

ОПТИМАЛНО ПЕРСОНАЛИЗИРАНЕ НА ДИРЕКТНИ МАРКЕТИНГОВИ ИНТЕРВЕНЦИИ ЧРЕЗ ИНКРЕМЕНТАЛНО МОДЕЛИРАНЕ

Доц. д-р Тодор Кръстевич, СА „Д. А. Ценов” – Свищов
Доц. д-р Маруся Смокова, СА „Д. А. Ценов” – Свищов

Резюме: В доклада се демонстрират възможностите на иновативна методология за предсказване на индивидуалните реакции на клиенти при алтернативни маркетингови въздействия, наричана инкрементално моделиране. Методологията позволява измерването на „нетния” ефект върху поведението на отделния потребител в резултат на персонифицираното маркетингово въздействие. Тази техника е сравнително нова и слабо позната в областта на аналитичното извличане на знания от данни. Основните ѝ сфери на приложение са свързани не толкова с предсказването на бъдещото поведение на клиента, колкото с оценяване (измерване) на действителния ефект в поведението му, дължащ се на конкретно маркетингово въздействие.

Ключови думи: инкрементално моделиране, аналитично извличане на знания от данни, директен маркетинг.

1. Въведение

Инкременталното моделиране (uplift modeling) е сравнително непопулярен метод за аналитично извличане на знания от големи масиви с данни. Същественото предизвикателство при набирането и подготовката на данни за прилагане на инкрементален модел е необходимостта от рандомизиран експериментален план (т.е. наличието на експериментална и контролна група, което не винаги е възможно при планирането на кампании за директен маркетинг). Целта на метода е, не просто оценяване на вероятността за реакция на конкретна директна маркетингова интервенция (типично за повечето предиктивни модели, основаващи се на надзиравано обучение), но и прогнозиране на инкременталния отговор (ефект) на маркетинговото действие. Под инкрементален отговор в случая разбираме разликите в поведението на клиентите от експерименталната и контролната групи в следствие на въздействието. Именно според посоката и интензивността на инкременталния отговор би следвало да се избират клиенти за таргетиране в процеса на маркетингова кампания. По този начин във фокуса

на внимание попадат хора, които е по-вероятно да купят в следствие на контакт и/или стимул, а не просто потребители, които имат висока вероятност за покупка (независимо от интервенцията).

2. Идея и основни принципи на инкременталното моделиране

Презумпцията при инкременталното моделиране е, че ефектът от една маркетингова кампания се получава като разлика между процента на реагиралите от експерименталната и контролната група. Така е възможно да се „изолира“ нетният ефект от използвания маркетингов инструмент за въздействие и да се постигне по-висока ефективност с по-малко интервениране (респ. по-нисък бюджет). Формално тази идея може да се опише по следния начин:

$$U = P(\text{покупка} \mid \text{въздействие}) - P(\text{покупка} \mid \text{без въздействие}),$$

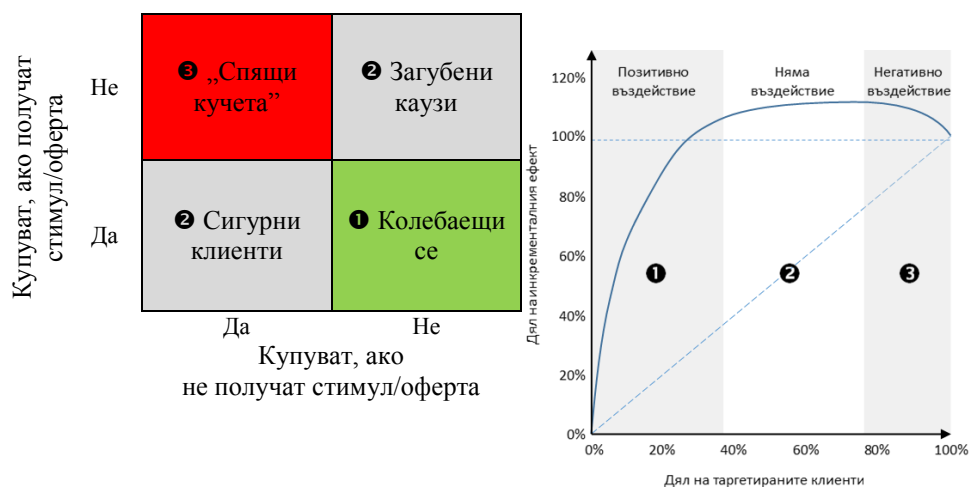
като с U бележим инкременталния ефект, а с P – вероятността за покупка.

