Computers & Industrial Engineering 63 (2012) 1107–1117

Contents lists available at SciVerse ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

A case-based reasoning system development for statistical process control: Case representation and retrieval ☆

Morteza Behbahani [a,*], Abbas Saghaee [a], Rassoul Noorossana [b]

[a] Department of Industrial Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran
[b] Department of Industrial Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran

ARTICLE INFO

Article history:
Received 23 February 2011
Received in revised form 4 July 2012
Accepted 11 July 2012
Available online 10 September 2012

Keywords:
Case-based reasoning
Statistical process control
Case representation
Case retrieval
Knowledge Engineering

ABSTRACT

Statistical process control (SPC) is a sub-area of statistical quality control. Considering the successful results of the SPC applications in various manufacturing and service industries, this field has attracted a large number of experts. Despite the development of knowledge in this field, it is hard to find a comprehensive perspective or model covering such a broad area and most studies related to SPC have focused only on a limited part of this knowledge area. According to many implemented cases in statistical process control, case-based reasoning (CBR) systems have been used in this study for developing of a knowledge-based system (KBS) for SPC to organize this knowledge area. Case representation and retrieval play an important role to implement a CBR system. Thus, a format for representing cases of SPC and the similarity measures for case retrieval are proposed in this paper.

1. Introduction

Case-based reasoning (CBR) is a type of knowledge-based systems (KBS). In CBR, information is stored for later retrieval in the form of cases rather than rules (Kendal & Creen, 2007). When the user encounters a new problem, the system searches within the case base to retrieve cases similar to the new problem and adapts the solutions of the old cases to the new one (Liao, 2005).

Statistical process control (SPC) consists of methods for understanding, monitoring and improving process performance over time (Woodall, 2000). SPC was developed by Walter Shewhart during the 1920s. It is now realized that SPC is not just a collection of techniques, but a way of thinking about quality improvement, and it is regarded in many organizations as an important element of Total Quality Management (TQM) (Caulcutt, 1995).

Control charts are important features of SPC. They are used to monitor the process, to discover the causes of the problem and also to check process stability (Caulcutt, 1995). Considering the successful results of control chart applications in various manufacturing and service industries, this area has attracted the attention of quality engineers and statistics experts. Therefore, many investigations related to SPC have been published in the journals and presented in scientific conferences; hence, the knowledge area related to SPC has expanded far beyond what was introduced by Shewhart as the control charts.

Although knowledge in this field has been expanded tremendously, it is hard to develop a structured knowledge base to cover such a broad area. Therefore, some issues have been raised, such as the uneven growth of different parts of the SPC (e.g., the variable control charts have gained greater concern than the attribute control charts).

Furthermore, new techniques and topics discussed in some recent investigations have not been practical. In addition, there are ambiguities to recognize which investigations about SPC belong to what part of this area. It is difficult to find out whether the investigations made in one part of SPC could be effective in some other parts of this knowledge area. Thus, despite many studies in this field, some potential research opportunities still exist. The main reasons for the above-mentioned issues include:

- Lack of a comprehensive model and perspective on the statistical process control.
- Lack of a proper structure for knowledge storing in this field that can be updated.
- Failure to organize the existing knowledge in this area for future problem solving.

Knowledge Engineering (KE) was defined by Feigenbaum, McCorduck, and Nii (1988) as a discipline to integrate knowledge

☆ This manuscript was processed by Area Editor H.-S. Jacob Tsao.

* Corresponding author. Address: Shahid Kolahdooz Street, Motaghed Alley, No. 56, Postal Code 1939644613, Tehran, Iran. Tel.: +98 22007492; fax: +98 2122007492.

E-mail addresses: m.behbahani@srbiau.ac.ir (M. Behbahani), rassoul@iust.ac.ir (R. Noorossana).

0360-8352/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cie.2012.07.007

into computer systems by building, maintaining and developing knowledge-based systems to solve complex problems. With regard to the KE activities (Awad, 1996; Kendal & Creen, 2007), developing a KBS for SPC leads to results such as eliciting and structuring knowledge as well as storing knowledge in the knowledge-base. In addition, it makes statistical process control easily available for organizations in the case of future problem solvings. Therefore, it seems that developing a knowledge-based system for SPC can eliminate the reasons mentioned above. Several knowledge-based systems have been developed for statistical process control (e.g., Cheng & Hubele, 1992; Evans & Lindsay, 1987; Hosni & Elshennavy, 1988; Masud & Thenappan, 1993, etc.); however, most of them are rule-based systems which focus only on a limited part of SPC knowledge area (e.g., selecting the type of control chart). In addition, a comprehensive system to cover the entire field cannot be found. Considering the features of SPC knowledge area and many cases in this field, CBR systems can be appropriated. In summary, in this survey, at first, the knowledge-based systems already presented for statistical quality control (SQC) will be reviewed. Then, the reasons for choosing a CBR system to represent and organize SPC knowledge will be explained and finally, a framework for case representation and retrieval to implement a CBR system for SPC will be proposed.

2. Knowledge-based systems applications in statistical quality control

Several knowledge-based systems have been developed for SQC. Some of these systems have been categorized and illustrated in Table 1. While reviewing these systems, the following points were observed:

- Almost all systems presented so far for statistical quality control are rule-based.
- The systems developed for SPC have focused only on a limited part of this knowledge area (e.g., selecting the control chart type), and comprehensive systems were not found.
- These systems have little flexibility. They are just able to help solve problems that are exactly set in the defined borders and are not able to solve other issues related to SPC areas that are out of these borders.
- Furthermore, the mentioned rule-based systems do not perform well in solving the problems with missing or incomplete data.
- In addition, most of these systems were developed many years ago. Many new studies have led to rapid expansion of relevant knowledge, and these systems need to be updated.

Therefore, to organize the knowledge area of SPC, especially the applied knowledge for problem solving, a KBS with a comprehensive vision for SPC and also with more flexibility should be developed with the property of easy updating over time.

3. Why CBR systems?

CBR systems act based on the human reasoning model in solving problems. The core idea of CBR is "Similar problems have similar solutions" (Kolodner, 1993). For instance, when an experienced

Table 1
Some knowledge-based systems for statistical quality control.

