

사용자의 행동과 점진적 기계학습을 이용한 쓰레기 편지 여과 시스템의 설계

김강민, 박은진, 김재훈
한국 해양대학교 컴퓨터 공학과

shoutout@mail.hhu.ac.kr bakeunjin@yahoo.co.kr jhoon@mail.hhu.ac.kr

Designing a Spam Mail Filtering System Using User Reaction and Incremental Machine Learning

Kang-Min Kim, Eun-Jin Park, Jae-Hoon Kim
Department of Computer Engineering, Korea Maritime University

요 약

본 논문은 쓰레기 편지를 여과하기 위해 대상 편지에 따른 사용자들의 행동(reaction)을 묵시적(implicitly)으로 수집한 후 이를 점진적(incrementally) 기계학습기의 자질(feature)로 사용하여 편지 여과 작업의 증거가 되는 단어들을 지속적으로 학습하면서 최적의 편지 여과 결과를 제공하는 기법과 시스템 구조를 제안한다. 사용자 개인의 컴퓨터에 행동 정보와 학습 데이터를 저장하도록 설계하여 묵시적 정보 수집에서 자주 제기되는 개인 프라이버시 문제를 해결하였으며, 점진적 기계학습 기법을 사용하여 개인 정보를 포함하는 대량의 편지 학습 데이터를 모으기 힘들다는 문제를 해결하였다. 또 향후 제안하는 시스템을 이용하여 여러 종류의 기계학습 기법 중 쓰레기 편지 여과 작업을 가장 효과적으로 수행할 수 있는 기법을 선택하는 작업을 수행할 계획이다.

1. 서 론

인터넷의 급속한 성장과 더불어 전자 편지 서비스(E mail service)는 이제 의사교환의 필수적인 매체로 사용되고 있다. 그러나 대량의 정보를 적은 비용으로 전달할 수 있다는 전자 편지의 장점에도 불구하고, 사용자들은 매일 편지 보관함에서 각종 성인 정보 또는 원하지 않는 광고 편지 등 이른바 ‘쓰레기 편지’를 선별해 내는 일로 하루 일과를 시작하는 실정이다. 이러한 피해를 방지하기 위해 많은 연구가 이루어졌고[1-4], 또 여러 기법을 사용한 시스템이 제안되었다[4-6]. 그 중에서도 관심을 가질 만한 것은 쓰레기 편지와 정보성 편지에 대한 사용자들의 행동(reaction)이 각각 다른 점을 이용, 사용자의 행동을 묵시적(implicit)으로 관찰한 후 이 정보를 편지 여과 작업에 활용하는 것이다[7-10]. 예를 들어 사용자는 쓰레기 편지로 간주되는 편지는 읽지 않고 제목만

확인하고 바로 삭제할 수 있다. 또 중요한 정보 편지일 경우 사용자는 편지를 따로 보관함을 만들어서 보관하거나, 다른 사람에게 전달 또는 답장을 보내는 행동을 하는 것이 보통이다. 시스템은 이러한 사용자의 행동을 저장하였다가 이후 여과 작업에 행동 정보를 이용하는 것이다. 이는 사용자에게 여과 규칙을 직접 마련하거나, 도착한 편지를 직접 판단하여 시스템에게 알려주어야 하는 등의 직접적인 참여를 요구하지 않는다는 점에서 충분히 매력이 있다. 하지만 사용자의 행동을 ‘관찰한다(observng)’는 점은 개인의 프라이버시 문제를 제기하였는데[11], 이러한 문제는 사용자의 행동에 관련되는 정보를 서버 측이 아닌 사용자의 컴퓨터에 저장하도록 설계하면 어느 정도 해결이 된다.

