

부정 탐지를 위한 이상치 분석 활용방안 연구 : 농수산 상장예외품목 거래를 대상으로*

김동성

한양대학교 일반대학원 경영학과
(paulus82@hanyang.ac.kr)

김기태

한양대학교 일반대학원 경영학과
(smash1005@hanyang.ac.kr)

김종우

한양대학교 경영대학 경영학부
(kjiw@hanyang.ac.kr)

박성기

IPCS Consulting
(steve@ipcs.kr)

기업 의사 결정 지원을 위하여 거래 데이터를 다양한 관점에서 분석하고 활용하려는 노력과 관심들이 증가하고 있다. 이러한 노력들은 고객 관리나 마케팅에만 국한되는 것이 아니라 부정행위에 대한 감시와 탐지를 목적으로도 다양한 분석 방안들이 연구되고 있다. 부정행위는 기술의 발전을 악용하여 다양한 형태로 진화하고 있으며, 이에 따라 목적에 맞는 부정탐지 방안 연구와 적용을 통하여 탐지 효율의 극대화를 위한 노력의 필요성이 증가하고 있다. 이러한 연구 동향의 일환으로 본 연구에서는 대용량 거래 데이터가 저장·관리되고 있는 국내 최대 농수산물 유통 시장의 2008년부터 2010년까지 상장예외품목의 거래 가격을 분석하여 부정 탐지 규칙을 도출하였으며, 전문가 검증을 통하여 도출된 규칙의 신뢰성을 확보하였다. 본 연구의 주요 부정거래 분석 방안으로는 정상적인 데이터들은 발생 확률이 높은 반면에 특이한 데이터들의 발생 확률은 낮다고 가정하는 통계적 접근을 통한 이상치 식별 방안을 활용하였다. 이에 따라 부정거래 분석 별로 정의된 Z-Score 값보다 클 경우 부정거래 탐지 대상이 된다. 다만 상장예외품목 거래의 경우 취급 가능한 중도매인의 수가 제한되어 있으며, 일반적인 상장품목의 거래보다 거래량이 적기 때문에 소수의 이상치가 품목의 평균에 미치는 영향이 크다. 그 예로 다른 소수의 중도매인들이 해당 품목을 정상적인 가격에 거래하였다 하더라도, 특정한 중도매인 한 명이 지나치게 비정상적인 가격에 거래할 경우 모든 거래들이 부정거래로 탐지될 가능성도 있다. 이러한 문제를 해결하기 위하여 기존의 Z-Score의 개념을 활용하여 수정된 Z-Score(Self-Eliminated Z-Score)를 사용하였다. 또한 부정 유형별 탐지 규칙 관리와 활용을 위한 시스템 프로토타입(prototype) 개발을 수행하였다. 이를 통하여 실제 부정거래 탐지 업무에 적용할 수 있는 효과적인 방안을 제시하였고, 농수산 유통시장의 공정성 및 투명성 확보를 위한 관리·감독의 기능 강화가 가능할 것이다.

주제어 : 부정 탐지, 이상치 검출, 상장예외품목

논문접수일 : 2014년 6월 17일 논문수정일 : 2014년 9월 11일 게재확정일 : 2014년 9월 12일
투고유형 : 국문급행 교신저자 : 김종우

1. 서론

정보 기술의 활용이 점차 고도화됨에 따라 다수의 업무들이 전산화되어 다양한 데이터들이 저장 및 관리되고 있으며, 데이터 분석을 통하여 유용한 정보를 찾아내고자 하는 노력이 여러 분

야에서 증가하고 있다(Lee, 2011). 데이터의 수집과 분석 및 활용에 대한 관심은 클라우드 컴퓨팅과 사물 인터넷이라 불리는 IoT (Internet of Thing)의 발전으로 인하여 더욱 가속화 될 전망이다. 반면에 이러한 기술의 발전을 악용하여 정보통신, 온라인 banking, 전자 상거래 등 많은 분야

* 이 논문은 2012년 한양대학교 교내연구비 지원으로 연구되었음(HY-2012-G)

에서 다양한 형태의 사기·부정행위가 증가하고 있으며, 전 세계적으로 매년 수십억 달러의 피해가 발생함에 따라 부정 탐지(fraud detection)는 지속적으로 연구되는 주요 관심 분야 중 하나이다(Kou et al., 2004, Ramamoorti, 2008, Whiting et al., 2012). 이러한 부정행위 탐지 노력은 온라인에서의 사기·부정행위를 비롯하여 은행, 신용카드, 보험 등의 민간 금융 분야와 감사업무, 사회복지, 조세 등 정부의 재정 지원·징수까지 다양한 분야에서 과학적 기법을 활용한 방안이 연구되고 있다(Abbasi et al., 2012, Bierstaker et al., 2006, Lee et al., 2010). 이러한 연구 동향의 일환으로 본 연구에서는 전자입찰 시스템과 통합정산 시스템을 통하여 하루 평균 거래 물량 약 7,300톤, 거래금액 약 104억 원에 이르는 거래 내역이 데이터베이스에 저장 및 관리되고 있는 국내 최대 농수산물 유통시장의¹⁾ 거래 기록을 활용하여 효과적인 부정 탐지 방안을 모색하였다.

농수산물 유통 거래 데이터는 시간에 따른 가격 차이와 거래 물량의 변화가 잦은 농수산물 품목의 특성으로 인하여 대용량 데이터의 특징을 가지며, 관리와 분석에 실질적으로 많은 어려움이 존재한다. 더불어 본 연구의 분석 대상인 상장예외 품목 거래는 물량 탈루 및 수수료 과다 징수, 거래 실적 편법 기재 등 거래의 공정성과 투명성 측면에서 다양한 문제가 지속적으로 논의되고 있으며, 부정행위 적발과 차단을 위한 대책 마련의 필요성이 꾸준히 제기되어 왔다(Gwon and Kim, 2012). 이에 따라 본 연구에서는 사전에 정의된 부정거래 의심 유형을 바탕으로 2008년부터 2010년까지 농수산물 상장예외품목의 일별, 월별, 분기별 거래 가격을 분석하여 부정거래 탐지