Application area	Reference	Target(s)	Description
SPC	Evans and Lindsay (1987)	• Determining assignable cause • Interpreting control charts	SPC is considered as a process with the following steps: quality characteristic measurements, chart construction, control chart interpretation, determination of assignable cause and managerial action
	Hosni and Elshennavy (1988)	Control chart selection	The development of a quality control system based on knowledge for selecting control graphs, both by attributes and by variables
	Ntuen, Park, and Kim (1989)	Pattern recognition	An expert system which makes use of pattern recognition in order to 'visualize' pieces under inspection
	Cheng and Hubele (1992)	• Monitoring • Interpreting • Diagnosing	An expert system framework for problem-solving aspects of statistical process control
	Masud and Thenappan (1993)	Monitoring and designing control chart	Acknowledge advisory system to represent knowledge about the design and monitoring of control charts due to Effective statistical quality control programs
	El-Shal and Morris (2000)	Fault detection	A fuzzy expert system for fault detection in SPC in a manufacturing industry
	Tatara and Cinar (2002)	Monitoring and diagnosing multivariate control chart	A knowledge-based system for multivariate statistical process monitoring and diagnosis
Quality assurance	Crawford and Eyada (1989)	Allocation of resource planning	An expert system aimed at planning allocation of resources for quality assurance program
	Affisco and Chandra (1990)	Quality assurance purposes in production organizations	A conceptual model of an expert system for the quality assurance purposes
	Eyada (1990)	Auditing	An expert system developed for auditing procedures of quality assurance involving both suppliers and producers in the process.
	Brink and Mahalingam (1990)	Quality evaluation	An expert system which evaluates the quality at manufacturing level to detect and correct defects occurring during the production process
Inspection and sampling	Fard and Sabuncuoglu (1990)	Determining the type of sampling	An expert system which seeks to select the type of sampling by attributes
	Paladini (2000)	Type of inspection	A rule-based expert system to adopt the inspection type
Design of experiments (DOE)	Chen (1991)	To aim implementation of DOE	A rule-based expert system aimed to answer the questions about the implementation of DOE

doctor visits a new patient, he recalls similar patients he visited in the past and may use some of the previous treatments to treat the new patient. Many prototypes of CBR systems have been built in different areas, e.g., CASEY, CYRUS, PROTOS and HYPO (Kolodner, 1993). CBR systems are an alternative in many situations. There are some reasons indicating that a CBR system is more efficient than other types of knowledge-based systems for SPC. Some of these reasons are expressed below:

- One of the time-consuming activities in implementing and developing a rule-based system is extracting knowledge and representing it in IF/Then format. Therefore, developing a rule-based system for a widespread area like SPC, especially if it is going to cover all of this area, is an intricate and difficult work or may be impossible. In CBR systems, with storing knowledge in case format, the required intensity of the knowledge extraction process decreases significantly. Also, a CBR system can begin with little experience and few cases and can grow over time. Moreover, by storing knowledge in case format, both tacit and implicit knowledge can be used to solve SPC problems.
- CBR systems can also be used in solving the problems with missing or incomplete data, and it is an important advantage over other types of KBSs.
- Some questions about the characteristics of a knowledge area are introduced in Table 2. These questions were raised by Pal and Shiu (2004) and can be used to determine whether case-based reasoning is applicable or not for a specified knowledge area. Answers to the mentioned questions, as shown in Table 2, indicates that using CBR for the knowledge storing and problem solving in the area of SPC is suitable.

Considering the advantages of a CBR system over rule-based systems, which have been previously developed for SPC (mentioned in Section 2), developing a CBR system for SPC seems to be logical. In addition, using the CBR system, in comparison with other types of knowledge-based systems, has more advantages as pointed out by Pal and Shiu (2004).

4. Developing a CBR system for statistical process control

Although the concept of CBR is understandable, researchers and practitioners need a practical model to develop a CBR system for SPC. Aamodt and Plaza (1994) introduced a process model of CBR Cycle to provide a better understanding of CBR. This model is commonly called the R4, which constitutes the following four processes: retrieve, reuse, revise, and retain.

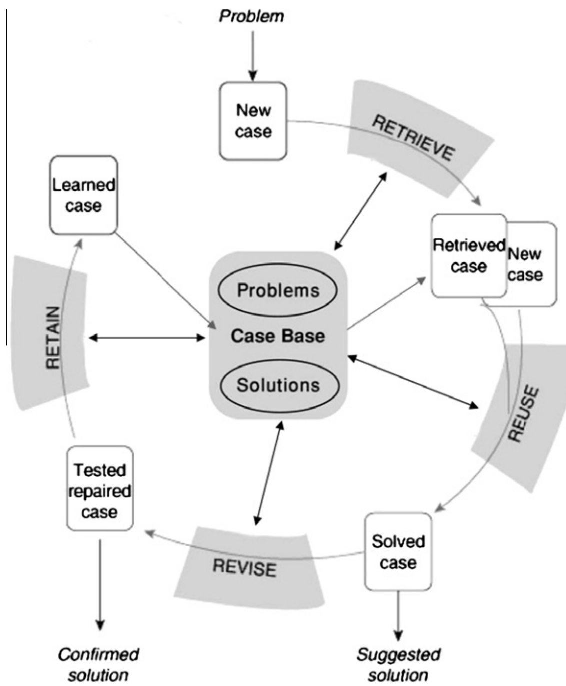

Fig. 1. Model R4 for CBR systems (Aamodt and Plaza).

As shown in Fig. 1, a new case is created by defining a new problem. This problem is used to *Retrieve* similar cases from the collection of previous cases. In CBR systems, the quality of the results mainly depends on the similarity measures used to retrieve similar cases. Through *Reuse*, the retrieved case is compounded with the new case for proposing a solution to the initial problem. Since it is usually unlikely that an exact match will occur in retrieval, the retrieved solution must be adapted. In the *Revise* process, the proposed solution is tested for example by being applied to the real-world environment or being evaluated based on general knowledge of SPC or by experts' judgments and the solution can be repaired if it fails. During the *Retain* process, useful experiences are retained for future reuse and the case base is updated with a new case. Cases are retained because they contain valuable knowledge or lessons; they are not just records of every event or experience.

5. Case representation

The main purpose of this paper is to create a base to store and use the knowledge of SPC to solve future problems by representing

Table 2
Comparison between features of SPC knowledge area and characteristics of candidate domains for CBR application.

