고급소프트웨어실습

중간 리포트

(CSE4152)

Due: 2021년, 10/20일, 23:59분

학번: 20172141

반 번호: 2반

이름: 김미소

**1. 1주차에서 다룬 두 가지 예제 외에 몬테카를로 시뮬레이션을 사용하여 해결할 수 있는 문제를 2가지 이상 들고 어떻게 몬테카를로 시뮬레이션을 적용할 수 있는지 설명하시오. (1페이지 이내)**

1) 자산 가격 변동 예측

자산 가격 변동을 예측하기 위한 두 가지 요소가 있는데, 그 중 하나는 자산 가격에 대한 현재의 변동 방향을 나타내는 Drift가 있고 다른 요소로는 예측에 대한 입력 값인 시장 가격의 변동이다. 과거의 자산 가격 데이터들을 통하여 자산 가격의 변동 정보와 표준 편차, 분산, 평균 가격의 변동과 같은 정보들을 알 수 있다. 이 정보들을 이용하여 몬테카를로 시뮬레이션을 모델링 할 수 있다. 과거의 자산 가격 데이터로부터 얻은 정보들을 통하여 정기 일일 수익이나 일일 평균 수익에 대한 수식을 세울 수 있고 이렇게 얻어진 식을 통하여 다음 날에 대한 수익을 예상하기 위한 수식을 얻을 수 있다. 그리고 이 식에 대하여 랜덤한 시장 가격들을 반복적으로 넣고 계산을 원하는 만큼 충분히 하고 나면 미래의 가격 변동 시나리오에 대하여 시뮬레이션 할 수 있다.

2) 스케줄 예측

한 가지 프로젝트를 진행할 때, 이 프로젝트가 완수되기까지 얼마의 시간이 걸릴 지 예측할 수 있다. 프로젝트의 Task마다 걸릴 수 있는 최소 시간, 최장 시간, 예상 시간(확률이 제일 높은)을 가정하면 프로젝트 완수 기간에 대하여 예측할 수 있다. Task A가 10%의 확률로 2일이라는 최소 시간이 걸리고, 20%의 확률로 7일이라는 최장 시간이 걸리고, 70%의 확률로 4일이라는 예상 시간이 걸린다는 가정을 task마다 두는 것이다. Task A가 2일(10%), 7일(20%), 4일(70%)의 시간을 갖고 Task B가 1일(25%), 5일(15%), 3일(60%)가 걸린다고 할 때, 프로젝트 완수 기간은 task마다 어떤 값이 도출되는지에 따라 (2+1)일, (7+1)일, (4+5)일 등 여러가지의 값이 나올 수 있다. 즉 이러한 계산을 충분한 횟수로 반복하여 계산하면 프로젝트 완수 기간에 대한 각 값의 frequency들을 알 수 있고 이를 통해 프로젝트가 얼마가 걸릴 지 어느정도 예측할 수 있게 된다.

**2. Pearson Correlation Coefficient와 Principal Component Analysis (PCA)는 유사한 목적으로 사용될 수 있다. 어떤 경우에 두 가지 방법이 유사한 목적으로 사용될 수 있는지 구체적인 예를 들어서 설명하시오. (1 페이지 이내)**

데이터 간 상관관계를 파악하기 위한 목적으로 Pearson Correlation Coefficient와 Principal Component Analysis를 사용할 수 있다.

피어슨 상관 계수는 데이터의 변수 X, Y간의 상관 관계를 따질 때 X, Y가 함께 변하는 정도에 X, Y가 각각 변하는 정도를 나누어 구할 수 있다. 피어슨 상관 계수가 0이면 상관 관계가 없는 것이고 1에 가까울수록 완벽한 양의 상관 관계, -1에 가까울수록 완벽한 음의 상관 관계가 된다. 피어슨 상관 계수는 X와 Y의 공분산에 (X 표준편차 \* Y 표준편차)로 나누어 계산할 수 있다. 이 공분산을 통하여 데이터 변수 하나 하나에 대한 상관 관계를 파악할 수 있다.

