

通信販売における顧客の名寄せ知識の構築

The construction of knowledge for customers name collection in the mail order

東 弘樹^{*1}
Hiroki Azuma

津田 和彦²
Kazuhiko Tsuda

^{*1} HAZS株式会社
Hazs Corporation

^{*2} 筑波大学
Graduate School of Business Sciences, University of Tsukuba

The unpaid customers in a mail order swing the descriptions of addresses, in order to be unable to identify an individual.. This paper presents investigating detect a difference on how to write the paid customer and unpaid customer, and construction of this knowledge.

1. はじめに

通信販売業界は、Web やモバイルなど申込手段の拡充と共に、ここ10年の長期デフレ環境下においても、売上は10年以上伸び続けている。(社)日本通信販売協会の2011年度の売上高調査結果[(社)日本通信販売協会 2012]によると、2011年度の通信販売業界全体の売上高は、推計で5兆0,900億円、対前年度の伸び率も9.0%を記録している。売上を拡大する要因のひとつに決済方法の多様化がある。顧客の利便性を高めるという事もさることながら、商品が届いてから決済を行う後払い方式は、顧客から見た取引の安全性をアピールすることができ、大幅な増収・増益が見込めることができる。

しかし後払い方式では、一般的に商品と共に振込用紙などを届けるため、代金の支払いは顧客の良心に依存しており、遅れや未払いが発生するという問題が生じる。社団法人日本通信販売協会によると、貸倒率は約0.5%と報告している。この数字は、督促状や課電などの債権回収をした後のもので、初期段階では、この数倍の未払いが生じている。

その中でも悪質なものの多くは、商品の届出先や連絡先の名前や住所の表記を揺らし、名寄せできないようにして申し込んでいる。この表記の揺らし方法が巧みになっており、現在利用されている販売管理システムなどの検索機能では、名寄せの精度が低く、対応しきれない。そこで、本論文では、自然言語処理を利用して、複雑化する通信販売業界における、正確な名寄せ・住所寄せを支援することを目的に、名寄せ知識の構築方法を提案する。

2. 通信販売における課題と対処

通信販売業界についての関連研究を外觀する。受注前の活動に関しては、Simester [Simester2006]の通信販売カタログの送付先の最適化モデル、君島[Kimijima2010]の通信販売における広告投入と広告効果タイムラグを考慮した広告効果モデル、[Conlin2007]の通販受注と気温との関係を分析、松田ら[Matsuda2001]の通信販売業界における受注予測モデルなど、多くの研究が存在する。

しかし、受注後の活動に関しては、高橋ら [Takahashi 2012]のランダムフォレストを用いた、地域性による顧客特徴の分析のみで債権管理や貸倒等、未払い顧客について研究されているものは、ほとんど存在しない。

そのため、現状では申込み内容を熟練した作業員が目視で

連絡先: 東 弘樹, HAZS(株), hirokiazuma2002@hazs.biz

名寄せを行い、未払い顧客を割り出すという属人的な対応が行われている。

3. 不正購買の具体的な手口

後払いの多くは、健康食品や化粧品で利用されている。そこで、化粧品の通信販売の受注・支払状況のデータを基に未払い顧客の分析を行った。その結果、未払い顧客は氏名や住所などの個人情報を曖昧に記述することで、個人を確定されないように対応している事例が数多く見受けられた。抽出できた事例を図1に示す。

1) 様方を利用するケース

氏名	受注日	〒+住所番号	住所表記	金額	商品
斎藤恵子	1/21	30300411541	佐藤市田中町丙 1541 高橋様方	32,118	C
松本和子	1/21	30300411541	佐藤市田中町丙 1541 高橋様方	32,118	C
斎藤恵子	2/20	30300411541	佐藤市田中町丙 1541 高橋様方	32,118	C
松本和子	2/20	30300411541	佐藤市田中町丙 1541 高橋様方	32,118	C
斎藤恵子	3/21	30300411541	佐藤市田中町丙 1541 高橋様方	32,118	C

