**데이터 분석 관련 기술 결정**

논문 제안서에서 제시한 분석 기법 3가지를 하나하나 살펴보고 본 주제와 가장 적합한 기법을 결정한다.

1. **회귀분석(Regression Analysis)**
   1. **정의**

회귀분석 기법은 데이터 분석 관련 기술 중 하나로 어떤 변수가 다른 변수에 의하여 설명된다고 보고, 그 함수 관계를 조사하는 통계학적 기법이다. 다시 말해, 변수 간의 상호 관련성을 찾아내고, 변수들의 변화로부터 다른 변수들의 변화를 예측하는 기법을 말한다. 변수 간 상호 관련성 분석 기법에는 상관 분석과 회귀 분석이 있는데 전자는 기본 변수와 목적 변수의 역할이 바뀔 수 있을 때나 역할의 구분이 없을 때 사용할 수 있으며, 후자는 두 변수의 관계가 분명한 방향이 있을 때 사용할 수 있다. 데이터 분석 기법에서는 주로 후자를 사용하는 경우가 많다.

* 1. **검토**

본 논문의 주제는 날씨 차이, 즉 평균 기온과 구제역 발생 지역 기온의 차(差)와 구제역 전파 경로의 상호 관련성을 찾아내는 것이다. 여기서 기온 차는 독립변수, 전파 지역은 종속변수로써 활용할 수 있고 이들을 통해 논문의 결론을 내릴 수 있을 것으로 기대된다. 그러므로 회귀분석 비법을 사용하기로 결정하였다.

1. **연관규칙학습(Association Rule Learning)**
   1. **정의**

연관규칙학습이란 상품, 서비스 간의 관계에서 유용한 규칙을 찾아내는 기법으로 간단히는 If A then B로 표현할 수 있다. 특정 문제에 대한 예측 값이나 특정 목적에 따라 분류하는 것이 아니라, 상품이나 서비스의 거래 기록 같은 데이터에서 연관성 정도를 측정하여 관성이 많은 데이터를 그룹화하는 클러스터링의 일종이다.

* 1. **검토**

본 논문에서는 기온 차와 구제역 발생 지역의 상관 관계를 통해 결론을 도출하고자 한다. 단, 기온 차를 검토할 때 회귀분석에서는 단순히 구제역 발생 지역과 평년 기온의 차이 정도를 측정할 뿐이지만 연관규칙학습에서는 “구제역 발생 지역의 기온이 평년 기온보다 낮을 경우 구제역 발생 확률이 높은가” 등의 자세한 사례를 검토해 결론을 보강할 수 있으리라 생각한다. 그러므로 연관규칙학습 또한 사용하기로 결정한다.

1. **유전알고리즘(Genetic Algorithms)**
   1. **정의**

유전알고리즘이란 자연계의 유전학에 기반한 병렬적, 전역적인 탐색 알고리즘으로, 다윈의 적자생존 이론을 기본으로 한다. 문제의 가능한 해를 정해진 형태의 자료구조로 만들고 점차적으로 변형하여 더 좋은 해를 만든다. 이때, 해를 나타내는 자료구조는 유전자, 변형에 의해 더 좋은 해를 만드는 과정을 진화로 볼 수 있다.

* 1. **검토**

당초 본 논문 제안서 작성 당시 유전알고리즘을 이용할 수 있으리라 기대했었다. 하지만 다양한 데이터 중에서 필요한 데이터만을 추출해 그 데이터를 일반화시키는 유전알고리즘은 본 논문에서 사용할 데이터 분석 기법과 맞지 아니하다는 결론을 내렸다. 그러므로 본 기법은 사용하지 않는다.