

사업 아이디어 매력도 평가를 위한 검색 데이터 활용에 관한 연구

심재후*, 최명길**

*중앙대학교 대학원 창업학과

**중앙대학교 사회과학대학 상경학부

e-mail: echangup@wm.cau.ac.kr

A Study on Use of Search Data for Evaluation of Business Idea Attractiveness

Jaehu Shim*, Myeonggil Choi**

*Dept. of Entrepreneurship, Chung-Ang University

**Dept. of Business Administration, Chung-Ang University

요 약

성공적인 창업을 위해서는 창업가의 준비가 선행되어야 하지만, 매력적인 사업 아이디어의 개발이 뒤따라야 한다. 그러나 지금까지의 창업연구는 창업행동과 사업성과에 영향을 미치는 창업가 요인에 치우쳐 있으며, 사업 아이디어의 개발과 평가에 대한 연구는 부족한 실정이다. 이 연구는 고객이 상품을 구매하기 전 인터넷 검색엔진에서 해당 상품에 대한 검색을 하는 경우가 일반화되고 있다는 사실과 고객이 검색엔진에 입력하는 키워드는 고객의 의도를 대변한다는 사실을 기초로, 키워드로 표현된 사업 아이디어의 매력도를 객관적으로 측정하는 방법을 제시하는 것을 목적으로 한다.

이 연구는 키워드로 표현된 사업 아이디어 매력도(BIA)를 구매의도를 가진 잠재고객의 자사 웹 사이트 방문수로 정의한다. 키워드로 표현된 사업 아이디어 매력도(BIA)는 [해당 키워드의 조회수(Q) × 구매의도 비율(R) / 경쟁 사이트의 수(S)]의 수식으로 나타낼 수 있으며, 수식을 구성하는 변수 중에서 해당 키워드의 조회수(Q)와 경쟁 사이트의 수(S)는 검색엔진에서 쉽게 제공 받을 수 있으므로, 구매의도 비율(R)만 알 수 있다면 BIA를 비교적 정확히 추정할 수 있다.

연구자는 특정 분야 키워드 100개를 선정하 다음, 전문가로 하여금 각 키워드의 구매의도 비율(R)을 추정하게 하고, 전문가 추정 없이도 구매의도 비율을 예측할 수 있도록 각 키워드의 구매의도 비율(R)을 예측하는 주요 데이터를 의사결정 나무 기법으로 도출하고, 의사결정 나무 기법으로 도출된 데이터로 구성된 회귀식을 제시함으로써 키워드로 표현된 사업 아이디어 매력도(BIA)를 객관적으로 평가하는 방법을 제시한다. 이 연구는 사업 아이디어의 개발과 평가에 대한 객관적인 기준을 제시함으로써 창업의 성공률을 높이는 데 기여할 수 있고, 창업연구에 새로운 방법론을 도입했다는 점에서 의의가 있다.

1. 연구의 배경과 목적

성공적인 창업을 위해서는 창업가의 준비가 선행되어야 하지만, 매력적인 사업 아이디어의 개발이 뒤따라야 한다. 그러나 지금까지의 창업연구는 창업행동과 사업성과에 영향을 미치는 창업가 측면의 특성에 치우쳐 있으며, 사업 아이디어의 개발과 평가에 대한 연구는 부족한 실정이다.

Ireland 등이 조사한 바에 따르면 1963년부터 2005년까지 Academy of Management Journal에 게재된 창업(Entrepreneurship) 관련 논문 52편 중에서 창업가의 행동(17편), 조직의 성장(10편) 및 회사의

성과(10편)에 대한 연구가 대다수를 차지하고 있으며, 사업 아이디어 개발과 평가에 대한 연구는 전혀 없다[5].

이 연구는 고객이 상품을 구매하기 전 인터넷 검색엔진에서 해당 상품에 대한 검색을 하는 경우가 일반화되고 있다는 사실과 고객이 검색엔진에 입력하는 키워드는 고객의 의도를 대변한다는 사실을 기초로, 키워드로 표현된 사업 아이디어의 매력도를 객관적으로 측정하는 방법을 제시하는 것을 목적으로 한다.

2. 이론적 배경

2.1. 사업 아이디어 매력도의 정의

아키야마(秋山) 등은 인터넷이 가져온 고객의 구매행동 변화를 설명하면서, 기존의 AIDMA 모델을 대체하는 AISAS 모델을 제안한다. 즉, 상품에 대한 주목(Attention) → 흥미(Interest) → 욕구(Desire) → 기억(Memory) → 구매(Action)에 이르는 일련의 구매과정이 인터넷 등장 이후에 상품에 대한 주목(Attention) → 흥미(Interest) → 검색(Search) → 구매(Action) → 정보공유(Share)로 바뀌고 있다는 것으로, AISAS 모델은 구매 이전 검색(Search)의 중요성과 구매 이후 정보공유(Share)의 중요성을 강조한다[4].