2) 建物の名前自体を揺らすケース

氏名	受注日	〒+住所番号	住所表記	金額	商品
和田まり	3/14	9550842213182	伊藤市島 2-13-18 コーポ吉田 2号	8,820	D
和田正美	3/26	9550842213182	伊藤市島 2-13-18 吉田コーポ 2号	7,770	E

3) 建物の表記を消し、半角全角を利用する複合ケース

氏名	受注日	〒+住所番号	住所表記	金額	商品
松田まい	2/22	886085369102	大西市門脇青葉西 69シテイコートF102	6,500	F
松田由美	7/20	886085369102	大西市門脇青葉西 69F102	17,745	G

4) 「字(あざ)」や「ノ」等、住所の特性を利用するケース

氏名	受注日	〒+住所番号	住所表記	金額	商品
中村夏美	4/16	260156127	鈴木市道下ノ町 127	7,820	A
小林とも	4/18	260156127	鈴木市道下町 127	8,820	B
加藤ふみ	4/18	260156127	鈴木市道字下町 127	8,820	B

図1 未払い顧客の個人情報記載事例

名前を揺らすケースとして、図 1-1) 住所地に旧姓や姉妹の名前、子供の名前+様方や実家+様方のように、様方をつける事で複数の申込を行うケースが多い。都心部では、マンションの集合ポストに表札を出さない事が多く、別人(他人・架空)名義でも宅配会社は配達する。この事から、全くの別人の名前を利用するケースも存在する。その他、ひらがなやカタカナ、英字を使い申込みを行うケースを確認している。

住所地を揺らすケースとして、図 1-2) マンション名の表記の構成を前後させることがある。また、ひらがな・カナで記載したり、外来字の表記を揺らすケースもある。(例レジデンス, デジレンス, レジデンス) 図 1-3) マンション名を記載しないケース, 略字で MS (マンション) と記載する。その他, 号室・号・室等をつけたりつけなかったりする。図 1-4) 字(あざ)をつけたり「ノ」をつけたり, 存在しない東棟・西棟などをつける, 旧住所名にする, 北海道や京都に見られる三条のような地名を数値に置き換えたり, 「上がる」「下がる」の表記で揺らしたり, -(ハイフン)を取ったり, 一(長音)に変えたり, 位置を変えたり, 101 号室を「1F右」のような表現をしたり, そもそもマンションでは無い建物をマンションのように見せかけたりして申込みが行われていた。今回対処方法の手順を用いて検出した結果, 6,700 万円の約 100 万円 (1.49%) の不正購入が存在した。

4. 集計方法

4.1 データの概要

主に基礎化粧品を取り扱う通販会社 T 社の協力の基, 2012 年度の後払いデータ約 30 万件 (売上規模約 18 億) について分析した。通販会社 T 社は, 2012 年度年間売上約 30 億円程度の中堅の通販会社である。初回購入は, 代引き及びカードでの購入に限定している為, 初回から後払いを承諾している企業に比べ, 債権内容は, 比較的良質債権が多い企業である。

顧客 DATA は, 顧客番号, 受注番号, 顧客氏名, カナ, 郵便番号, 住所 1・2・3, 電話番号, 受注日, 入金日, 商品名, 商品価格を抽出した。

4.2 集計手順

不正受注が発生する場合の多くは, 架空の氏名を使い, 住所地の表記を揺らして, 名寄せを困難にするため, 文字情報での意図的な操作が多い。そのため, 数値情報のみを使い住所地を名寄せした上で, 未払い顧客の名前を割り出した。未払い顧客の定義として住所地が同じで名前の相違し, 且つ未払いの状態であるものとした。

不正受注が発生する場合の多くは, 架空の氏名を使い, 住所地の表記を揺らして, 名寄せを困難にするため, 文字情報での意図的な操作が多い。そのため, 数値情報のみを使い住所地を名寄せした上で, 未払い顧客の名前を割り出した。