쓰레기 편지 여과 시스템의 경우 시대의 흐름을 반영할 수 있도록 하는 요소가 필수적이다. 시간이 지남에 따라 다양한 형태의 쓰레기 편지들이 새롭게 나타나게 되고, 여과 시스템은 이러한 쓰레기 편지를 사용자의 특별한 조작 없이 적절하게 처리할 수 있는

기능을 갖추어야 한다. 따라서 쓰레기 편지의 분류 작업에 사용되는 모델을 자동으로 갱신할 수 있도록 점진적 기계학습 기법을 이용하였다[12]. 일반적으로 학습을 위해서는 미리 마련된 학습 데이터가 필요하지만 편지의 경우 개인적인 내용이 포함된 경우가 많으며, 같은 편지가 사용자에게 따라 쓰레기 편지가 될 수도 있고, 정보성 편지가 될 수도 있기 때문에 학습 데이터를 마련하는 것이 쉽지 않다. 하지만 미리 마련한 대량의 학습 데이터를 통해 한꺼번에 학습을 하는 방식 대신, 새롭게 수신되는 편지들과 그 편지들에 대한 사용자의 행동을 이용하여 여과를 위한 분류 모델을 점진적으로 학습한다면 대량의 학습 데이터를 마련해야 하는 고민을 해결할 수 있다[13].

결국 사용자는 무의식적으로 편지 여과 시스템을 사용해 가면서 점점 정확한 여과 결과를 제공 받게 되는 것이다.

본 논문의 흐름은 다음과 같다 우선 2 절에서는 앞서 설명한 요소들의 관련 연구들에 대해 간단히 설명하고, 3 장에서는 제안하는 편지 여과 시스템의 전체적인 구성과 주요 모듈의 개념 및 기능을 설명한다 4 장에서는 결론 및 향후 연구 방향을 제시하고자 한다.

2. 관련 연구

2.1 묵시적 피드백(implicit feedback)

편지 여과 시스템에서 대상 편지가 쓰레기 편지인지 정보 편지인지 여부를 판별하기 위해 시스템은 사용자의 행동을 관찰하고, 그 결과를 이용하여 여과 작업을 수행하게 되는데, 이를 ‘묵시적 피드백(implicit feedback)’ 이라 한다[10]. 우선 Morita 와 Shinoda 는 그들의 연구[8]에서 사용자가 문서를 읽는데 소비하는 시간과 정보 요구 사이의 상관관계를 발견하였으며, ‘GroupLens’ 시스템은 ‘문서를 읽는데 투자하는 시간’ 을 곧 사용자의 문서에 대한 관련 정도로 평가하였다. 또 다른 사용자 행동 요소로 저장, 삭제, 출력 또는 전달 등의 행동 정보를 이용하기도 하는데, [9]에서는 사용자의 하이퍼 텍스트(hyper text)에 대한 반응, 마우스 스크롤 행동을 포함하는 브라우징 행동 등을 관찰하여 여과작업을 수행하기도 한다. 특히 Oard 는 최근 사용자의 행동을 검토(examine), 저장(retain), 참조(reference), 주석 첨가(annotate)로 분류하고 이를 묵시적 피드백의 증거 데이터로 활용하는 방법을 제안하였다[7].

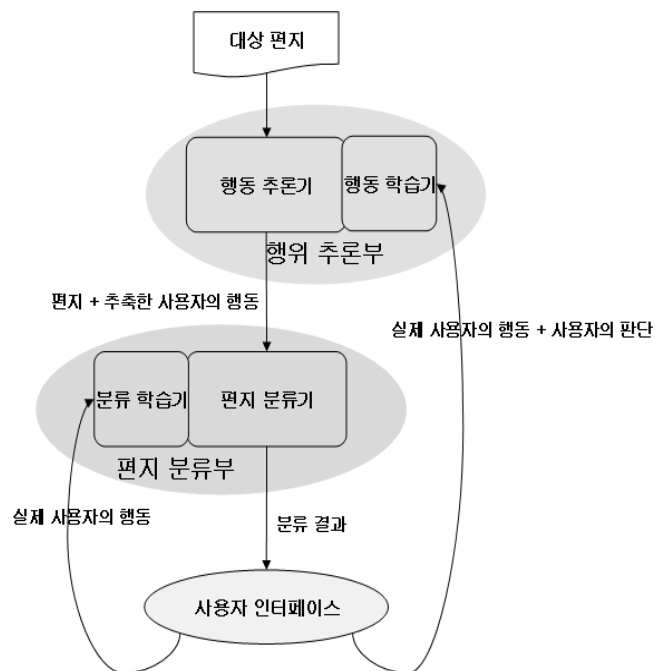
2.2 기계학습(machine learning)

여러 가지 기계학습 기법을 이용하여 자동적으로 모델의 갱신이 가능한 지능적(intelligent) 편지 여과 시스템을 설계할 수 있다. 그 중에서도 많이 사용되고 있는 기법으로 나이브 베이시안 분류자(naïve Bayesian classifier)가 있다[12]. 이것은 규칙 기반 여과 시스템의 한계를 벗어나기 위해 문서 분류, 쓰레기 여과 시스템에 사용되고 있는 대표적인 기계학습 방법 중의 하나이다. 최근에는 베이시안 분류 기법에 가중치가 부여하는 쓰레기 편지 여과 시스템도 제안되었다[14].