규칙을 도출하였으며, 전문가 검증을 통하여 도출된 규칙의 신뢰성을 확보하였다. 또한 부정거래 유형별 규칙 관리 및 활용을 위한 부정거래 탐지 시스템의 프로토타입(prototype) 개발을 통하여 과학적이고 자동화된 부정 탐지 방안의 개발을 꾀하였다. 이를 바탕으로 대용량 데이터에서 목적에 맞는 데이터 추출과 거래 기록의 비교 및 추이 파악 등 다차원 분석이 가능하며, 시간과 자원의 제한을 받지 않는 전수 조사와 상시 검사가 가능하다. 나아가 부정거래의 사전 예방으로 도매시장 통제 기능의 강화와 거래의 공정성·투명성 확보가 가능 할 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 농수산물 상장예외품목 거래제도, 정보기술을 활용한 부정 탐지(fraud detection) 현황, 이상치 검출(outlier detection) 기법에 대하여 검토한다. 3장에서는 이상치 검출 기법을 활용한 상장예외 품목의 부정거래 적출 방안을 제안하며, 4장에서는 실증 분석을 바탕으로 부정거래 탐지 규칙을 도출한다. 5장에서는 도출된 부정거래 탐지 규칙의 활용을 위한 시스템 프로토타입(prototype)에 대하여 확인하며, 마지막 6장에서는 결론과 본 연구의 한계점 및 향후 연구 방향을 제시한다.

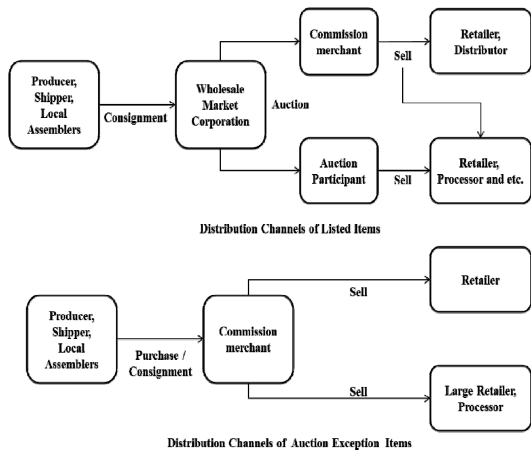
2. 연구배경

2.1 농수산물 상장예외품목 거래제도

도매시장 매매방법은 경매 거래를 기본 원칙으로 하나 중도매인이 직접 생산지에서 수집 및 거래가 가능한 상장예외품목 거래 제도가 존재한다. 상장예외 거래 제도는 경매 거래의 보완장

1) <http://www.garak.co.kr/gongsa/jsp/gs/main.csp>

치로써 특정 품목에 대하여 한시적이고 일시적으로 도입되는 거래 형태이다. 상장예외품목 지정은 해당 품목의 거래량이 적거나 취급하는 중도매인이 소수인 경우, 중도매인의 해당 농수산물 매입에 어려움 존재 등을 도매시장 개설자가 인정해야 하며, 거래 품목과 중도매인에 대하여 1년 단위로 허가·관리되고 있다. 하지만 상장예외품목 거래 제도는 취급하는 중도매인의 수 및 거래물량 증가로 인한 효율적인 관리의 어려움과 수의매매에 의한 가격 결정으로 유통정보의 생산과 이용의 어려움 등 공정성과 투명성 측면에서 한계가 존재한다(Wi and Kwon, 2009). 이에 따라 상장예외품목이 아닌 품목들의 거래 발생, 생산 농가들의 시장대응력 약화, 거래 실적 누락과 세금 탈루 등의 문제점들이 제기되고 있는 현황이다(Gwon and Kim, 2012). 상장품목 거래 유통 경로와 상장예외품목 거래 유통 경로의 가장 큰 차이점은 생산자, 출하자 및 산지유통인과 중도매인의 거래 시 도매시장의 농수산 품목 상장을 통한 경매거래 이행 여부이다(<Figure 1> 참조).



(Figure 1) Distribution Channels of Listed Items and Exception Auction Items

상장예외품목 거래의 대표적인 부정거래 의심 유형으로는 시장 평균 거래 가격에 비해 지나치게 높거나 낮은 판매가 차이, 중도매인과 생산자 간 실제로 이루어지지 않은 거래의 송품장 허위 등록, 계절적 특성으로 인해 중도매인이 취급하기 어려운 품목의 거래 발생 또는 특정 품목 거래 물량의 급격한 증감 발생 등이 있다.

2.2 정보기술을 활용한 부정 탐지 (Fraud Detection)

The Association of Certified Fraud Examiners (ACFE)는 부정행위에 대하여 '소속 된 조직에서의 지위를 이용하여 자신의 이익을 위한 자원이거나 자산의 고의적인 악용'이라고 정의한다 (Singleton and Singleton, 2010). 부정의 유형과 그에 따라 발생하는 피해와 사고는 각각 유형별로 다양하고 광범위하므로 여러 가지 형태로 부정에 대한 정의와 분류가 가능하다. 부정의 종류로는 대표적으로 자산 횡령(misappropriation of assets), 허위 보고(fraudulent statements), 부패(corruption)의 3가지로 분류된다(Lee and Chang, 2012).

정보의 디지털화로 인하여 다양한 자료들이 컴퓨터 내부에 저장되고 전산처리 과정에 의존하는 업무 변화로 인하여 디지털 데이터에 대한 부정 탐지 필요성이 증가되고 있다. 이에 따라, 부정예방(fraud prevention) 및 부정 탐지(fraud detection)를 목적으로 국내외 민간분야 및 공공분야에서는 데이터마이닝과 통계분석, 인공지능 기법 등을 활용하며, 정형화된 데이터뿐만 아니라 다양한 데이터를 바탕으로 성과검사, 합법성 검사 등의 범위로 확대되고 있는 추세이다.

정보기술을 활용한 부정 탐지의 국내외 주요

동향을 살펴보면 민간 분야에서 카드 회원의 이용 패턴 분석을 통한 부정사용 탐지와 온라인 은행 업무 시 부정 탐지 및 위험 관리(Raj and Portia, 2011, Li et al., 2012), 데이터마이닝기법을 활용한 주식시장에서의 이상매매 탐지(Hong et al., 2006), 손해배상금과 같은 보험금 부정 수급에 대한 탐지 연구(Ham and Hong, 2008)가 있으며, 공공 분야에서는 세금 탈세 탐지 및 공공 보조 지원금의 부정 수급(Linoff and Berry, 2011) 등이 존재한다. 이 밖에 부정탐지 관련 선행연구들은 다음과 같다(<Table 1> 참조).