Characteristics of candidate field	Description	Features of SPC
Does the domain have an underlying model?	If the domain is impossible to understand completely, CBR allows us to work on past experience without a complete understanding of the underlying mechanism	Considering the broadness of SPC area, it is so hard to provide a comprehensive underlying model for SPC and it is not found in our study
Are there exceptions and novel cases?	In domains where new experiences and exceptions are encountered, CBR system is an important alternative	Many cases are implemented in the field of SPC, and they are accessible in article databases
Do cases recur?	When experiences are similar enough to be compared and adapted successfully, it might be better to build a CBR system to derive the solution	In this filed the similar cases occur frequently
Is there significant benefit in adapting past solutions?	A CBR can be helpful where there is a significant benefit to create a solution through modifying a similar solution rather than creating a solution to a problem from scratch	Since the methods of SPC are based on statistical and mathematical science, using similar cases is helpful to find a problem solving method
Are relevant previous cases obtainable?	Is it possible to obtain characteristics of past cases? Do the recorded cases contain the relevant features of the problem influencing the outcome of the solution?	In this research, we are trying to present a structure that can represent the characteristics of SPC problems and case retrieval

and retrieving the SPC cases. The representation of cases is important for CBR. A case is a piece of knowledge representing an experience and typically comprises a problem and a solution. Thus, in many practical CBR applications, cases are usually represented as attribute–value pair that represents the problem and solution features. A collection of cases is called a *case base*, which can be described in terms of problem space and solution space (Fig. 1). The problem space is used for retrieving similar cases in the case base.

As Stoumbos, Reynolds, Ryan, and Woodall (2000) noted: "The increasing complexity of problems encountered provides an opportunity to narrow the gaps between applications and applied and theoretical SPC research." Therefore, as the problems become complicated, the theoretical studies will have more impact on applications for solving SPC problems. Thus, some studies with theoretical aspects in some day have found application aspects. According to the above-mentioned purpose of this CBR system, the theoretical knowledge in addition to the application area of SPC knowledge has also been considered. Therefore, the concept of a 'case' in this article includes practical cases and applied instances as well as theoretical studies that are available in solving applied problems of SPC.

The cases of SPC are usually stored in databases in an unstructured format. CBR uses indexes to represent cases and speed up retrieval. A flat case base is a common structure of case representation. In this method, indexes are chosen to represent the important aspects of the case, and retrieval involves comparing the query case features to each case in the case base. Case indexes should be predictive, address the purposes of the case used, be abstract enough to allow for widening the future use of the case base, and be concrete enough to be recognized in the future (Watson, 2003).

It is very difficult to make a decision about what features of a case must be represented and to decide about the indexing vocabulary which requires knowing what the important features of statistical process control are. Therefore, we reviewed some important sources of SPC to extract main features of SPC problems, considering that the problem space is used for case retrieval. Furthermore, some indexes to represent these important features were selected. The reviewed sources and the procedure to select the indexes will be explained in the next section. Also, with respect to this fact that different words are used for some concepts in SPC, defining and labeling these indexes cause unification of the literature related to SPC.

5.1. An overview of statistical process control

The purpose of this section is to provide an overview and to detect and justify some important features for case indexing. There are some ambiguities in this area. Sometimes, SPC is limited to and equated with control charts; on the other hand, there is a view that equates SPC with Total Quality Management (TQM). However, SPC should not be described in such general terms. Hardly any comprehensive descriptions of SPC approach can be found in literature. Some of these descriptions focus on the methodological and organizational aspects of implementing SPC (e.g., Berger & Hart, 1986; Does, Schippers, & Trip, 1997; Gaafar & Keats, 1992; Parks, 1983). In these studies, some steps have been determined for implementation of SPC approach, such as initial activities to develop an SPC system, process analysis and measurement system development, control charting, out of control action plan, and process capability analysis.

SPC is a practical approach for reducing variability of processes. Therefore, understanding the variation in the process is of primary importance in SPC. Control charts are used to detect variation due to an assignable cause. In many applications of control charting (e.g., ANSI/ASQ, 1996; Montgomery, 2009; Woodall, Spitzner, Montgomery, & Gupta, 2003, etc.), it is useful to distinguish between Phase I (retrospective phase) and Phase II (prospective phase) methods and applications. In Phase I, control chart limits are often calculated using parameters estimated from an in-control reference sample. In Phase I, the primary interest is to better understand the process and assess process stability and variability. In addition to the use of various statistical tools (e.g., graphs and testing hypotheses), control charts play an important role in Phase I analysis (Chakraborti, Graham, & Human, 2009). Juran and Godfrey (1998) have presented the following steps for developing and designing the control charts:

- Choose the quality characteristic to be charted.
- Choose the type of control chart.
- Choose the centerline of the chart and the basis for calculating the control limits.
- Choose the rational subgroup or sample.
- Provide a system for collecting data.
- Calculate the control limits and provide adequate instruction to all concerned with the meaning and interpretation of the results.

In Phase II, new samples are compared with the estimated control limits to detect possible assignable causes and departures from in-control state. If these causes can be eliminated from the process, variability will be reduced and the process will be improved. In Phase II, process monitoring using on-line data can quickly detect shifts in the process from the baseline established in Phase I. Thus, the success of process monitoring in Phase II depends critically on the success of the Phase I analysis.

Montgomery (2009) described the process improvement activities using control charts, as illustrated in Fig. 2. Moreover, some references – Woodall (2000), Stoumbos et al. (2000), Woodall and Montgomery (1999), Banks (1993), etc.— have reviewed some topics and expressed some important materials and issues related to SPC; studying them can provide a proper insight to SPC for readers. By reviewing these sources and collecting the main topics that have been discussed, we summarize the general approach of statistical process control in Fig. 3.

5.2. Case indexing

The main steps of SPC approach are illustrated in Fig. 3. Various activities need to be done in each step to solve SPC problems or implement SPC in organizations. Many new techniques and several related issues have been developed around each step of SPC over time because of its rapid growth and widespread applications (Montgomery & Woodall, 1997). They may be used for effective similarity-based case retrieval. Main concepts and important fea-

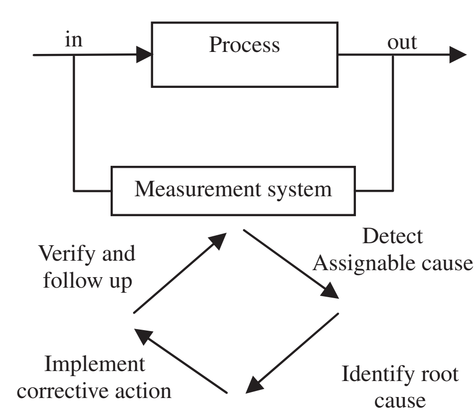

Fig. 2. Quality improvement using control charts (Montgomery).

