing tool for advertisement and conveying information regarding unique features of new products and services, the study provides a unique opportunity to explore the desirable features in line with customer demands and needs. The outcome of the research leads to information and actionable knowledge that can be useful for decision makers to make an effective marketing strategy. The focal point of our work is to formulate the online review based fashion marketing and sales platform as a Bayesian Network Structure Learning problem and has been addressed by a novel hybrid metaheuristic to prioritize real time attributes that define the customer preference based on deeper comprehension of their structural relationships.

The findings of the paper aligns with concepts of marketing theory, consumer behavior theory and psychological theories. In this regard, the general marketing theory emphasizes on the value and psychological parameters that drive the customer need and demand. The study also revealed that psychological satisfaction, historical revival, seasonal information and facts and figure based reviews are major components of information in fashion blogs promoting the success of a blogger based effective advertising platform. People's preference are highly based on the cultural and historical heritage of the region. Furthermore, the findings of this study are supported by well-known theories such as the McCarthy's extended 7 P theory (Perreault and McCarthy 2002), and social exchange theory (Posey et al. 2010), and the Luhmannian social systems theory (Herting and Stein 2007).

8. Conclusion and future work

The recent past has witnessed a dramatic growth of natural language text data, including web pages, news articles, scientific literature, emails, enterprise documents, and social media such as blog articles, forum posts, product reviews, and tweets. Text data are unique in the sense that they are usually generated directly by humans rather than a computer system or sensors, and are thus especially valuable for discovering knowledge about people's opinions and preferences, in addition to many other kinds of information that we encode in text. In this context, this research attempts to exploit the abundantly available text data coming from online social media blogs in Germany. The problem is formulated as a Bayesian Network Structure Learning (BNSL) problem with multiple product attributes as nodes and their corresponding parent-child relationships as directed arcs connecting the nodes. The proposed NP-hard problem is solved by using four hybrid metaheuristics GAK2, ACOK2, PSOK2 and SLPSOK2 tailored as applicable to the context. Here, it is found that on an average, the novel integrated approach, SLPSOK2 outperforms GAK2, ACOK2, and PSOK2 by 31%, 17%, and 49% respectively with respect to computational time . All the approaches return almost identical K2scores with SLPSOK2 performing slightly better validating the global optima. Nine fashion attributes are analyzed under 18 problem sets covering three scenarios to understand the underlying structural relationships between variables (attributes). Experimentations on German case study reveal that psychological satisfaction, historical revival, seasonal information and facts and figure based reviews are major components of information in fashion blogs that influence customers.

The present work can be extended to learn structures with tertiary and higher degree discrete state variables. However increasing the number of a states of a node drastically increases the computational time to arrive at optimal solutions. Thus, this research provides the future possibility to explore better machine learning algorithms to solve computationally complex problems. To conduct a cross cultural analysis of multiple market scenario using multiple agent based optimization and simulation approaches is also an opportunity for future research in this area.

Acknowledgement

This research is conducted under a collaborative effort made by Department of Industrial & Systems Engineering IIT Kharagpur, West Bengal, India and SCI-STI-DK group of Institute of Mechanical Engineering IGM, EPFL, Switzerland. The authors are also thankful to Prof. Klaus-Dieter Thoben, BIBA, Germany, for facilitating real time data source to carry out the research. It was supported and funded by an ongoing European project FALCON bearing grant agreement number 636868.

References

Alkahtani, M., A. Choudhary, A. De, and J. A. Harding. 2018. “A decision support system based on ontology and data mining to improve design using warranty data.” *Computers & Industrial Engineering*, Accepted.

Aouay, S., S. Jamoussi, and Y. B. Ayed. 2013. “Particle Swarm Optimization Based Method for Bayesian Network Structure Learning.” In *2013 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO)*, 28-30 Apr, Hammamet, Tunisia, 1-6.

Askari, M. B. A., and M. G. Ahsaee. 2018. “Bayesian network structure learning based on cuckoo search algorithm” In *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, 28 Feb – 2 Mar, Kerman, Italy, 127-130.

Cano, A., M. G. -Olmedo, A. R. Masegosa, and S. Moral. 2013. "Locally Averaged Bayesian Dirichlet Metrics for Learning the Structure and the Parameters of Bayesian Networks." *International Journal of Approximate Reasoning* 54 (4): 526-540.

Cao, Q., and M. J. Schniederjans. 2004. "Empirical Study of the Relationship between Operations Strategy and Information Systems Strategic Orientation in an E-Commerce Environment." *International Journal of Production Research* 42 (15): 2915-2939.

Chaudhry, S. S., and W. Luo. 2005. "Application of genetic algorithms in production and operations management: A review.", *International Journal of Production Research* 43 (19): 4083-4101.

Chen, H. X., Q. Zheng, T. Lei, and S. L. –Peng. 2007 "Learning Bayesian Network structures with discrete particle swarm optimization algorithm." *IEEE Symposium on Foundations of Computational Intelligence*, 47-52.