(Table 1) Review of Related literature on Fraud Detection

Authors (year)	Research overview
Lee, J. W, D. H. Lee, and I. S. Kim (2013)	To detect the smishing, they analyzed the characteristics of the contents of the message and the sender's address using SVM (Support Vector Machine).
Lee, J. W., I. A. Choi, W. C. Ji, H. Y. Park, and H. J. Shin (2008)	They proposed a new evaluation index that can be quantified the degree of outliers for fraud detection of medical health care benefits.
Chau, D. H and C. Faloutsos (2005)	They suggested a method to detect auction fraudsters on eBay, through analyzing the fraudulent transaction history by decision tree induction technique.
Durtschi, C., W. Hillison, and C. Pacini (2004)	They tries to detect the fraud in accounting data, using Benford's law.
Fawcett, T. and F. Provost (1997)	They proposed a rule learning algorithm using call records to detect cloning fraud of cellular phone.
Ghosh, S., and D. L. Reilly (1994)	In order to detect fraud of credit card, they proposed neural network approach using the analysis results of past fraudulent behavioral data.

부정 탐지 방안에 대한 국내외 기존 연구들은 탈세나 부정 수급과 같이 사전에 명확하게 정의된 부정행위를 탐지하는 반면에, 본 연구는 수의 매매로 인하여 시장에서의 기준 매매 가격과 거래 물량에 대한 투명성 측면에서 문제가 제기되고, 부정행위에 대한 정의가 명확하지 않은 농수산 상장예외품목 거래를 대상으로 부정 탐지 방안을 연구하며, 이러한 점이 기존 연구와의 차이점이라고 할 수 있다.

3. 분석방안

3.1 분석 데이터 및 선정

본 연구 수행을 위하여 2008년부터 2010년까지 농수산 상장예외품목에 대한 중도매인 거래 내역 데이터를 수집하였으며, 날씨의 영향이나 신선도 유지기간 등 농산품목과 수산품목이 갖는 특성 차이로 인하여 거래내용에 차이가 발생함에 따라 농수산 상장예외품목 중 농산 품목만을 연구 대상으로 선정하였다. 또한 분석의 효과성과 효율성을 위하여 최소 거래 기준을 설정하여 거래 금액이 연간 1억 2천만 원 이상이 되는 품목들에 한하여 분석 데이터로 선정하였다.

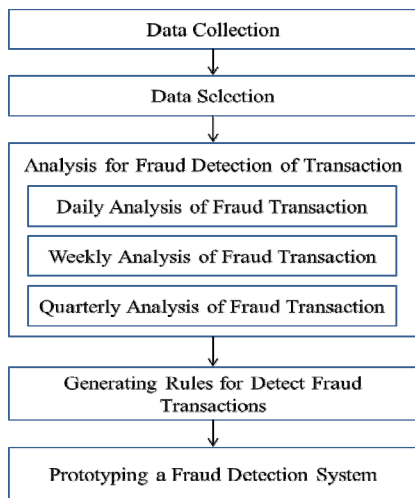
이에 따라 실제 분석에 사용된 농산 품목의 연도별 거래 데이터의 특징은 2008년 거래 수 307,423건, 총 거래 품목 90개, 총 거래 금액 3,223억 원, 2009년 거래 수 346,558건, 총 거래 품목 103개, 총 거래 금액 3,608억 원, 2010년 거래 수 392,582건, 총 거래 품목 115개, 총 거래 금액 4,911억 원으로 매년 거래 수와 거래 금액이 증가하는 추세를 확인할 수 있다(<Table 2> 참조).

〈Table 2〉 Analytical Data (2008~2010)

Year	2008	2009	2010
The Number of Transactions	307,423	346,558	392,582
The Number of Commission merchants	333	337	328
The Number of Items	90	103	115
Total Amount of Transactions	3,233 Billion won	3,608 Billion won	4,911 Billion won

3.2 분석 절차

본 연구는 다음과 같은 분석 단계를 거쳐 수행되었다(〈Figure 2〉 참조).



〈Figure 2〉 Research Procedure

부정거래 분석은 시간에 따른 가격 변동과 거래 추이 등을 고려하기 위하여 일간, 주간, 분기별로 분석한 결과를 바탕으로 각각의 부정거래 탐지 규칙을 정의하였다. 부정거래 분석의 일간,

주간, 분기별 구분의 주요 요인은 실제 가락시장 상장예외품목 거래의 부정탐지 업무환경을 우선적으로 고려하였다. 또한, 상장예외품목 거래의 경우 각각의 품목을 취급하는 중도매인의 수가 제한적인 이유로 기준 가격이 없고, 매수자인 중도매인이 가격결정권을 갖는 거래 환경을 고려하여 기간별 세분화 된 매매가격 분석이 필요함도 주요 요인 중 하나이다.

3.3 부정거래 분석

본 연구의 각각의 기간별 부정거래 분석 방안은 ‘품목별 가격차이’ 분석, ‘송품장별 판매가격 차이’ 분석, ‘판매가격 변화 추이’ 분석, ‘비계절 품목 거래’ 분석으로 다음과 같다(〈Table 3〉참조).

〈Table 3〉 Analysis Methods for Fraud Detection of Transaction by Period

Period	Analysis Methods	Details
Daily	Price Difference of Product Category	Comparative analysis of the average market price of the product category by day
	Difference between Selling Price per invoice	From the recoding of invoice of Commission merchant per day, the analysis of the sale price of the same product
Weekly	Price Difference of Product Category	Comparative analysis of the average market price of product category by week
	Trend of Changes in Selling Price	Comparative analysis of the average selling price of weekly three-year deals by the Commission merchant
Quarterly	Non-seasonal Item Trading	Trading frequency analysis of the Commission merchant by product category
	Changes in Sales Price	Comparative analysis of the average selling price of quarterly three-year deal by the Commission merchant

품목별 가격 차이와 송품장별 판매가격 차이는 분석 기간이 길어짐에 따른 평균 가격 차이 결과가 분석 목적에 맞지 않으므로 분기별 분석을 제외한 일간 및 주간별 분석만 수행하였다. 또한 과거 3년간의 거래 기록을 활용한 품목별 판매가격 변화 추이 분석 시에는 분석 대상이 되는 날짜가 과거와 동일하더라도 요일 차이가 발생하여 유통 시장의 개장일과 휴일이 비교가 될 수 있으며, 특히 음력 날짜를 사용하여 매년 날짜가 변화하는 명절은 농산물 특성상 가격변화가 민감하기 때문에 이러한 사항들을 고려하여 주간과 분기별 분석만을 대상으로 한정하였다. 이 밖에도 비계절 품목의 거래 분석을 위하여 계절적 요인을 최대한 반영 할 수 있도록 분기별 거래 빈도에 대한 분석을 수행하였다.