또한 편지는 지극히 개인적인 사생활 정보를 많이 포함하고 있기 때문에 학습을 위한 데이터를 모으는 것이 쉽지 않다. 따라서 대량의 학습 데이터를 통해 한꺼번에 학습하는 것이 아니라 실시간으로 학습 모델을 갱신시켜 나가는 점진적 학습(incrementally learning) 기법을 사용하는 편이 유리하다. 실제로 최근 전자 상거래 서비스를 제공하는 업체에서는 고객의 행동(구매 활동, 관심 물품 선정 등)을 지속적으로 관찰하여, 최적의 서비스를 제공하기 위해 점진적 기계학습 기법을 활용하고 있다[14].

3. 쓰레기 편지 여과 시스템

제안하는 시스템의 전체적인 골격은 [그림 1]과 같다.



[그림 1] 쓰레기 편지 여과 시스템의 전체 구조

시스템은 행동 추론부(3.3 절에서 자세히 설명)를 통해 편지 내용으로부터 사용자의 행동 정보를 추론한다. 편지 분류부는(3.4 절에서 자세히 설명) 추론된 사용자의 행동과 편지의 내용을 이용하여 사용자에게 분류된 결과를 제공한다. 이 두 부분은 각각 사용자의 실제 행동 데이터를 전달 받아 행동 추론 모델과 편지 분류 모델을 점진적으로 학습해 나가는 학습기를 내장하고 있다.

3.1 사용자 행동 모델링과 학습 대상 결정

본 논문은 수신된 편지에 대한 사용자의 행동 패턴을 다음과 같이 정의하였다.

- A. 읽기(Reading) + 삭제(Delete)
- B. 읽기(Reading) + 분류(classification)
- C. 읽기(Reading) + 전달(Forward)
- D. 읽기(Reading) + 답장(Reply)
- E. 무시(Ignore) + 분류(classification)
- F. 무시(Ignore) + 삭제(Delete)

- G. 쓰레기 편지에서 정보 편지로 이동
- H. 정보 편지 -> 쓰레기 편지로 이동

사용자는 쓰레기 편지 또는 정보 편지 각각에 대해 상이한 반응을 보인다. 시스템은 사용자 인터페이스를 통해 사용자의 행동 데이터 수집하고, 이 데이터를 자질 벡터에 전달한다. 행동 추론부는 편지에서 추출한 편지 데이터(제목, 내용, 헤더정보에서 추출한 단어)를 학습 데이터로 삼아서 A-F의 데이터를 추론하는 학습 모델을 내장하고 있다. 이에 비해 편지 분류부는 메일 데이터와 A-F의 행동 데이터를 학습 데이터로 삼아서 G, H를 결정하는 학습 모델을 내장하고 있다. 다시 말해 편지 분류부는 사용자가 명확하게 쓰레기 편지와 정보 편지를 구분하는 행동을 할 경우에만 학습이 가능하다.

3.2 자질 벡터(feature vector)의 구성

자질 벡터는 편지를 분류하기 위한 입력 데이터로서 3.1 절에서 정의된 사용자의 행동 정보를 저장한 ‘행동 벡터부’와 편지에서 추출한 정보의 모델인 ‘편지 벡터부’로 구성되어 있다([그림 2] 참조). 편지 벡터부는 대상 편지에서 추출한 정보로 다음과 같은 정보를 저장한다.

- A. 제목에서 추출한 각 단어 빈도
- B. 본문에서 추출한 각 단어 빈도
- C. 보내는 사람에서 추출한 단어 빈도
- D. 받는 사람에서 추출한 단어 빈도
- E. 편지 본문 URL에서 추출한 단어 빈도
- F. 편지 본문 내의 이미지 파일의 개수

편지A	...	편지F	행동.위기	...	행동.삭제	...
4	...	3	1	...	0	...