김연호는 검색엔진에서 제공하는 특정 키워드에 대한 조회수와 경쟁 사이트 수를 이용하여 잠재 수요 고객을 추정할 수 있다고 주장한다[1].

Jansen 등은 검색엔진에서 사용자의 검색 의도는 정보적(informational), 항해적(navigational), 거래적(transactional) 의도로 구분할 수 있다고 주장하면서, 검색엔진에서 제공받은 152만 여개의 검색 로그를 분석한 결과 80.6%가 정보적 의도, 10.2%가 항해적 의도, 나머지 9.2%가 거래적 의도라고 분석했다[6].

이러한 주장을 종합하면 인터넷의 등장으로 잠재 고객들은 상품 구매 전 인터넷 검색엔진에서 검색을 하는 경우가 많아졌으며, 검색엔진이 제공하는 특정 키워드에 대한 조회수와 경쟁 사이트 수 및 해당 키워드 검색 시 상품을 구매하고자 하는 거래적 의도를 분석함으로써 사업 아이디어의 매력도를 평가하는 것이 가능하다는 추론이 가능하다.

이 연구는 키워드로 표현된 사업 아이디어 매력도(BIA: Business Idea Attractiveness)를 '구매의도를 가진 잠재고객의 자사 웹 사이트 방문수'로 정의하고, 아래의 수식으로 나타낸다.

$$BIA = \frac{Q \times R}{S}$$

Q: 해당 키워드에 대한 조회수

R: 전체 조회수 중 거래적 의도 비율

S: 해당 키워드에 대한 경쟁 사이트의 수

해당 키워드에 대한 조회수(Q)와 경쟁 사이트의 수(S)는 검색엔진에서 쉽게 제공 받을 수 있으므로, 구매의도 비율(R)만 알 수 있다면 사업 아이디어 매력도(BIA)를 비교적 정확히 추정할 수 있다.

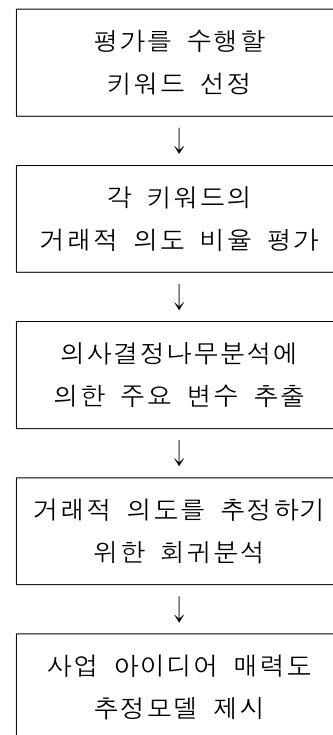
2.2. 웹 마이닝의 정의와 분류

웹 마이닝(Web mining)은 웹 구조를 검색하고 웹 내용의 중요도를 등급 매기고 웹 내용의 규칙성과 역동성을 발견하고 웹 접근 패턴을 마이닝하는 작업이다. 일반적으로 웹 마이닝 과제는 웹 콘텐츠 마이닝(Web content mining), 웹 구조 마이닝(Web structure mining), 웹 사용 마이닝(Web usage mining)의 세 가지 범주로 분류될 수 있다[3].

웹 서비스의 일부인 검색엔진에서 '사업 아이디어 매력도'라고 하는 유용한 정보를 추출하는 과정은 웹 마이닝 과정이라고 할 수 있고, 웹 마이닝 중에서도 웹 사용 마이닝에 가깝다고 할 수 있다. 그러나 사용자의 검색엔진 로그(사용기록) 및 웹 로그를 토대로 정보를 추출하는 기존의 웹 사용 마이닝과는 다른 방법으로 접근하고 있다는 점에서 본 논문의 방법론이 차별화 된다.

3. 연구방법

연구자는 아래와 같은 순서로 연구를 수행했다.



3.1. 키워드 선정과 전문가 평가

연구자는 ‘영어’ 관련 키워드 100개를 선정한 다음, 전문가로 하여금 각 키워드 조회 시 사용자들의 거래적 의도 비율(R)을 평가하게 했다.