본 연구의 주요 부정거래 분석 방안으로는 정상적인 데이터들은 발생 확률이 높은 반면에 특이한 데이터들의 발생 확률은 낮다고 가정하는 통계적 접근을 통한 이상치 식별 방안을 활용하였다. 이에 따라 우선 품목별 거래 단가에 대하여 정규분포를 만들고 품목별 거래 각각의 경우가 정규분포 상 어느 위치에 분포하는지 확인하였으며, 일간-주간 부정거래 분석 별로 정의된 Z-Score 값보다 클 경우 부정거래 탐지 대상이 된다.

다만 상장예외품목 거래의 경우 중도매인의 수가 제한되어 있기 때문에 일반적인 상장품목의 거래에 비하여 거래량이 상대적으로 적은 편이다. 그렇기 때문에 소수의 이상치가 품목의 평균에 미치는 영향이 크다. 그 예로 다른 소수의 중도매인들이 해당 품목을 정상적인 가격에 거래하였더라도, 특정한 중도매인 한 명이 지나치게 비정상적인 가격에 거래할 경우 모든 거래들이 부정거래로 탐지 될 가능성도 있다. 이러한 문제를 해결하기 위하여 본 연구에서는 두 가지

방법을 사용하였다. 첫 번째로는 하나의 송장 내에 들어있는 같은 품목의 거래는 하나의 거래로 취급하였다. 이를 통해 농수산물 하루 내에 가격이 심하게 바뀔 수 있다는 문제를 처리하여 정상적인 거래가 부정거래로 탐지되는 것을 줄일 수 있었다. 두 번째로는 기존의 Z-Score의 개념을 활용하여 수정된 Z^{se} (Self-Eliminated Z-Score)를 사용하였다. Z^{se} 는 해당 품목의 전체 거래에서 비교하고자 하는 특정 거래를 제외한 거래 집합의 평균 가격에서 산출한 Z-Score이다. 이를 사용하여 부정거래의 평균 가격이 Z-Score의 기준이 되는 전체 평균 가격에 미치는 영향을 줄이고자 하였다.

Z^{se} 를 구한 방법은 다음과 같다. 우선 탐지 대상 품목의 전체 거래 집합을 a 로 정하고, Z^{se} 를 구하려는 특정한 한 건의 거래를 m , 전체 거래 집합 a 에서 특정한 거래 m 을 제외한 새로운 거래 집합을 m' 이라 정한다. 그리고 함수 $n(\cdot)$ 은 해당 집합 내의 거래의 수를 결과 값으로 가진다. (단, $n(a) \geq 2$, $n(m) = 1$)

탐지 대상 품목의 전체 거래 집합 a 에서 특정한 한 건의 거래 m 을 제외한 집합의 평균 거래 가격 표준편차 s_r 을 구하기 위하여, 전체 거래 집합 a 의 평균 거래 가격의 표준편차인 s_a 를 이용하여 편차제곱합인 S'_a 을 구한다(<Formula 1> 참조).

$$S'_a = (s_a)^2 \times (n(a) - 1) \quad (1)$$

다음으로 탐지 대상 품목의 전체 거래 내 평균 거래 가격인 X_a 와 제외하고자 하는 특정한 한 건의 거래의 평균 거래 가격인 X_m 을 이용하여 특정한 거래의 편차제곱인 S'_m 을 구하였다(<Formula 2> 참조).

$$S'_m = (X_a - X_m)^2 \quad (2)$$

위에서 구한 S'_a 에 S'_m 을 이용하여 m 을 제외한 거래 집합 $\neg m$ 의 편차제곱합 $S'^{se}_{\neg m}$ 을 계산한 후, 이를 바탕으로 한 표준편차 $s^{se}_{\neg m}$ 에 대한 계산식은 다음과 같다(<Formula 3> 참조).

$$s^{se}_{\neg m} = \sqrt{\frac{S'_a - S'_m}{n(a) - 2}} \quad (3)$$

다음으로 m 을 제외한 전체 거래 집합의 평균 거래 가격 $X^{se}_{\neg m}$ 를 구하기 위하여, 전체 거래 집합의 평균 거래 가격 X_a 와 특정한 거래의 평균 거래 가격인 X_m 을 이용하였다(<Formula 4> 참조).

$$X^{se}_{\neg m} = \frac{X_a \times n(a) - X_m}{n(a) - 1} \quad (4)$$

마지막으로 전체 거래 집합 a 의 평균 거래 가격과 표준편차 대신 a 에서 m 이 제외된 거래 집합 m 의 평균 거래 가격과 표준편차를 이용하여, 특정 중도매인의 품목 거래 m 을 제외한 수정된 Z-Score인 $Z^{se}_{\neg m}$ 를 산출할 수 있다(<Formula 5> 참조).

$$Z^{se}_{\neg m} = \frac{X_m - X^{se}_{\neg m}}{s^{se}_{\neg m}} \quad (5)$$

3.3.1 일간 부정거래 탐지

일간 부정거래 분석 시에는 다음과 같이 품목별 판매가격 차이 분석, 송품장별 판매가격 분석을 수행한다.

(1) 품목별 탐지 분석

시장 가격 차이 분석을 위하여 부정 탐지 대상이 되는 중도매인이 거래한 품목의 가격을 제외한 해당 품목의 전체 시장 거래 가격 평균과 중도매인의 거래 가격을 비교한다. 이를 위하여 품목별 가격의 Z^{se} 를 도출 후 전체 시장 평균 가격 범위에서 일정 수준을 초과하는 가격으로 거래한 중도매인을 탐지한다.