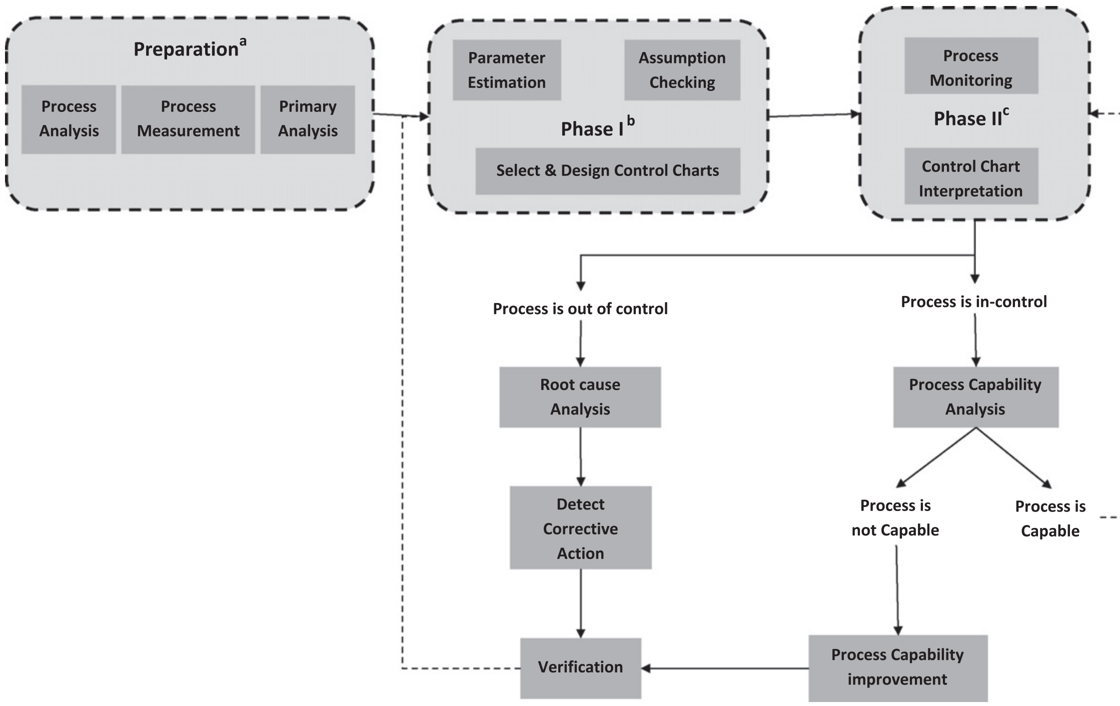

Fig. 3. The general approach of statistical process control. [a]To establish a SPC system, various preliminary activities are performed in "Preparation", such as selecting QCs, developing a measurement system, selecting a model for quality control etc. [b]In Phase I, control chart is designed using parameter estimation from in-control samples. [c]In Phase II, new samples are compared with the estimated control limits. If the process was in control, root causes must be identified and managerial actions must be done to correct them and if process was in-control, process capability will be analyzed and improved, if necessary.

tures of these issues must be extracted in order to determine the indexes used for representing the cases of SPC.

Therefore, to extract these contents, we first collected some important educational SPC books (DeVor, Chang, & Sutherland, 1992; Montgomery, 2009; Oakland, 2003; Ryan, 1989; Smith, 1998; Wheeler & Chambers, 1992, etc.) and used their materials, especially their tables of contents. Then, to avoid ignoring other important contents for indexing, we listed the keywords of top 1000 articles about SPC in terms of the citation number. We investigated these materials and determined their relations with the main steps of the SPC approach. Some of these are shown in Fig. 4. Finally, by using these materials as well as comments of some quality engineering experts, 15 indexes have been chosen to represent SPC cases. These indexes are classified in three different groups as shown in Fig. 5.

- *Functional index* describes the function of the case in the field of SPC, and 'Functional Roles' is the only sub- index of it.
  - *Functional roles*: Each case, based on the defined problem, goes through a path to address the problem. This path determines its functional roles. A case in SPC can take one or more functional roles among the following set which are specified among the extracted materials of SPC mentioned above:

    Monitoring, pattern recognition, interpretation, root cause analysis, data gathering, data reduction, data clustering, change point estimation, parameter estimation, process capability analysis, signal analysis, forecasting, assignable cause detection, chart selection, characteristic selection, model selection, accuracy calculation, fault diagnosis, chart designing, risk analysis, coast analysis, chart designing, specifying sampling method, etc.

    The studies which are totally theoretical and without application aspects were not compatible with the proposed system, considering the concept of a case in our study as mentioned in Section 5. Therefore, only cases which have a given functional roles can be entered into this system.
- *Structural indexes* represent features which describe the structural components of a case. These indexes include: *Process Features*, *Qualitative Characteristics (QCs)* and *Model Features*. *Process features* indexes analyze the structure of the process (*Quality level*, *Process Run-Time*, *Other Process features*). The features of the QCs of a process are represented by the *qualitative characteristics* indexes (*Number of QCs*, *Type of QCs*, *Sample size*, *Sample frequency*, *Distributions of QCs*). And *model features* describe some characteristics of the SPC models that have been applied in the case (*Autocorrelation*, *Concentration parameter*, *Scattering Parameter*, *Correlation*). These indexes, which are effective in the case retrieval and representation, have been illustrated respectively in Tables 3–5.
- *Applied indexes* describe the application area and the date of case implementation. Therefore, there are two sub-indexes, namely, *Application area* and *implementation date*.
  - *Application area*: SPC has been applied in various areas of science and technology. The application area of SPC cases can be useful in the similar case retrieval. A case may be implemented in one or more of the following areas[1]:

    Engineering; Business; Management and Accounting; Decision Sciences; Medicine; Computer Science; Material Science; Chemical Engineering; Mathematics; Biochemistry; Genetics and Molecular Biology; Economics; Econometrics and Finance; Nursing; Social Sciences; Health Professions; Environmental Science; Pharmacology, Toxicology and Pharmaceutics Energy; Physics and Astronomy; Earth and Planetary Sciences; Chemistry; Agricultural and Biological Sciences

[1] This classification for application areas has been taken from SCOPUS.