Chickering, D. M., D. Heckerman, and C. Meek. 2004. "Large-Sample Learning of Bayesian Networks Is NP-Hard." *The Journal of Machine Learning Research* 5: 1287-1330.

Cho, S., H. –B. Jun, and D. Kiritsis. 2016. "Heuristic Algorithms for Maximising the Total Profit of End-of-Life Computer Remanufacturing." *International Journal of Production Research* 55 (5): 1350-1367.

Chong, A. Y. –L., F. T. S. Chan, M. Goh, and M.K. Tiwari. 2013. "Do Interorganisational Relationships and Knowledge-Management Practices Enhance Collaborative Commerce Adoption?" *International Journal of Production Research* 51 (7): 2006-2018.

Chong, A. Y. L., E. Ch'ng, M. J. Liu, and B. Li. 2015. "Predicting Consumer Product Demands via Big Data: The Roles of Online Promotional Marketing and Online Reviews." *International Journal of Production Research*, 55 (17): 5142-5156.

Colorni, A., M. Dorigo, and V. Maniezzo. 1992. "Distributed optimization by ant colonies*." First European Conference on Artificial Life*, 134-142.

Cooper, G. F., and E. Herskovits. 1992. "A Bayesian Method for the Induction of Probabilistic Networks from Data." *Machine Learning* 9 (4): 309-347.

Degeratu, A. M., and A. Rangaswamy, J. Wu. 2000. "Consumer choice behaviour in online and traditional supermarkets: The effects of brand name, price, and other search attributes." *International Journal of Research in Marketting* 17: 55-78.

Dolgui, A., D. Ivanov, and B. Sokolov. 2017. "55th Anniversary of Production Research." International Journal of Production Research 1–2.

Franceschetti, A., E. Demir, D. Honhon, T. V. Woensel, G. Laporte, and M. Stobbe. 2017. "A metaheuristic for the time-dependent pollution-routing problem." *European Journal of Operations Research* 259 (3): 972-991.

Garces, G. A., A. Rakotondranaivo, and E. Bonjour. 2016. "Improving users' product acceptability: an approach based on Bayesian networks and a simulated annealing algorithm." *International Journal of Production Research* 54 (17): 5151-5168.

Gheisari , S., and M. R. Meybodi. 2016. "BNC-PSO: Structure learning of Bayesian Networks by particle swarm optimization." *Information Sciences* 348: 272-289.

Goldberg, D. 1989. "Genetic algorithms in search, optimization, and machine learning." Addison-Wesley: MA.

Herting, S., and L. Stein. 2007. “The Evolution of Luhmann’s Systems Theory with Focus on the Constructivist Influence.” *International Journal of General Systems* 36 (1): 1-17.

Huang, Y. –S., C. –S. Hsu, and J.-W. Ho. 2014. “Dynamic pricing for fashion goods with partial backlogging” *International Journal of Production Research* 52 (14): 4299-4314.

Injadat, M., F. Salo, and A. B. Nassif. 2016. “Data Mining Techniques in Social Media: A Survey.” *Neurocomputing* 214: 654-670.

Ji, J., C. Yang, J. Liu, J. Liu, and B. Yin. 2017. “A comparative study on swarm intelligence for structure learning of Bayesian networks.” *Soft Computing* 21 (22): 6713-6738.

Ji, J., R. Hu, H. Zhang, and C. Liu. 2011. “A Hybrid Method for Learning Bayesian Networks Based on Ant Colony Optimization.” *Applied Soft Computing* 11: 3373-3384.

Jiang, Y., Z. Liang, H. Gao, Y. Guo, Z. Zhong, C. Yang, and J. Liu. 2018. “An improved constraint-based Bayesian network learning method using Gaussian kernel probability density estimator.” *Expert Systems With Applications* 113: 544-554.

Jin, J., P. Ji, and R. Gu. 2016a. “Identifying comparative customer requirements from product online reviews for competitor analysis.” *Engineering Applications of Artificial Intelligence* 49: 61-73.

Jin, J., Y. Liu, J. Ping, and H. Lui. 2016b. “Understanding Big Consumer Opinion Data for Market-Driven Product Design.” *International Journal of Production Research* 54 (10): 3019–3041.

Kennedy, J, and R. Eberhart. 1995. “Particle Swarm Optimization.” *Neural Networks, 1995. Proceedings., IEEE International Conference on* 4: 1942–1948.

Krishna, G. J., and V. Ravi. 2016. “Evolutionary computing applied to customer relationship management: A survey.” *Engineering Applications of Artificial Intelligence* 56: 30-59.

Larrafiaga, P., Mike P., Y. Yurramendi, R. H. Murga, and C. M. H. Kuijpers. 1996. “Structure Learning of Bayesian Networks by Genetic Algorithms : Performance Analysis of Control Parameters.” *IEEE Transactions on Patttern Analysis and Machine Intelligence* 18 (9): 912–926.