1) 住所 DATA の統合と数値化及び住所寄せ

①住所情報 1・2・3 を結合した上で, 数値情報だけを抽出した後, 郵便番号を結合し新たな住所番号を作成した。

②住所番号を昇順でソートし同じ番号を IF 関数を使い同じ住所地を判明した。

③個人情報に該当するため, ランダムマスキングプログラムにより情報(氏名, 住所地名)を変更したものを分析に利用した。

2) 統計

①マンション名の相違は, 住所番号と住所が相違するものを抽出した後, マンション名が相違する DATA を抽出し統計した。

②名前ひらがなは, 氏名列を昇順で並べ変えた上で, 氏名欄が全てひらがなのものだけ統計した。

③様方は, 住所列のテキストフィルターで, 「様方」を検索し, 統計した。

④最初の一文字がひらがな・カナ・英字は, 氏名列を昇順で並べ変えた上で, 先頭一文字を抽出し統計した。

⑤字は, 住所列のテキストフィルターで, 「字」を検索し, 統計した。

5. 検出内容

4.2 の手順で抽出した DATA について, 支払済顧客と未払い顧客の割合を計算し, それぞれの割合を算出した。その一例を表 1 に示す。

表 1 支払済顧客と未払い顧客の差異

	金額(百万円)			記述率		倍率
	Total	支払済	未払い	支払済	未払い	
平均値	1,771.68	1,687.66	84.015			
マンション名	2.102	1.934	0.168	0.1146%	0.2000%	1.74
名前ひらがな	0.446	0.412	0.034	0.0244%	0.0405%	1.66
様方	11.398	10.635	0.763	0.6302%	0.9082%	1.44
住所のひらがな・カナ・英字	2.991	2.804	0.187	0.1661%	0.2226%	1.34
字, ノ等の使い	62.921	59.821	3.1	3.5446%	3.6898%	1.04

その結果, 「マンション名」の記載が無い場合は, 未払い率が 1.74 倍になるなど, 支払済顧客と未払い顧客では記載方法の差異があることが判明した。

また, 悪意を持つ注文にも流行のようなものが存在し, 抽出した名寄せを困難とする記載方法は, 特定パターンが特定の時期に集中している場合が多いこともわかった。

6. おわりに

抽出した記載方法の差異は僅かであるが, 悪意を持つ注文はこれらの記載方法を組み合わせていることが多い。そのため, これらの記載方法を抽出することでも悪意を持つ注文を抽出できる。実際, この方法で未払い顧客の 10% 以上を抽出できた。未払い顧客には悪意のない注文も存在するため, 悪意を持つ注文の抽出率は, 更に高いと思われる。

今後は, 統計手法を用いて悪意を持つ注文に存在する記載方法の特徴を分析し, 精度の向上を目指したい。

参考文献

- [(社)日本通信販売協会 2012] 社団法人 日本通信販売協会, 第 19 回全国通信販売利用実態調査, 売上高調査 2011, 2012. 5
- [Simester2006] Simester, D. I., Sun, P. and Tsitsiklis J. N.: Dynamic Catalog Mailing Policies, Management Science, 52(5), 683-696(2006)
- [Kimijima2010] Kimijima M.: A Study on Measuring Input-Output Process on Order-Getting Costs for Direct Marketing, Yokohama National University Departmental Bulletin Paper, 16(1), 21-39(2010) (Japanese)
- [Conlin2007] Conlin M., O'Donoghue T. and Vogelsang T. J.: Projection Bias in Catalog Orders, American Economic Review, 9(4), 1217-1249(2007)
- [Matsuda2001] Matsuda Y., and Ebihara J.: Forecasting Model in the Mail-Order Industry, UNISYS Technology Review, 71, 52-68, Nov(2001) (Japanese)
- [Takahashi 2012] Masakazu Takahashi, Hiroaki Azuma, Masanori Ikeda, and Kazuhiko Tsuda: Building Knowledge for Characterization of the BadDebt Customers in the Mail Order Industry with Random Forest, Advances in Knowledge-Based and Intelligent Information and Engineering Systems, Frontiers in Artificial Intelligence and Applications, (Vol.243, pp.867-847,) IOS Press, 2012.