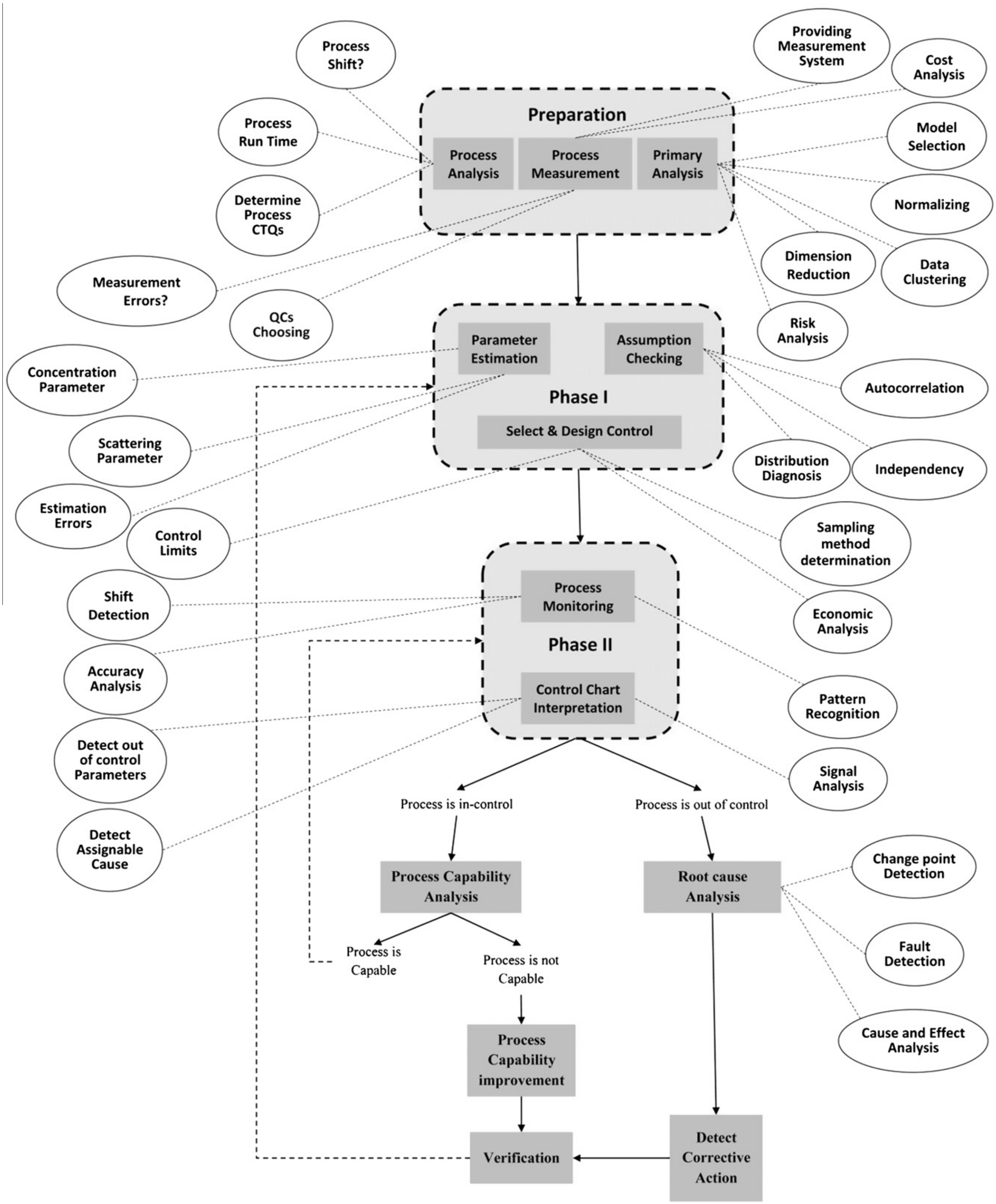

Fig. 4. Some issues and activities related to different parts of SPC approach.

– *Implementation date*: Since the knowledge area of SPC is growing, retrieving cases closer in terms of time can be more useful. This index is displayed as a number indicating the year of case implementation.

Each of these 15 indexes represents the features of SPC cases. It should be mentioned that these features are relevant to the problem space not to solution space. These indexes have been approved by several SPC experts. However, indexes can be updated and redefined if new findings are introduced in this field by adding cases to the case base. Due to the lack of information in most cases which is one of the restrictions of this research, although using more accurate measures can result in better efficiency, it seems that this level of information would be sufficient. Additionally, it

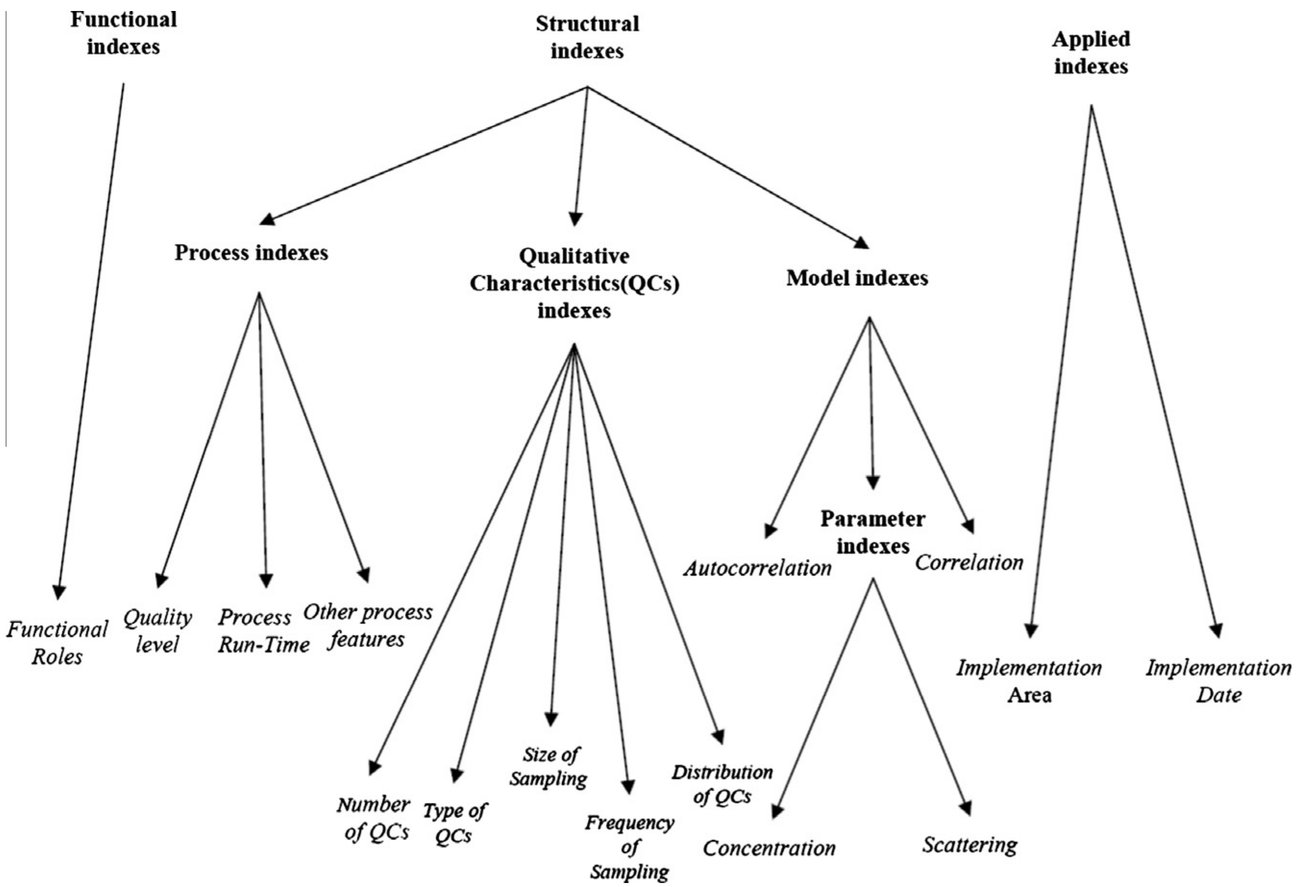

Fig. 5. The selected indexes for SPC case representation.