이미 알려진 결과를 예측하는 데 적합한 변수를 추출하고 모델링하는 지도학습(Supervised learning)을 활용한 데이터 마이닝 등, 일반적인 데이터 마이닝 작업의 경우, 전문가에 의한 변수 평가는 필요하지 않다. 각 키워드에 대한 거래적 의도 비율은 이미 알려져 있는 데이터가 아니기 때문에, 전문가로 하여금 각 키워드에 거래적 의도 비율을 평가하게 한 다음, 이를 추정하는 주요변수를 추출하고 회귀식을 도출한다.

3.2. 의사결정나무분석에 의한 주요변수 추출

의사결정나무분석(decision tree induction)은 클래스 식별이 있는 훈련용 데이터에서 의사결정나무를 학습하는 것이다. 의사결정나무는 흐름도와 유사한 나무구조이며, 내부노드에는 속성에 대한 검사를 표시하고, 각 가지는 검사의 결과를 나타내며, 잎 노드는 클래스 레이블의 분포를 나타낸다. 의사결정나무가 구축될 때, 속성선택 척도는 서로 다른 클래스 데이터를 가장 잘 분리하는 속성을 선택하는 데 사용된다[3].

본 연구는 키워드 길이, 키워드에 포함된 단어 수 등 키워드 내부 요소와 특정 기간 동안 검색엔진에서의 키워드 조회수, 블로그 글 수, 뉴스 기사수 등 키워드 관련 정보의 수 등 키워드 외부 요소를 의사결정나무 구축에 투입하여 어떤 속성이 거래적 의도 비율과 연관되어 있는지 확인한다.

3.3. 회귀분석과 사업 아이디어 매력도 추정

연구자는 의사결정나무분석에서 추출된 주요변수를 설명변수로 하고 거래적 의도 비율을 종속변수로 하는 다중 선형 회귀분석을 실시한다.

다중 선형 회귀분석(multiple linear regression model)은 정량적인 종속변수와 여러 독립변수 사이의 선형관계를 모델링하기 위해 사용된다. 이를 수식으로 나타내면 다음과 같다[2].

$$R = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \epsilon$$

R: 전체 조회수 중 거래적 의도 비율

β: 각 독립변수의 회귀계수

ε: 잔차

4. 연구결과

3.3.에서 도출한 거래적 의도 비율(R)을 설명하는 회귀식을 사업 아이디어 매력도(BIA: Business Idea Attractiveness) 추정 공식에 대입하면 다음과 같다.

$$BIA = \frac{Q}{S}(\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \epsilon)$$

Q: 해당 키워드에 대한 조회수

S: 해당 키워드에 대한 경쟁 사이트의 수

β: 각 독립변수의 회귀계수

ε: 잔차

이 추정공식을 활용하여 키워드의 조회수, 해당 키워드에 대한 경쟁 사이트의 수 및 검색엔진에서 확인할 수 있는 데이터를 활용하여 누구나 손쉽게 사업 아이디어의 매력도를 평가할 수 있다.

5. 결론

이 연구는 키워드로 표현된 비즈니스 아이디어에 대해 전문가가 추정한 구매의도 비율을 예측하는 데 활용될 수 있는 주요 데이터를 의사결정나무기법으로 도출하고, 의사결정 나무 기법으로 도출된 검색 데이터로 구성된 사업 아이디어 매력도 추정 회귀식을 제시함으로써 키워드로 표현된 사업 아이디어 매력도(BIA)를 평가하는 방법을 제시한다.

이 연구는 개발된 사업 아이디어의 매력도를 누구나 쉽게 평가할 수 있는 객관적인 방법을 제시함으로써 창업의 성공률을 높이는 데 기여할 수 있고, 창업연구에 새로운 방법론을 도입했다는 점에서 의의가 있다.

참고문헌

- [1] 김연호, 인터넷 게릴라 마케팅, e비즈니스, 2009.
- [2] Galit Shmueli 등저, 선택수 등역, 비즈니스 인텔리전스를 위한 데이터마이닝, 사이텍미디어, 2009.
- [3] Han J. 등저, 강창완 등역, 데이터 마이닝, 사이플러스, 2007.
- [4] 秋山隆平 등, ホリスティック コミュニケーション, 2004.

- [5] Ireland, R. D., et al., "Entrepreneurship Research in AMJ: What Has Been Published, and What Might the Future Hold?", *Academy of Management Journal*, Vol. 48, No. 4, pp. 556-564, 2005.
- [6] Jansen, B. J., et al., "Determining the Informational Navigational and Transactional Intent of Web Queries", *Information Processing and Management* Vol. 44, pp. 1251-1266, 2008.