주성분 분석은 공분산 행렬을 구한 후 Eigenvalue Decomposition을 통해 eigen value와 eigen vector를 통해 분산이 가장 큰 값, 즉 상관 관계가 가장 큰 축을 찾아낼 수 있다. 주성분 분석은 데이터 변수 하나 하나에 대한 상관 관계를 파악하기 보다는 여러 데이터들이 모여 분포를 이루어 낼 때 이 분포의 주성분이 무엇인지 분석해주는 것이 주 목적이다. 그러나 공분산 행렬을 이용한다는 점에서 공분산 행렬을 통해 변수 간 상관 관계를 파악하는 것이 가능하다. 데이터가 고차원인 경우, 주성분 분석을 통해 저차원으로 변환시켜 데이터 상관관계가 잘 드러나게 하고 변환된 데이터를 피어슨 상관 계수를 통해 변수 간 상관 관계를 분석하는 것도 방법일 것 같다.

**3. https://www.math.utah.edu/software/c-with-fortran.html 문서를 읽고 C/C++ 언어와 FORTRAN 언어의 차이를 요약한 후, C/C++ 프로그램에서 FORTRAN 함수를 호출할 때 주의할 점을 기술하시오. (1 페이지 이내)**

FORTRAN 언어는 프로그래밍 언어를 정의하는 것에 대한 좋은 이해가 형성되기 전에 만들어진 언어이기 때문에 데이터 타입이 제한적이었다. FORTRAN이 처음 만들어진 이후에 여러 데이터 타입을 추가하긴 했지만 처음부터 다양한 데이터 타입을 가지고 생성된 C/C++와 비교했을 때 공통적이라고 여길 수 있는 데이터 타입은 C/C++에서 int, float, double 타입 뿐이므로 유의해야 한다.

FORTRAN은 배열 저장 방식이 column-major order임에 반해 C/C++는 row-major order로 배열을 저장한다. 뿐만 아니라 FORTRAN은 인덱스가 1부터 시작하여 N개의 인덱스에 대하여 1~N까지 인덱싱할 수 있지만 C/C++는 0~N-1까지 인덱싱할 수 있다. 따라서 배열에서 특정 값을 가져올 때 C/C++의 A[i][j]를 FORTRAN에서는 A[j+1][i+1]과 같이 행, 열 순서를 바꾸고 인덱스 값도 조정해야 한다.

FORTRAN의 함수는 오로지 스칼라 값만 리턴할 수 있다. C/C++ 함수는 스칼라 값 뿐만 아니라 구조체나 클래스를 리턴할 수 있다.

FORTRAN의 파일은 binary, text 두 가지 형태의 파일이 있으며 language와 런타임 라이브러리가 인식할 수 있는 가장 작은 단위인 record 단위로 read, write 된다. C/C++에서는 파일을 unstructured byte stream으로 보고 fgetc()함수나 fputc()함수를 통해 바이트 단위로 파일에 접근할 수 있다. 그리고 런타임 라이브러리의 영향을 받지 않고 파일에 대하여 컨트롤 할 수 있다. FORTRAN과 C/C++ 사이에서는 text file만 공통적으로 사용할 수 있다.

C/C++는 동적으로 메모리를 할당하는 데에 자유롭지만 FORTRAN에서는 포인터 타입 자체가 잘 쓰이지 않으며 쓰인다 하더라도 C의 동적 메모리 사용 기능에 비하여 훨씬 못 미친다. FORTRAN에서는 problem size를 결정하고 working array를 스크래치 스페이스에 대한 오프셋을 계산하여 함수나 서브루틴에 전달한다. 그러므로 C/C++에서도 스크래치 배열을 FORTRAN 루틴을 호출하기 전에 동적 메모리에 할당하고 FORTRAN 루틴에서 리턴한 직후에 메모리를 free해주는 방식을 사용해야한다.

FORTRAN에서는 매개변수를 주소로 전달하고 C/C++에서는 매개변수를 값 혹은 주소로 전달할 수 있다. 따라서 FORTRAN 함수는 호출된 함수에서 값이 바뀌면 메인에서도 값이 바뀔 수 있지만 C/C++의 경우 매개변수를 값으로 전달한 경우 호출된 함수에서 값이 바뀌어도 메인에서는 아무런 영향을 받지 않는다. C/C++에서 FORTRAN 함수를 호출하기 위해서는 매개변수를 전부 주소 전달 형식으로 바꾸어야 한다.

FORTRAN 수치 라이브러리에 재귀에 관한 속성을 루틴에 선언하지 않았을 수 있으므로 C/C++에서 FORTRAN 루틴을 통한 재귀가 발생하지 않도록 유의하여야 한다.