(2) 송품장별 판매가격 분석

중도매인은 모든 거래에 대하여 송품장을 작성하여 신고 및 제출해야 하며, 하나의 송품장에는 다수의 거래 기록 작성이 가능하다. 이러한 송품장 기록의 확인 결과 특정 송품장 내에서 동일 품목이라도 거래 가격의 차이가 발생하며, 이러한 현상은 생산자와 품목이 동일한 경우에도 존재한다. 송품장 내에서의 동일 품목의 가격 차이가 일정 범위 이상을 벗어난다면 이는 특정 부정거래를 정상 거래처럼 보이기 위한 가격 조정으로도 의심 할 수 있다. 송품장 내의 동일 품목의 가격 차이를 동일한 일시에 동일 품목을 취급한 다른 중도매인들의 송품장 내에서 발생하는 가격 차이 평균과 비교하여 상대적으로 높게 이루어지는 중도매인을 분석한다.

3.3.2 주간 부정거래 분석

주간 부정거래 분석 시에는 다음과 같이 품목별 판매가격 차이 분석, 품목별 3년간 거래 가격 차이 분석을 수행한다.

(1) 품목별 판매가격 차이 분석

시장에서 거래되는 품목별 가격에 대하여 시간의 흐름에 따른 변화를 고려하기 위해 한 주

동안 중도매인 자신이 거래한 품목의 가격을 제외하고 해당 품목의 한 주 동안 전체 시장 거래 가격 평균과 비교한다. 거래 가격의 비교를 위하여 품목별 가격의 표준 점수 Z^{se} 를 도출하며, 한 주 동안 시장에서 이루어진 평균 가격 차이 범위에서 일정수준 이상의 범위에 해당하는 가격으로 거래한 중도매인을 탐지한다. 일간 탐지와 동일하게 분석 대상이 되는 중도매인의 한 주 동안의 거래 가격은 전체 시장 거래 가격 평균에도 영향을 미치게 되므로 중도매인 자신의 거래 기록은 시장 거래 가격 평균 도출 시 반영하지 않는다.

(2) 품목별 3년간 거래 가격 차이 분석

중도매인의 과거 거래 기록과는 다른 형태의 거래가 발생하였다면 부정거래로 의심할 수 있다. 이에 따라 과거 3년간 주간 거래 기록 평균과 비교하여 가격 증가가 큰 품목들을 분석한다.

3.3.3 분기별 부정거래 분석

분기별 부정거래 분석 시에는 다음과 같이 중도매인의 특정 품목 거래 횟수 분석, 품목별 3년간 거래 가격 차이 분석을 수행한다.

(1) 중도매인의 특정 품목 거래 횟수 분석

품목의 계절적 특성을 고려하였을 때, 중도매인이 실제 취급하기 어려운 품목들의 거래 기록은 중도매인의 실적 문제, 기타 사유로 인한 허위 거래로 판단할 수 있으며, 계절적 요인을 고려하기 위하여 분기별로 중도매인의 특정 품목 거래 횟수가 일정 수준 이하인 중도매인만을 분석한다.

(2) 품목별 3년간 거래 가격 차이 분석

중도매인의 3년간 분기별 거래 가격 평균 차이 발생이 전체 시장에서의 품목 거래 가격 변화로 발생한 것일 수 있으며, 이를 고려하여 3년간의 분기별 거래 가격 분석을 수행한다. 과거 3년간 분기별 거래 기록 평균가와 비교하여 일정 수준 이상 거래 가격의 증가가 발생한 품목들을 분석한다.

4. 분석 결과

4.1 부정거래 탐지 규칙 도출

부정거래 분석과 전문가 검증을 통하여 최종 도출된 부정거래 탐지 규칙은 다음과 같다 (<Table 4> 참조).

<Table 4> Generated Rules for Fraud Detection of Transaction by Period

Period	Analysis Methods	Details
Daily	Price Difference of Product Category	$Z^{se} \geq 2.57$
	Difference between Selling Price per invoice	$Z^{se} \geq 2.33$
Weekly	Price Difference of Product Category	$Z^{se} \geq 2.57$
	Trend of Changes in Selling Price	The current average selling price \geq Three-year average selling price * 2
Quarterly	Non-seasonal Item Trading	The number of transactions of the item per quarter = 1
	Changes in Sales Price	The current average selling price \geq Three-year average selling price * 2

본 연구에서는 사전에 부정거래에 대한 학습 데이터가 존재하지 않은 관계로 도출된 탐지 규칙과 결과의 정확도 확인에 어려움이 존재하였으며, 이에 따라 실제 상장예외품목 부정거래 탐지 관련 업무의 전문가 검증을 통하여 도출된 탐지 규칙의 유의성을 검증하였다.

도출된 부정 거래 탐지 규칙 중 이상치 검출 방안을 활용한 품목별 가격 차이와 송품장별 판매가격 차이 분석에서 각각 정규 분포에 대하여 신뢰수준 99%의 양측 분석과 단측 분석으로 인하여 탐지 규칙 값의 차이점이 존재한다. 품목별 가격 차이 분석은 판매에 따른 거래 수수료 회피를 목적으로 상대적으로 낮은 거래 금액이 발생할 수 있으며, 중도매인의 상장예외품목 거래 허가를 위한 최소 거래 실적 유지를 위한 높은 거래 금액 발생도 부정 의심 유형으로 확인되어 양측 분석을 실시하였다. 송품장별 판매가격 차이 분석은 동일 품목의 낮은 거래 금액과 높은 거래 금액을 송품장 내에 작성함으로써 송품장 단위에서의 특정 품목 판매 가격 평균에 영향을 미치는 부정 의심 유형이 확인되어 단측 분석을 실시하였다.

4.2 일간 부정거래 탐지 결과

2008년부터 2010년까지 일간 부정거래 탐지 결과는 다음과 같다. 2008년 품목별 시장 평균 거래 가격과 송품장별 판매 가격 차이 탐지 결과 각각 11,897건(평균 39.7건/일), 10,638건(평균 35.5건/일)으로 나타났으며, 2009년 품목별 시장 평균 거래가격과 송품장별 판매 가격차이 탐지 결과는 15,437건(평균 51.5건/일), 11,470건(평균 38.2건/일)으로 확인되었다. 2010년 품목별 시장 평균 거래가격과 송품장별 판매 가격차이 탐지

결과 각각 11,897건(평균 39.7건/일), 10,638건(평균 35.5건/일)으로 나타났다.

