Analise da Viabilidade do Uso de Reamostragem Bootstrap em Bases de Dados Utilizando Funções Nativas do SGBD Oracle

RESUMO

O conhecimento é um dos mais importantes recursos de uma organização, tornando possíveis ações inteligentes nos planos organizacional e individual, induzindo a inovações e capacidade de continuamente. Construir o conhecimento, por inúmeras vezes, implica em processamento analítico em bases de dados dos mais diversos tamanhos. Em decorrência dos mais diversos fatores, o acesso aos dados completos dessas bases é por muitas vezes dificultado ou possui custo processual elevado. Tendo em mente essas observações, o presente trabalho propõe-se a validar a aplicabilidade da técnica de reamostragem Bootstrap em bases de dados transacionais, com o objetivo de verificar se analises estatísticas realizadas em um fragmento reamostrado de uma base de dados apresentam uma mesma linha de tendência da base original. A metodologia aplicada propõe que em uma base de dados qualquer, com distribuição populacional desconhecida, uma amostra seja extraída e reamostrada ***n*** vezes. A cada reamostra, análises estatísticas serão realizadas e ao final a media dessas estatísticas será comparada com o resultado da base de dados original. A escolha do Bootstrap como método de geração das reamostras deve-se a simplicidade de sua implementação e por desconsiderar a distribuição da populacional. Todos os processos propostos no trabalho utilizarão funções nativas do Sistema Gerenciador de Banco de Dados Oracle 10g. Até o momento, as funções de randomização mostraram-se bastante robustas na geração de reamostras com e sem reposição de elementos. Ademais, resultados iniciais em bases de dados experimentais têm mostrado que, uma vez selecionado o método estatístico adequado, os resultados obtidos por meio de reamostragem são, no mínimo, equivalentes aos obtidos com uso da amostra original.

PALAVARAS CHAVE. Bootstrap, Base de dados, Oracle, Reamostragem.

Analysis of the Feasibility of Using Bootstrap Resampling in Databases Using Native Functions in Oracle DBMS

ABSTRACT

Knowledge is one of the most important resources of an organization, with this is possible intelligent actions in organizational and individual plans, inducing to innovation and ability to continuously. Build knowledge, on some occasions, involves analytical processing in databases of different sizes. Some factors make difficult or promove the large cost to access to complete data for these databases. Thinking about these observations, this study aims to validate the applicability of the technique of bootstrap resampling in databases, to verify if the statistical analysis performed on a resampled fragment of a data base have a similar trend line of the original database. The methodology proposes that in any one database, with unknown population distribution, a sample is extracted and resampled n times. Each resampling, statistical analysis will be performed and the media at the end of these statistics will be compared with the results of the original database. The bootstrap resampling method was choice due to its simplicity of implementation and to disregard the distribution of population. All processes in the proposed work will use native functions System Manager Oracle Database 10g. At this moment, the functions of randomization are robust to generate resampling with and without replacement of elements. Furthermore, initial findings in experimental databases have shown that, once selected the appropriate statistical method, the results obtained by resampling are at least equivalent to those obtained using the original sample.

KEYWORDS. Bootstrap, Data Base, Oracle, Resampling.