При традиционния подход за таргетиране с помощта на „конвенционални“ предиктивни модели се оказва въздействие само върху тези потенциални клиенти, чиито профили са оценени с висока вероятност за реакция (независимо, дали са „активно“ таргетирани в миналото потребители). Част от потенциалните клиенти обаче, върху които не се въздейства „активно“ в миналото, също осъществяват покупки, т.е. вероятността за възникване на позитивна реакция без въздействие е ненулева. От тук, основното предизвикателство е да се идентифицират тези потенциални клиенти, които имат най-високо нарастване на вероятността за покупка в следствие на въздействието (т.е. нар. „колебаещи се“) и респ. да се „филтрират“ и елиминират от списъците за въздействие онези, чиято реакция в следствие на интервенцията е негативна (т.нар. „спящи кучета“). „Колебаещите се“ са потенциални клиенти, които не биха купили, ако с тях не бъде осъществен контакт (респ., ако не им се въздейства). Таргетирането на клиенти, попадащи в категориите „сигурни“ или „загубени каузи“, не води до инкрементален ефект и само увеличава разходите за кампанията. От тук, идеята на инкременталното моделиране може да бъде илюстрирана по следния начин (Вж. Фиг. 1).

3. Моделиране на хетерогенността на потребителските реакции

Потребителите не реагират по един и същи начин на конкретни маркетингови стимули. Поради присъщата хетерогенност на съвременните пазари, маркетинговите интервенции биха могли да предиз-

викват желан и очакван положителен ефект при голяма част от целевия пазар, докато при конкретни подгрупи (сегменти), отличаващи се със специфични характеристики, те да са неефективни, дори вредни (нежелани) за предложителя. По аналогия, предлагането на нов продукт или услуга би могло да не предизвика по-добър ефект от досега предлагания продукт или услуга сред масата потребители, но е възможно при определени малки, латентни потребителски сегменти да се наблюдава учудващо позитивни реакции. Поради тези причини, в случаите, когато са налични детайлни клиентски бази данни, таргетирането на базата на предефинирани сегменти като че ли ще придобива все по-малко значение и ще се измества от оптималното персонализираното таргетиране. Под „оптимално” персонализирано таргетиране се разбира такова маркетингово въздействие, което максимализира вероятността за възникване на желаното поведение (например конкретно изборно решение и/или покупка).



Фиг. 1. Идея на инкременталното моделиране.
(Източник: Portrait Software©)

От гледна точка на моделиране с цел предсказване на изборната вероятност върху клиентски бази данни, тук има сериозно предизвикателство. Докато реакцията на потребителите на минали въздействие или при минали обстоятелства може да се наблюдава (и следователно, да се моделира), то оптималната опция за маркетингово въздействие на практика е неизвестна величина. В литературата има сполучливи опити за систематизиране и проверка предиктивната точност на раз-