4.3 주간 부정거래 탐지 결과

품목별 주간 시장 평균 거래 가격 차이 분석에 따라 2008년 1,579건(평균 29.3건/주), 2009년 2,221건(평균 41.1건/주), 2010년 1,579건(평균 29.3건/주)으로 나타났다. 주간 부정거래 분석 중 과거 3년간의 판매 가격 변화 추이 분석은 수집된 데이터의 제한으로 인하여 2010년 거래 내역만이 분석 가능하였으며, 그에 따른 탐지 결과는 789건(평균 14.6건/주)로 확인된다.

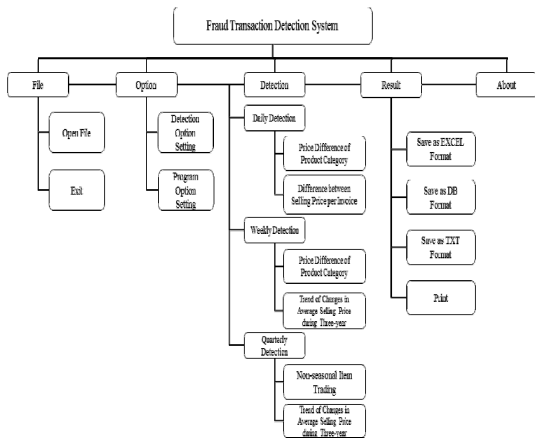
4.4 분기별 부정거래 탐지 결과

중도매인의 품목별 거래 빈도 탐지 결과 2008년 726건(평균 181.5건/분기), 2009년 1,069건(평균 267.3건/분기), 2010년 726건(평균 181.5건/분기)으로 탐지되었다. 분기를 고려한 판매 가격 변화 추이 분석은 수집된 분석 데이터를 고려하여 2010년 186건(평균 46.5건/분기)으로 탐지되었다.

5. 부정거래 탐지 시스템 프로토타입 (Prototype) 개발

본 연구에서는 이상치 검출 방안을 적용하여 부정거래를 탐지하는 규칙을 도출하였으나 실제 농산물 유통시장 환경은 거래 대상이 되는 상장예외품목의 변화와 부정거래 유형 변화 등 시간에 따른 변화가 가능하며, 이를 고려한 부정거래 탐지 규칙의 관리와 활용이 가능해야 한다.

따라서 본 연구에서는 부정거래 탐지 규칙의 지속적인 사용 환경 구축을 위하여 부정거래탐지 시스템의 프로토타입(prototype)을 개발하여 이를 제시한다. 제시한 시스템은 델파이(Delphi)를 기반으로 하는 범용 프로그램 개발 툴을 사용하여 개발하였다. 해당 시스템의 기능적인 개요는 다음과 같다(<Figure 3> 참조).



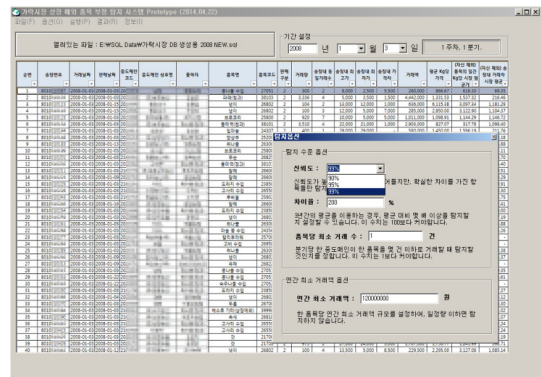
<Figure 3> Functional Decomposition of the Prototype Fraud Detection System

본 시스템은 크게 파일, 옵션, 실행, 결과, 정보라는 다섯 가지의 기능으로 구성되어있다. 이 중 본 탐지 시스템의 주요 기능은 연구 수행을 통하여 도출한 부정거래 탐지 규칙을 활용할 수 있는 탐지 옵션 설정과 실행 기능이다.

탐지 옵션 설정 기능은 기준을 설정할 수 있는 하위 창을 호출한다. 이 창에서는 부정거래 탐지 규칙 신뢰도, 차이율, 품목당 최소 거래 수, 연간 최소 거래액 항목들의 기준 값을 설정할 수 있다. 신뢰도는 품목별 가격 차이와 송품장별 판매 가격 차이에서 사용되는 Z^{se} 의 임계점을 결정하며, 차이율은 판매 가격 변화 추이의 기준을 결

정한다. 품목당 최소 거래 수는 비계절 품목 거래에서 분기별 거래량의 횡수를 결정하며, 연간 최소 거래액은 모든 부정거래 탐지 규칙에서 연간 거래액이 설정한 금액 이하인 품목들을 결과에서 제외한다. 사용자는 이를 통하여 분석 환경에 따른 탐지 대상 거래 데이터 선택이 가능하다. 실행 기능은 사용자가 도출한 부정거래 탐지 규칙들을 일간, 주간, 분기별로 탐지 명령을 내릴 수 있는 기능이다. 실행 기능의 메인 창에서는 사용자가 탐지를 원하는 기간을 설정하고 부정 탐지를 실행하면, 시스템은 탐지 옵션 설정 기능에서 설정된 기준을 참고하여 원하는 수준의 부정 거래들을 찾아준다. 본 시스템 프로토타입의 실제 화면 예시는 다음과 같다(<Figure 4> 참조).

본 연구에서 도출한 규칙을 사용하여 이를 시스템화함으로써, 반복적이고 정형화된 업무에 소요되는 시간과 노력은 최소화시키고, 의사결정과 같이 보다 가치 있는 업무에 더 많은 자원 배분이 가능 할 것이다. 또한 규칙에 대한 정의만으로도 부정거래 탐지 시스템의 사용이 가능하기 때문에, 사용자가 실제 업무에 대한 이해나



<Figure 4> A Screen of the Prototype Fraud Detection System

전문적 지식, 관련 경험이 부족하더라도 시스템에 원하는 명령을 내리거나 결과를 해석하는 정도의 보다 적은 양의 교육만으로도 해당 업무의 처리가 가능할 것이다.