Larranaga, P., H. Karshenas, C. Bielza, and R. Santana. 2013. “A review on evolutionary algorithms in Bayesian Network learning and inference tasks.” Information Sciences 233: 109-125.

Li, C., S. Yang, and T. T. Nguyen. 2012. “A Self-Learning Particle Swarm Optimizer for Global Optimization Problems.” *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society* 42 (3): 627–646.

Li, H, and A. K. Parlikad. 2016. “Social Internet of Industrial Things for Industrial and Manufacturing Assets.”, *3rd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology*, Biarritz.

Li, X. –L., and X. –D. He. 2014. “A hybrid particle swarm optimization method for structure learning of probablisic relational models.” *Information Sciences* 283: 258-266.

Lin, C. –H. 2013. “A rough penalty genetic algorithm for constrained optimization.” *Information Sciences* 241: 119-137.

Liu, C., A. R. -Serrano, and G. Yin. 2011. “Customer-driven product design and evaluation method for collaborative design environments.” *Journal of Intelligent Manufacturing* 22: 751-764.

Mogale, D. G., S. K. Kumar, and M. K. Tiwari. 2018. "An MINLP model to support the movement and storage decisions of the Indian food grain supply chain." *Control Engineering Practice* 70: 98-113.

Mehrjoo, M., and Z. J. Pasek. 2016. "Risk assessment for the supply chain of fast fashion apparel industry: A system dynamics framework." *International Journal of Production Research* 54 (1): 28–48.

Ojha, R., A. Ghadge, M. K. Tiwari, and U. S. Bititci. 2018. "Bayesian network modelling for supply chain risk propagation." *International Journal of Production Research* 1-25.

Pavon, R., F. Diaz, and V. Luzon. 2008. "A model for parameter setting based on Bayesian Networks." *Engineering Applications of Artificial Intelligence* 21: 14-25.

Pearl, J., and T. S. Verma. 1995. "A theory of inferred causation." *Studies in Logic and the Foundations of Mathematics* 134: 789-811.

Perreault, W. D., and E. J. McCarthy. 2002. "Basic Marketing: A Global-Managerial Approach," 48–50.

Pitiot, P., T. Coudert, L. Geneste, and C. Baron. 2010. "Hybridization of Bayesian Networks and evolutionary algorithms for multi-objective optimization in an integrated product design and project management context." *Engineering Applications of Artificial Intelligence* 23: 830-843.

Posey, C., P. B. Lowry, T. L. Roberts, and T. S. Ellis. 2010. "Proposing the Online Community Self-Disclosure Model: The Case of Working Professionals in France and the U.K. Who Use Online Communities." *European Journal of Information Systems* 19 (2): 181–195.

Sheridan, M., C. Moore, and K. Nobbs. 2006. “Fast Fashion Requires Fast Marketing: The Role of Category Management in Fast Fashion Positioning.” *Journal of Fashion Marketing and Management* 10 (3): 301–315.

Shoham, A. 2003. “Determinants of fashion attributes’ importance.” *Journal of International Consumer Marketing* 15 (2): 43-61.

Neto, R. F. T., and M. G. Filho. 2013. “Literature review regarding ant colony optimization applied to scheduling problems: Guidelines for implementation and directions for future research.”, *Engineering Applications of Artificial Intelligence* 26: 150-161.

Wang, F., H. Zhang, K. Li, Z. Lin, J. Yang, and X. –L. Shen. 2018. “A hybrid particle swarm optimization algorithm using adaptive learning strategy.” *Information Sciences* 436: 162–177.

Wang, J., and S. Liu. 2018. “Novel binary encoding water cycle algorithm for solving Bayesian network structures learning problem.” *Knowledge-based Systems* 150: 95-110.

Wang, Y. –M. 2012. “Assessing the relative importance weights of customer requirements using multiple preference formats and nonlinear programming.” *International Journal of Production Research* 50 (16): 4414–4425.

Wong, M. L., W. Lam, and K. S. Leung. 1999. "Using evolutionary programming and minimum description length principle for data mining of Bayesian networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (2): 174-178.

Yang, C., J. Ji, J. Liu, J. Liu, and B. Yin. 2016. "Structural learning of Bayesian networks by bacterial foraging optimization." *International Journal of Approximate Reasoning* 69: 147-

167.

Yang, D., T. Xiao, T. –M. Choi, and T. C. E. Cheng. 2015. “Optimal reservation pricing strategy for a fashion supply chain with forecast update and asymmetric cost information.” *International Journal of Production Research* 52 (2): 1–22.

Zhang, X., S. Jia, X. Li, and C. Guo. 2018. “Learning the Bayesian Networks Structure based on Ant Colony Optimization and Differential Evolution.” In *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*, 20-23 Apr, Auckland, Newzealand, 354-358.

Zhou, Y., N. Fenton, and C. Zhu. 2016. "An empirical study of Bayesian network parameter learning with monotonic influence constraints." *Decision Support Systems* 87: 69-79.