← 편지 벡터부
→ 행동 벡터부

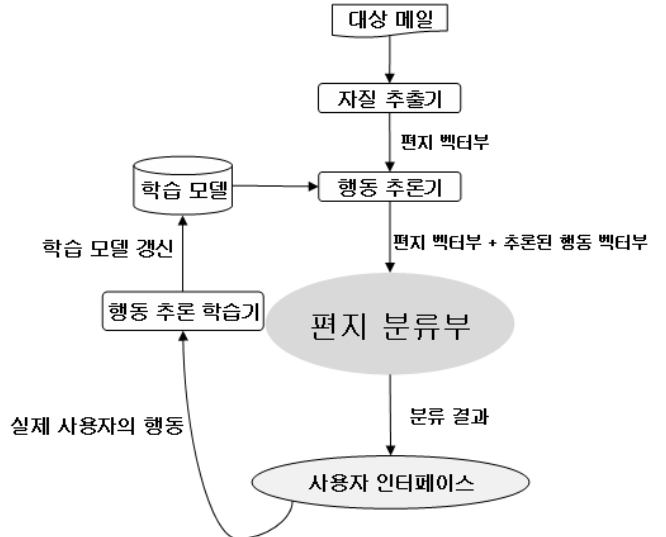
[그림 2] 자질 벡터의 구조

그리고 같은 편지에서 추출한 같은 단어라도 그 단어의 위치에 따라 그 비중이 다르다. 예를 들어 제목에서 추출한 단어는 본문에서 추출한 단어에 비해 비중이 높을 수 있다. 따라서 단어의 위치에 따라 차별적으로 가중치를 적용하여 여과 성능을 향상시킬 수 있다. 또 편지 본문 내에 이미지 파일이 많을 경우, 쓰레기 편지일 가능성이 높아진다는 경험적 논리에 근거하여 편지 본문 내의 이미지 파일의 개수도 편지의 정보성 여부를 결정하기 위한 증거로 삼았다.

3.3 행동 추론부

행동 추론부는 편지에서 추출한 편지 데이터(제목, 내용, 헤더정보에서 추출한 단어의 빈도)를 이용하여 사용자의 행동을 추론하는 기능을 수행한다. 자세히 설명하자면 편지 분류부는 여과 작업을 수행하기 위해 자질 벡터를 입력으로 한다. 자질 벡터는 대상 편

지에서 직접 추출한 정보와 사용자의 행동 정보로 구성되는데, 입력이 되는 대상 편지는 당연히 사용자 행동 정보가 없다. 따라서 편지의 내용으로 사용자의 행동 정보를 추론할 기능이 요구되고, 이 기능을 행동 추론부가 수행한다([그림 3]참조).

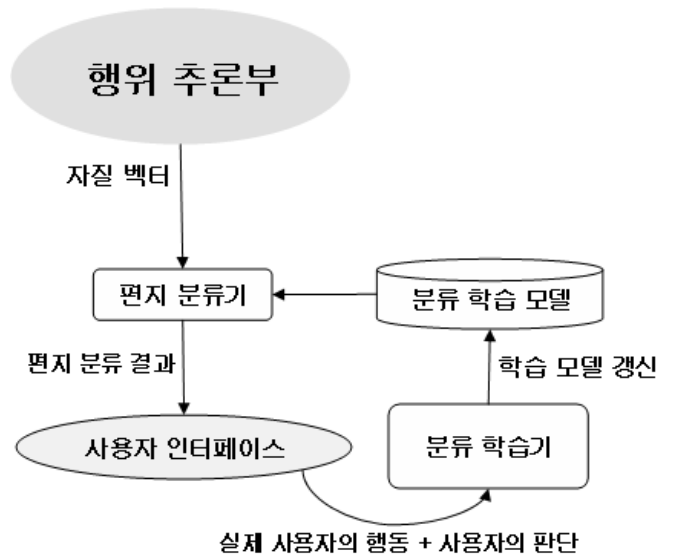


[그림 3] 행동 추론부의 구성

[그림 3]의 행동 추론 학습기는 편지 내용에서 추출한 정보와 실제 사용자의 행동을 학습 데이터로 삼아서 사용자의 행동을 추론하고 추론된 정보로 학습 모델을 수정하는 기능을 수행한다.