Table 3
The indexes related to process features.

Name	Explanation	Assigned value
Process run time	A process may be in a short-run or not. In a short-run process, because of short run-time, data to estimate the process parameters and control limits may not be available. Del Castillo, Grayson, Montgomery, and Runger (1996) and Del Castillo and Montgomery (1994), etc. have dealt with this issue	$= \begin{cases} 1 & \text{if process is ShortRun} \\ 0 & \text{otherwise} \end{cases}$
Level of process quality	Advanced technologies today keep the occurrences of defects at very low levels. In these high yield processes, because of the low defect counts, traditional control charting would encounter practical difficulties. Refer to Chan, Xie, and Goh (1997), Xie, Xie, and Goh (1995), and Goh (1987, 1991), etc. for further studies	$= \begin{cases} 1 & \text{if level of process quality is high} \\ 0 & \text{otherwise} \end{cases}$
Other process specifications	A process may be multistage or not or can be batch process or online (continuous). These issues effect on methods of SPC. These issues have been considered in some studies such as: Tsung, Li, and Jin (2008), Shu and Tsung (2003), Ündey and Çinar (2002), and Martin and Morris (1996)	$= \begin{bmatrix} \begin{cases} 1 & \text{if the process is Multi-stage} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if process is Batch-Process} \\ 0 & \text{if process is Online-Process} \end{cases} \end{bmatrix}^T$

is noteworthy that because of limited information in the cases, just some of these case indexes can often take values.

6. Case retrieval and similarity relation

In CBR systems, the quality of the results mainly depends on the similarity measures used to retrieve similar cases. During the retrieval procedure, the current problem is matched against the problems stored in the case base. Matching is the process of comparing two cases with each other and determining their degree of similarity (DOS). DOS is the value expressing how much a case is similar to another one. Regarding the low probability of a case that exactly matches the current problem and also because of different views on the same issues, the exact matching systems confront with difficulties. However, using the similarity concept in CBR improves the retrieval. Each of the retrieved cases is similar to the current problem in some way; the represented solution in these cases can be used as a guide to solve the current problem. To determine how to measure the similarity between the two cases of SPC, some basic notions have been defined as follows:

Definition 1. Binary relation S on a non-empty set X is called a similarity relation when:

(R) $\forall x$, xSx
(S) If xSy, then ySx
(T) If xSy, $y\,S\,z$, then xSz

The conditions (R), (S), and (T) are the reflexive, symmetric, and transitive laws. If xSy, we say that x and y are similar.

Table 4
The indexes related to features of qualitative characteristics.

Name	Explanation	Assigned value
Number of qualitative characteristics (QCs)	The number of QCs which are used for process measurement and process control in a case may be one or more (Bersimis, Psarakis, & Panaretos, 2007; Cheng & Thaga, 2006)	= (Number)
Kind of QCs	The QCs used for process measurement and control in a case may be of different kinds such as: variable, attribute, fuzzy or functional data. For more studies, Noorossana, Saghaei, and Amiri (2011), Zarandi, Turksen, and Kashanm (2006), and Kim, Mahmoud, and Woodall (2003), Woodall (1997), etc. could be referred	$\subseteq \{\text{variable}, \text{Attribute}, \text{Fuzzy}, \text{Functional Data}\}$
Size of sampling	Depending on the number of QCs used for process measurement and control, size of sampling for each of them can be Single, Multiple or Variable	$\subseteq \{\text{single}, \text{multiple}, \text{variable}\}$
Frequency of sampling	Also, sampling intervals of QCs in a case can be determined as high, medium or low also they may be fixed or variable. The effects of these issues have been considered by some literature such as: Sefik (1998), Prabhu, Runger, and Montgomery (1997), and Baxley and Robert (1995)	$= \begin{bmatrix} \text{sample frequency}: \\ \frac{\frac{\text{low}\to 1}{\text{medium}\to 2}}{\text{high}\to 3} \\ \frac{\text{fixed}\to 0}{\text{variable}\to 1} \end{bmatrix}^T$
Distribution of QCs	Traditional control charts are designed under the normality assumption but depending on the number of QCs in a case, each of them may have normal or non-normal distribution or even be non-parametric. The effects of the loss of this assumption have been discussed by Bakir (1998, 2006), Albers and Kallenberg (2004a), Shore (2004), and Vermaat, Ion, Does, and Klaassen (2003), etc.	$\subseteq \mathit{Normal}, \mathit{Non\text{-}Normal}, \mathit{Non\text{-}Parametric}$

Table 5
The indexes related to model characteristics.

Name	Explanation	Assigned value
Autocorrelation	A basic assumption usually made in constructing control charts is that the process data are independent. Autocorrelation has a significant impact on standard control charts designed under this assumption. It increases the probability of false alarm. The effects of autocorrelation have been topics of frequent discussion in SPC literature. For example, Noorossana, Amiri, and Soleimani (2008), Boyles (2000), Lu and Reynolds (1999), and Faltin, Mastrangelo, Runger, and Ryan (1997), etc.	$= \begin{cases} 1 & \text{if autocorrelation exists} \\ 0 & \text{otherwise} \end{cases}$
Concentration parameter	Parameter estimation is one of the basic activities in Phase I of constructing control charts. Mean and variance are the most important parameters for measuring concentration and spread of the process. Characteristics of these parameters and errors in measurement and estimation of them effect on method and model of SPC. Some of the most important issues related to these parameters are considered for case indexing here. For more information refer to the references such as: Jensen, Jones-Farmer, Champ, and Woodall (2006), Shu, Tsung, and Tsui (2005), Albers and Kallenberg (2004b), Macii, Carbone, and Petri (2003), and Hong and Elsayed (1999)	$= \begin{bmatrix} \begin{cases} 0 & \text{if parameter is fixed} \\ 1 & \text{if parameter is random variable} \end{cases} \\ \begin{cases} 1 & \text{if there is estimation error} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if there is measurement error} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if there is other type of errors} \\ 0 & \text{otherwise} \end{cases} \end{bmatrix}^T$
Spread parameter		$= \begin{bmatrix} \begin{cases} 1 & \text{if there is estimation error} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if there is measurement error} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if there is between error} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if there is within error} \\ 0 & \text{otherwise} \end{cases} \end{bmatrix}^T$
Correlation	In a case of SPC when we control the process by more than 1 QC, these QCs may be correlated to each other. When controlled QCs are of type of Functional Data like PROFILES, correlation relationship certainly exists. So this index will be activated when controlled QCs in a case are more than 1 or be kind of functional data. And the components of this index will be different accordingly. The effects of correlation have been topics of frequent studies in SPC literature, such as: Mastrangelo and Montgomery (1995), Harris and Ross (1991), Kai and Hancock (1990), and Neuhardt (1987)	$= \begin{bmatrix} \begin{cases} 1 & \text{if there is correlation between QCs} \\ 0 & \text{otherwise} \end{cases} \\ \begin{cases} 1 & \text{if expositor variable is a variable with fixed values} \\ 0 & \text{if expositor variable is a random variable} \end{cases} \\ \begin{cases} 1 & \text{if correlation relationship is linear} \\ 0 & \text{if correlation relationship is non-linear} \end{cases} \end{bmatrix}^T$