6. 결론

본 연구에서는 국내 최대 농산물 유통 거래 시장에서의 2008년부터 2010년까지 상장예외품목 거래내역을 대상으로 부정탐지 분석 방안을 연구하였다. 이는 보험금 부당 청구나 탈세와 같이 사전에 명확하게 정의된 부정행위를 탐지하는 기존 연구와는 다르게 공급과 수요에 따른 변화가 민감한 품목의 실제 거래에 대하여 분석하고 부정 탐지 규칙을 도출하였으며, 전문가 검증을 통하여 부정거래 탐지 규칙의 유의성과 부정 탐지 결과의 신뢰성을 확보한 점이 기존 부정 탐지 관련 연구들과 비교하여 본 연구가 갖는 차이점이라 할 수 있다.

본 연구가 갖는 학문적 시사점으로는 부정거래의 품목별 가격 탐지와 송품장별 판매가격 차이 탐지 시 Z-Score의 개념을 활용하여 수정된 Z^{se} (Self-Eliminated Z-Score)를 사용하였다. Z^{se} 는 해당 품목의 전체 거래에서 비교하고자 하는 특정 거래를 제외한 거래 집합의 평균 가격에서 산출한 Z-Score로써, 이를 사용하여 부정거래의 평균 가격이 Z-Score의 기준이 되는 전체 평균 가격에 영향을 보다 적게 미치도록 하였다. 또한 탐지 규칙의 활용을 위한 부정거래 탐지 시스템의 프로토타입을 통하여 도출된 탐지 규칙들은 탐지 환경 변화에 따라 데이터마이닝의 패턴 학습이나 모형 생성과 같은 일련의 과정 없이 부정 유형에 따른 기준값 변화만으로도 부정 탐

지가 가능함을 제시한 점에서 실무적인 시사점을 갖는다고 할 수 있다.

본 연구의 한계점으로는 도매시장으로의 일별 농산물 반입 물량의 변화 주기가 잦아 정확한 측정 및 분석이 용이하지 않았다. 또한 농수산 품목이 갖는 특성으로 인하여 품목별 수확량의 정확한 예측이 요구되었으며, 거래 물량의 급격한 증감이 부정거래로 발생된 것이 아닌 생산지에서의 수확량 증감으로 인하여 발생된 경우일 수 있기 때문에 거래 물량을 고려한 부정 탐지 방안에 어려움이 존재하였다. 이 밖에도 분석에 필요한 기타 정산 기록 및 신뢰성 있는 데이터 확보에 어려움이 존재하여 다양한 부정거래 분석 방안 연구 및 개발에 제한이 있었다. 이러한 문제점으로 인하여 향후 연구 방안으로는 거래 물량에 대한 시계열 분석 기법 활용, 부정거래로 적발된 중도매인과의 거래인 또는 기타 이해 관계자들에 대한 사회네트워크 분석 기법 적용 방안과 같이 분석 데이터의 제약을 극복하고 데이터 특성을 고려한 실질적이고 효과적인 부정 탐지 방안의 연구로 진행되어야 할 것이다.

참고문헌(References)

- Abbasi, A., C. Albrecht, A. Vance, and J. Hansen, "MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud," *MIS Quarterly*, Vol.36, No.4(2012), 1293~1340.
- Bierstaker, J. L., R. G. Brody, and C. Pacini, "Accountants' perceptions regarding fraud detection and prevention methods," *Managerial Auditing Journal*, Vol.21, Issue.5(2006), 520~535.

- Chau, D. H. and C. Faloutsos, "Fraud detection in electronic auction," *Proceedings of the European Web Mining Forum at ECML/PKDD*, (2005), 87~97.
- Durtschi, C., W. Hillison, and C. Pacini, "The effective use of Benford's law to assist in detecting fraud in accounting data," *Journal of forensic accounting*, Vol.5, No.1(2004), 17~34.
- Fawcett, T. and F. Provost, "Adaptive fraud detection." *Data mining and knowledge discovery*, Vol.1, No.3(1997), 291~316.
- Ghosh, S. and D. L. Reilly, "Credit card fraud detection with a neural-network." *Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*, Vol. 3(1994), 621~630.
- Gwon, S. G. and C. H. Kim, "Current Status and Improvement Works of Unlisted Trading System [written in Korean]," *Proceedings of Winter Conference on Korea Food Marketing Association*, Vol.2011(2012), 37~65.
- Ham, S. O. and J. S. Hong, "A Study on the Fraud Detection of Industrial Accident Compensation Insurance," *Proceedings of Fall Conference on The Korean Operations Research and Management Science Society*, (2008), 342~345.
- Hong, C. H., S. M. Ahn, and K. W. Wee, "Detection of Stock Price Manipulation: A data Mining Approach," *Journal of Intelligence and Information Systems*, Vol.12, No.4(2006), 15~37.
- Kou, Y., C. T. Lu, S. Sirwongwattana, and Y. P. Huang, "Survey of Fraud Detection Techniques," *Proceedings of the 2004 IEEE international conference on Networking, sensing and control*, Vol.2(2004), 749~754.
- Lee, H. K., S. B. Han, and W. C. Jhee, "Illegal Cash Accommodation Detection Modeling Using Ensemble Size Reduction," *Journal of Intelligence and Information Systems*, Vol.16, No.1(2010), 93~116.
- Lee, J. W., D. H. Lee, and I. S. Kim, "Method of Detecting SmiShing using SVM," *Journal of Security Engineering*, Vol.10, No.6(2013), 655~668.
- Lee, J. W., I. A. Choi, W. C. Ji, H. Y. Park, and H. J. Shin, "A Quantitative Measure Integrating the Diverse Degree of Anomaly for Hospital Bill," *Proceedings of the Conference on Korean Institute of Industrial Engineers*, (2008), 342~345.
- Lee, M. G. and S. J. Chang, "The Necessity for Education to Deter Accounting Fraud and the Directions of Accounting Fraud Education," *The Korean Institute of Certified Accountants*, Vol.54, No.2(2012) 41~75.
- Lee, M. J., "Big Data and the Utilization of Public Data," *Internet and Information Security*, Vol.2, No.2(2011), 47~64.
- Li, S. H., D. C. Yen, W. H. Lu, and C. Wang, "Identifying the Signs of Fraudulent Accounts Using Data Mining Tech," *Computers in Human Behavior*, Vol.28, Issue 3(2012), 1002~1013.
- Linoff, G. S. and M. J. A. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Third Edition, Wiley Publishing, Inc., Indiana, 2011.
- Raj, S. B. E. and A. A. Portia, "Analysis on Credit Card Fraud Detection Methods," *Proceedings of 2011 International Conference on Computing, Communication and Electrical Technology*,

- (2011), 152~156.
- Ramamoorti, S., "The psychology and sociology of fraud: Integrating the behavioral sciences component into fraud and forensic accounting curricula." *Issues in Accounting Education*, Vol.23, No.4(2008), 521~533.
- Singleton, T. W. and A. J. Singleton. *Fraud auditing and forensic accounting*, 4th ed., John Wiley & Sons, New Jersey, 2010.
- Whiting, D. G., J. V. Hansen, J. B. McDonald, and C. Albrecht, "Machine Learning Methods for Detecting Patterns of Management Fraud," *Computational Intelligence*, Vol.28, No.4(2012), 505~527.
- Wi, T. S. and S. K. Kwon, "Reorganization of the Agricultural Wholesale Market," *Korean Journal of Food Marketing Economics*, Vol.26, No.3(2009), 75~93.