лични методи за машинно обучение, третиращи тази проблематика¹. В настоящия доклад се демонстрира практическото приложение на един от вариантите за реализация на метода, а именно – техниката на класификационните дървета. Целта на тази техника е измерването на „нетния” ефект върху поведението на отделния потребител в резултат на персонифицираното маркетингово въздействие (например чрез кампания за директен маркетинг). Тази техника е експериментална и сравнително слабо позната в областта на аналитичното извличане на знания от данни. Основните ѝ сфери на приложение са свързани не толкова с предсказването на бъдещото поведение на клиента, колкото с оценяване (измерване) на действителната промяна в поведението му, дължащ се на конкретно маркетингово въздействие. Техниката на класификационните дървета спада към групата методи и алгоритми за надзиравано машинно обучение (supervised learning). Целта при нея е да се предскаже стойността на интересуваша ни наблюдаваема зависима променлива (номинална или най-често дихотомна) на базата на набор от други, наблюдаваеми независими променливи (предиктори). Значенията на предикторите е възможно да бъдат както наблюдаваеми стойности (например определени признаци на клиентите като пол, възраст, местоживееене, навици, собственост и т.н.), така и поддаващи се на корекция от вземащия решенията (например начинът на контакт с клиента – с имейл, по телефона, пощенска пратка или посещение на място). В подобни случаи целта на предиктивното моделиране е не само да прогнозира значението на зависимата променлива колкото се може по-точно, но и да се избере оптималното решение за въздействие върху всеки клиент, на базата на неговите индивидуални характеристики² (т.е. да се намери решение, което би максимализирало вероятността за постигане на желанния резултат).

4. Данни, резултати и интерпретация

За демонстрация на възможностите на инкременталното моделиране бе използвана клиентска база данни, съдържаща 74449 записа (потребители, притежаващи клиентска карта за пазаруване в търговска верига)³. При издаването на картата са били изисквани някои ти-

¹ Вж. Guelman, L., Guillén, M., Pérez-Marín, A. M., Optimal personalized treatment rules for marketing interventions: A review of methods, a new proposal, and an insurance case study. Working paper 06, UB Riskcenter, Barcelona, 2014, p. 1

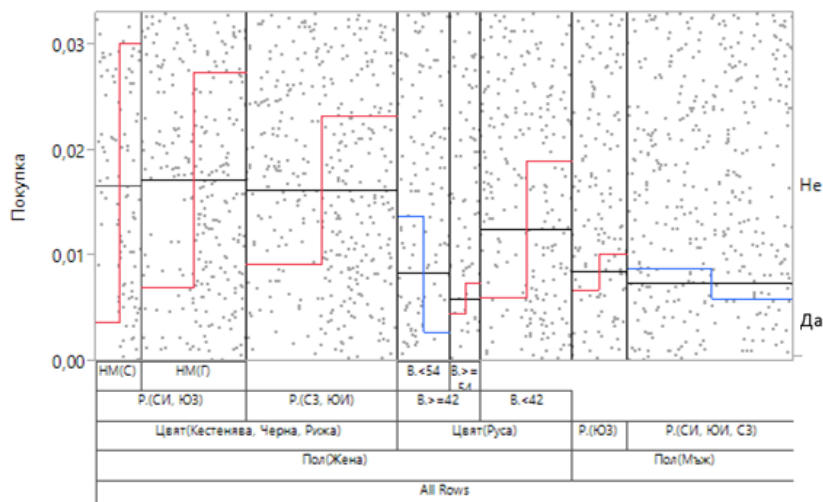
² Вж. Žliobaitė, I., Pechenizkiy, M., Learning with actionable attributes: Attention – boundary cases! ICDMW '10 Proceedings of the 2010 IEEE International Conference on Data Mining Workshopsp 2010, p. 1021.