Definition 2. A case *A* is a triple and is defined as follows:

A = <*F*, *V*, *W*>

where *F*, a finite set of features of a case reflecting their nature (set of indexes); *V*, a set of values of indexes; and *W* is a set of weights corresponding to features of a case reflecting the importance of indexes.

The cases that have identical *F* and *W* are considered as one *type*.

Definition 3. For two cases *A* and *B* of one type, the overall similarity is defined as:

$$Sim(A,B) = \frac{\sum_{i=1}^{n} sim(a_i, b_i) \cdot w_i}{\sum_{i=1}^{n} w_i} \quad (1)$$

$VA = \{a1, a2, \ldots, an\}$ and $VB = \{b1, b2, \ldots, bn\}$ are sets of values, where equal indexes determine the corresponding indexes of the cases, sim_i is the degree of similarity and w_i is the weight of importance of i-feature.

Thus, to calculate the similarity between two cases:

First, corresponding features in the compared cases must be found. The values of corresponding indexes must be available in both compared cases while measuring the DOS. Otherwise, the indexes are ignored. Then, DOS between the corresponding indexes of the compared cases is computed. Finally, the obtained value is multiplied by the corresponding important factors which can be identified by the user for each index and then they add up to get overall similarity value.

6.1. Computing the degree of similarity for corresponding features

There are several approaches to calculate the degree of similarity for corresponding features of two compared cases. The notion of similarity is opposite of the definition of difference. Therefore, a way of measuring DOS is to determine the distance between two values.

The partial similarity l for two corresponding features a and b is defined as:

$$Sim(a,b) = l(a,b) = 1 - d(a,b) \tag{2}$$

where $d(a, b)$ is the distance function.

The value of distance d ranges from 0 to 1. If the features are completely different, then d equals 1 (Avramenko & Kraslawski, 2006).

6.2. Distance measurements

As mentioned previously, knowledge area of SPC is broad. Therefore, different types of indexes are used to represent the features of a case in this field such as numeric, vector, set of elements and qualitative indexes. Depending on the types of these indexes, the distance measures may be different.

Avramenko and Kraslawski (2006) collected some distance measures for different types of data applied in other investigations for nearest-neighbor retrieval. Some of these distance measures have been used in this research for computing the DOS between two SPC cases.

6.2.1. Numeric values

For the numbers, the distance is determined as:

$$d = \frac{|a-b|}{\text{range}} \tag{3}$$

where range is the range of values of variables, a and b and $|a - b|$ is the absolute value of difference between numbers.

6.2.2. Vectors

For n-dimensional vectors $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$, the distance is calculated as:

$$d = \frac{L_k(\vec{a},\vec{b})}{L_k(\sum_{i=1}^{n}\vec{e}_i)} \tag{4}$$

where L_k is the Minkowski metric:

$$L_k(a,b) = \left(\sum_{i=1}^{n}|a_i - b_i|^k\right)^{1/k} \tag{5}$$

With $k = 1$, it is called the city block (or Manhattan) distance.

If all the coordinates are equally important, it is necessary to make normalization, i.e., the real values of coordinates are converted to relative ones, belonging to the range [0, 1].

$e1 = (1, 0, \ldots, 0)$, $e2 = (0, 1, \ldots, 0)$, $\ldots$ $en = (0, 0, \ldots, 1)$ are basic vectors. Each basic vector determines the maximum possible distance along one of the coordinates.

6.2.3. Sets

The value of the distance function for the sets is determined by the number of elements in the sets, which are not common. The distance for the sets a and b is determined as

$$d = 1 - \frac{|a \cap b|}{\max(|a|,|b|)} \tag{6}$$

It is the Levenshtien formula.

6.2.4. Qualitative values

The qualitative values can be encoded using integer numbers assigned to each qualitative category, and the similarity measure between two qualitative variables can be computed by determining the distance between these integer numbers.

The type of indexes is determined in Table 6. Thus, it is recognizable which distance measure has been used to measure similarity between the corresponding indexes of two cases.

7. Testing system

We developed a software tool using C#.net 2008 and Access 2007. The software has two major modules, including *Case Registration* and *Case retrieval*. In *Case Registration* module, the implemented cases can be indexed and stored in the database. In *Case retrieval* module, specifications of the current problems are entered by the user (i.e., assigning values to the indexes and determining the importance of each index). Then, the system searches cases similar to the current problem in case base by measuring the degree of similarity with each of the stored cases.

To check the performance of this software, we have indexed 100 cases published in four major journals related to SPC (*Journal of*

Table 6
Type of indexes.

Index	Type
Functional roles	Set of elements
Process run time	Numeric (0 or 1)
Quality level of process	Numeric (0 or 1)
Other process specifications	Vector ($n = 2$)
Number of QCs	Numeric (natural)
Type of QCs	Set of elements
Size of sampling	Set of elements
Frequency of sampling	Qualitative vector ($n = 2$)
Distribution of QC	Set of elements
Autocorrelation	Numeric (0 or 1)
Concentration parameter	Vector ($n = 4$)
Scattering parameter	Vector ($n = 4$)
Correlation	Vector ($n = 3$)
Application area	Set of elements
Implementation date	Numeric (Year)

Table 7
Features of a sample problem.