Abstract

A Study on the Application of Outlier Analysis for Fraud Detection: Focused on Transactions of Auction Exception Agricultural Products

Dongsung Kim* · Kitae Kim* · Jongwoo Kim** · Steve Park***

To support business decision making, interests and efforts to analyze and use transaction data in different perspectives are increasing. Such efforts are not only limited to customer management or marketing, but also used for monitoring and detecting fraud transactions. Fraud transactions are evolving into various patterns by taking advantage of information technology. To reflect the evolution of fraud transactions, there are many efforts on fraud detection methods and advanced application systems in order to improve the accuracy and ease of fraud detection. As a case of fraud detection, this study aims to provide effective fraud detection methods for auction exception agricultural products in the largest Korean agricultural wholesale market. Auction exception products policy exists to complement auction-based trades in agricultural wholesale market. That is, most trades on agricultural products are performed by auction; however, specific products are assigned as auction exception products when total volumes of products are relatively small, the number of wholesalers is small, or there are difficulties for wholesalers to purchase the products. However, auction exception products policy makes several problems on fairness and transparency of transaction, which requires help of fraud detection.

In this study, to generate fraud detection rules, real huge agricultural products trade transaction data from 2008 to 2010 in the market are analyzed, which increase more than 1 million transactions and 1 billion US dollar in transaction volume. Agricultural transaction data has unique characteristics such as frequent changes in supply volumes and turbulent time-dependent changes in price. Since this was the first trial to identify fraud transactions in this domain, there was no training data set for supervised learning. So, fraud detection rules are generated using outlier detection approach. We assume that outlier transactions have more possibility of fraud transactions than normal transactions. The outlier transactions are identified

* Dept. of Business Administration, Graduate School, Hanyang University

** Corresponding Author: Jongwoo Kim

School of Business, Hanyang University

222 Wangsimni-ro, Seongdong-gu, Seoul 133-791, Korea

Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

*** IPCS Consulting, President

to compare daily average unit price, weekly average unit price, and quarterly average unit price of product items. Also quarterly averages unit price of product items of the specific wholesalers are used to identify outlier transactions. The reliability of generated fraud detection rules are confirmed by domain experts.

To determine whether a transaction is fraudulent or not, normal distribution and normalized Z-value concept are applied. That is, a unit price of a transaction is transformed to Z-value to calculate the occurrence probability when we approximate the distribution of unit prices to normal distribution. The modified Z-value of the unit price in the transaction is used rather than using the original Z-value of it. The reason is that in the case of auction exception agricultural products, Z-values are influenced by outlier fraud transactions themselves because the number of wholesalers is small. The modified Z-values are called Self-Eliminated Z-scores because they are calculated excluding the unit price of the specific transaction which is subject to check whether it is fraud transaction or not.

To show the usefulness of the proposed approach, a prototype of fraud transaction detection system is developed using Delphi. The system consists of five main menus and related submenus. First functionalities of the system is to import transaction databases. Next important functions are to set up fraud detection parameters. By changing fraud detection parameters, system users can control the number of potential fraud transactions. Execution functions provide fraud detection results which are found based on fraud detection parameters. The potential fraud transactions can be viewed on screen or exported as files.

The study is an initial trial to identify fraud transactions in Auction Exception Agricultural Products. There are still many remained research topics of the issue. First, the scope of analysis data was limited due to the availability of data. It is necessary to include more data on transactions, wholesalers, and producers to detect fraud transactions more accurately. Next, we need to extend the scope of fraud transaction detection to fishery products. Also there are many possibilities to apply different data mining techniques for fraud detection. For example, time series approach is a potential technique to apply the problem. Even though outlier transactions are detected based on unit prices of transactions, however it is possible to derive fraud detection rules based on transaction volumes.

Key Words : Fraud Detection, Outlier Detection, Auction Exception Products

Received: June 17, 2014 Revised: September 11, 2014 Accepted: September 12, 2014

저 자 소개



김 동 성

현재 한양대학교 일반대학원 경영학과 경영정보시스템 전공 석사과정에 재학 중이며, 미래창조과학부에서 지원하는 클라우드 서비스 정책 연구 센터에서 소속되어 연구를 수행하고 있다. 협성대학교 경영정보학과에서 학사를 마쳤으며, 주요 연구 관심분야는 데이터마이닝 기법과 응용, 빅 데이터, 오피니언 마이닝, 사회 네트워크 분석, 클라우드 컴퓨팅 서비스 등이다.



김 기 태

현재 한양대학교 일반대학원 경영학과 경영정보시스템 전공 석사과정에 재학 중이며, 미래창조과학부에서 지원하는 클라우드 서비스 정책 연구 센터에 소속되어 연구를 수행하고 있다. 한양대학교에서 경영학과와 정보시스템학과를 취득하였으며, 주요 연구 관심분야는 데이터마이닝 기법과 응용, 빅 데이터, 클라우드 컴퓨팅 서비스, 시스템 분석 및 설계 등이다.



김 종 우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 오피니언 마이닝, 상품추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석, 클라우드 컴퓨팅 서비스 등이다.



박 성 기

현재 본인이 창업한 농업유통분야 전문컨설팅 업체인 IPCS의 대표이사로 재직 중이다. 연세대학교 경영학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 광운대학교에 박사학위를 수료하였다. 주요 컨설팅 분야는 농업 유통 분야 정책 및 제도 분야, 정보시스템 분야이며, 국내뿐만 아니라 개발도상국 대상의 정보시스템 컨설팅도 수행 중에 있다.