3.4 편지 분류부

편지 분류부는 편지 벡터부와 행동 벡터부로 구성된 자질벡터를 이용하여 대상 편지를 ‘쓰레기 편지’ 또는 ‘정보 편지’로 분류하는 기능을 수행한다. 사용자 인터페이스는 사용자에게 편지 여과 결과를 제공하고, 사용자의 행동을 수집하여 분류 학습기의 학습 데이터로 제공한다([그림 4]참조).



[그림 4]편지 분류부의 구성

[그림 4]의 분류 학습기는 편지로부터 추출한 정보와 사용자의 행동 정보와 실제 사용자의 결정(쓰레기 편지와 정보성 편지를 직접 구분하는 행동)을 학습 데이터로 삼아서 이후 사용자의 결정이 빠져있는 데이터로부터 이를 추론한다. 결과적으로 이것은 변화하는 사용자의 정보 요구를 지속적으로 학습하고 이전의 학습 모델을 갱신하는 기능을 수행하게 되는 것이다.

4. 결론 및 향후 연구방향

본 논문에서는 사용자의 행동 정보를 활용하여 분류 모델을 점진적으로 학습해 나가면서 다음에 수신되는 편지를 좀더 정확하게 여과하는 시스템을 제안하였다.

향후에는 쓰레기 편지와 정보성 편지의 분류에 활용할 수 있는 다양한 사용자 행동 증거들이 존재한다는 사실에 착안하여 이러한 사용자의 행동을 심층적으로 연구하고, 최고의 여과 성능을 발휘할 수 있도록 각 행동 증거들 간의 가중치를 조절한다. 마찬가지로 다양한 기계학습 기법 중에서 쓰레기 편지 여과 작업을 가장 효과적으로 수행하는 기법을 찾아서 편지 분류기의 성능을 좀 더 향상시킬 수 있다.

참고문헌

- [1] Uri Hanani, Bracha Shapira and Peretz Shoval., "Information filtering: Overview of issues, research and systems", *User Modeling and User - Adapted Interaction. Dordrecht*, vol.11, no.3, pp.203-259, 2001.
- [2] Douglas W. Oard and Gary Marchionini, *A conceptual framework for text filtering*, Technical Report, CLIS-TR-96-023, College of library and information service, University of Maryland, (<http://www.ee.umd.edu/med-lab/filter/papers/filter.ps>), 1996.
- [3] 한국 정보보호 진흥원, 이메일 추출 방지 프로그램의 원리및 기능분석, 2002.
- [4] Paul Wolfe, Charlie Scott and Mike Erwin, *Anti-spam tool kit*, McGraw-Hill Osborne Media, 2004.
- [5] 한국 정보보호 진흥원, 스팸메일 차단 솔루션 활용 가이드, 2004.
- [6] 한국 정보보호 진흥원, 이메일 서비스 업체별 스팸 방지기술 활용 현황, 2002.
- [7] Kim Jinmook and Douglas W. Oard, "Observable behavior for implicit user modeling: A framework and user studies", *한국문헌정보학회지*, vol.35, no.3, pp. 173-189, 2001.
- [8] Masahiro Morita and Yoichi Shinoda, "Information filtering based on user behavior: Analysis and best match text retrieval", *Proceedings of SIGIR*, pp. 272-281, 1994.
- [9] Jeremy Goecks, and Jude Shavlik, "Learning users' interests by unobtrusively observing their normal behavior", *Proceedings of the 5th International Conference on Intelligent User Interfaces*, pp. 129-132, 2000.
- [10] Judd D. Antin, *Information filtering behavior: Understanding sociocultural and behavioral adaptations for information filtering in everyday life*, 2004.
- [11] Douglas W. Oard, Anton Leuski, and Stuart Stubblebine, "Protecting the privacy of observable behavior in distributed recommender systems", *Proceedings of SIGIR'03 Workshop on Implicit Measures of User Interests and Preferences*, 2003.
- [12] Thomas M. Mitchell, *Machine Learning*, McGraw-Hill Higher Education, 1997.
- [13] Ray J. Solomonoff. "Progress in incremental machine Learning", pre-liminary report(<http://world.std.com/rjs/nips02.pdf>), 2002.
- [14] 김현준, 정재은, 조근식, "가중치가 부여된 베이지안 분류자를 이용한 스팸 메일 필터링 시스템" *한국정보과학회 논문지* vol. 31, no. 8, pp. 1092-1100, 2004.