Index	Value	Importance (specified by user in the range of [0–10])
Functional roles	Monitoring, change point detection	10
Process run time	More	5
Quality level of process	–	–
Other process specifications	–	–
Number of QCs	Functional data	10
Type of QCs	Normal	6
Size of sampling	–	–
Frequency of sampling	–	–
Distribution of QC	–	–
Autocorrelation	–	–
Concentration parameter	–	–
Scattering parameter	–	–
Correlation	–	–
Application area	–	–
Implementation date	–	–

Table 8
The profiles of most similar articles with the sample problem retrieved.

Retrieved article	Author(s)	Journal	Degree of similarity
A change point methods for linear profiles	Mahmud, Parker, and Woodall (2006)	Quality and reliability engineering international	0.85
Using control charts to monitor process and product quality profiles	Woodall et al. (2003)	Journal of quality technology	0.6
On the monitoring of linear profiles	Kim et al. (2003)	Journal of quality technology	0.5

Quality Technology; Technometrics; Quality Engineering; Quality and Reliability Engineering International) and stored them in the case base. Then, we used this case base to solve several hypothetical problems about SPC. In all the efforts, the solutions of similar cases retrieved from the database were helpful to solve new problems. For example, a problem with features indexed and illustrated in Table 7 entered into the software. Later, the degree of similarity between the problem and the articles stored in case base was calculated, and the three articles with the highest DOS were shown as retrieved cases. Profiles of these articles are illustrated in Table 8. Furthermore, the efficiency of the retrieval method which has been used in this CBR system was evaluated by some methods proposed by Gu and Aamodt (2006).

8. Conclusion

In this paper, a CBR system was proposed for problem solving in SPC due to some reasons such as expansion of the SPC knowledge area, lack of a comprehensive insight into SPC approach, and numerous cases in this field. Regarding these reasons, CBR systems have more advantages over other types of knowledge-based systems. Case representation and retrieval play an important role in the implementation of CBR systems. Therefore, after studying the various features of SPC problems, some indexes were chosen for case representation. Indexes can take different types of data. The proposed retrieval system based on the similarity concept is able to cope with the SPC cases represented by different types of data, and this is the strength of this system as compared with 'Exact Matching' search systems. Regarding the low probability of a case that exactly matches with the current problem and also because of the different views on the same issues, the exact matching search systems face difficulties, but by using the similarity concept, the proposed CBR system can retrieve some cases, each of them is similar to the current problem in some ways. The represented solution in these cases can be used as a guide for solving the current problem.

Defining different features for representing cases makes the proposed CBR system able to search SPC cases with more detailed properties. For developing this system some indexes were used to represent SPC cases. Defining and labeling of these indexes led to unification in literature related to SPC, while the keyword search systems encountered some conflicts in finding suitable cases. Furthermore, the degree of similarity measured by this system is a criterion for ranking system outputs (retrieved cases).

Lack of accurate information with missing or incomplete data in the cases was one of the difficulties in this research, but the CBR systems due to their nature are able to work under these conditions. Furthermore, since cases can simultaneously include tacit and explicit knowledge, both types of knowledge can be used to solve SPC problems by CBR systems.

Developing and presenting this system as an example can be regarded as a beginning for organizing SPC area. Therefore, the set of proposed indexes, which consider different insights and viewpoints on SPC area, can be revised and updated and can be completed in the future studies.

The proposed representation and retrieval system can be used as a virtual expert to help practitioners and researchers working on SPC subjects to solve new problems. The CBR system proposed in this article has been designed for the body of knowledge related to SPC, especially applied aspects of knowledge; furthermore, this software is capable of being customized specially for storing and using the knowledge and experiences related to SPC for different manufacturing and service organizations, and can be used for their practical problem-solving process.

CBR systems can begin with a little experience and can grow over time. Their power will increasingly grow by adding new cases to the case base. Therefore, providing a web-based version of this software can help to wide registration of SPC cases around the world and can be helpful for future problem-solvings in the domain of SPC.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications, 7*, 39–59.

Affisco, J. F., & Chandra, M. (1990). Quality assurance and expert systems – A framework and conceptual model. *Expert Systems with Applications, 1*(2), 147–153.

Albers, W., & Kallenberg, W. C. M. (2004a). Empirical non-parametric control charts: Estimation effects and corrections. *Journal of Applied Statistics, 31*(3), 345–360.

Albers, W., & Kallenberg, W. C. M. (2004b). Estimation in Shewhart control charts: Effects and corrections. *Metrika, 59*(3), 207–234.

ANSI/ASQ. B2-1996 (1996). *Control charts method of analyzing data*. Milwaukee, WI: American Society for Quality.

Avramenko, Y., & Kraslawski, A. (2006). A similarity concept for case-based design in process engineering. *Computers and Chemical Engineering, 30*, 548–557.

Awad, E. M. (1996). *Building expert systems: principles, procedures, and applications*. St. Paul: West Publishing Company.

Bakir S. T. (1998). Bibliography and review of distribution-free quality control charts. In *Joint statistical meetings of the American statistical association. August 9-13, Dallas, TX*.

Bakir, S. T. (2006). Distribution-free quality control charts based on signed-rank-like statistics. *Communications in Statistics – Theory and Methods, 35*(4), 743–757.

Banks, D. (1993). Is industrial statistics out of control? *Statistical Science, 8*(4), 356–377.

Baxley Jr, & Robert, V. (1995). Application of variable sampling interval control charts. *Journal of Quality Technology, 27*(4), 275–282.

Berger, R. W., & Hart, Th. H. (1986). *Statistical process control: a guide for implementation*. New York: Dekker.

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International, 23*(5), 517–543.

Boyles, R. A. (2000). Phase I analysis for autocorrelated processes. *Journal of Quality Technology, 32*(4), 395–409.

Brink, J. R., & Mahalingam, S. (1990). *An expert system for quality control in manufacturing*. *USF report* (pp. 455–466).

Caulcutt, R. (1995). The rights and wrongs of control charts. *Applied Statistics, 44*(3), 279–288.

Chakraborti, S., Graham, M. A., & Human, S. W. (2009). Phase I statistical process control charts: An overview and some results. *Quality Engineering, 21*, 52–62.

Chan, L. Y., Xie, M., & Goh, T. N. (1997). Two-stage control charts for high yield processes. *International Journal of Reliability Quality and Safety Engineering, 4*(2), 149–165.

Chen, F. T. (1991). A personal computer-based expert system framework for the design of experiments. *Computers and Industrial Engineering, 21*(1–4), 197.

Cheng, C. S., & Hubele, N. F. (1992). Design of a knowledge-based expert system for statistical process control. *Computers and Industrial Engineering, 22*(4), 501–517.

Cheng, S. W., & Thaga, K. (2006). Single variables control charts: An overview. *Quality and Reliability Engineering International, 22*(7), 811–820.