³ Данните са достъпни на <http://bit.ly/2cMePW3>

пични социодемографски признаци от клиентите като: пол (мъж, жена), възраст (години), регион по местоживееене (северозападен, югозападен, североизточен и югоизточен), вид на населеното място (град, село), цвят на косата (руса, кестенява, черна, рижа). На 38088 случайно подбрани клиенти е била изпратена промоционална рекламна брошура, предлагаща нова марка боя за коса. Тези потребители представляват експериментална група, а останалите – респ. контролна група. В период от три месеца след получаването ѝ са били регистрирани пробните продажби (поне една покупка от новата марка за периода) сред цялата клиентела. За всеки клиент от базата е направен запис, дали е купил поне веднъж през периода (да/не) и дали му е била изпратена рекламна брошура (да/не). Чрез съставянето на обикновена кростаблица лесно може да се установи, че броят на реагиралите (т.е. осъществили поне една покупка през наблюдавания период) е само 890 (1,2%). От тях, от експерименталната група това са 615 (1,66% от получените стимула), а от контролната – съответно 275 (0,74% от получените стимула). Интересно е да се отбележи, че 49,58% от представителите на контролната група имат положителна реакция (купуват), въпреки че не са били обект на стимулиране (т.е. би следвало да се интерпретират като „сигурни“ клиенти). Разходите, направени за тяхното активиране, са били на практика излишни. Сходен извод може да се направи и за 30,90% от получените брошурата, които въпреки това не са реагирали позитивно (отново ненужно изразходване на средства). От тук, логично възниква въпросът, какъв е профилът на тези клиенти, които реагират положително на рекламния стимул и респ. на тези, които биха реагирали негативно при стимулиране?

Отговорът на този въпрос се търси с помощта на инкрементален модел, позволяващ идентифициране на четирите категории клиенти на индивидуално равнище – т.е. да се изчисли инкременталната стойност и посоката на въздействие на всеки потребител от базата данни. Разбира се, най-висок интерес за последващо таргетиране ще представлява сегментът на „колебаещите се“. За целта е приложен директен подход за съставяне на инкрементален модел с помощта на дърво на решенията, при което са изчислени индивидуалните инкрементални ефекти на всеки отделен респондент. Целева променлива в модела е наблюдаваното поведение при купуване (дихотомна променлива със значения купил/некупил по време на наблюдавания период). Като влияеща променлива, зависеща от решението на мениджърите, е използвана дихотомната променлива, показваща на кого е изпратена рекламната брошура.

Най-общ аналитичен обзор на резултатите е възможно да бъде направен с помощта на йерархична диаграма (Вж. Фиг. 2.). От диаграмата, както и от стойностите от Таблица 1, е видно, че най-висок нетен положителен инкрементален ефект (0,026) е идентифициран при групата на жените с цвят на косата, различен от рус, живеещи в села, локализиращи в североизточен и югоизточен регион на страната, следвани от жени със същия профил, но живеещи в градовете (вж. червената начупена линия, както и стойностите, на базата на които тя е представена графично от Таблица 2). Интересна особеност обаче се наблюдава при клиентите „руси жени, между 42- и 54-годишна възраст, както и при мъжете (с изключение на тези, локализиращи постоянно в югозападния регион). При тези две клиентски групи е налице негативен инкрементален ефект – т.е. в следствие на това, че са били подложени на стимул, те реагират негативно (пропорцията на некупуващите нараства). Това ясно се вижда от отрицателните знаци в колона „нетен ефект” от Таблица 1, както и от диаграмата на Фиг. 2 (сините линии). Относителната важност на предикторите при предсказване на инкременталните стойности са представени на Таблица 2.



Фиг. 2. Йерархична диаграма на инкременталния модел.






Ясно се вижда, че най-важната за предсказване на инкременталната стойност променлива е полът на клиента, следван от възрастта. Цветът на косата, регионът и населеното място са с повече от два пъти по-ниско предиктивното значение от първите две променливи. На базата на изчислените индивидуални инкрементални ефекти е възможно да се извърши сортиране на цялата база данни в низходящ ред. След

това, съобразно ограниченията на разполагаеми бюджет, могат да се подберат за последващо въздействие само тези клиенти (респ. сегменти), които имат най-висока положителна инкрементална оценка (т.е. тези, които реагират най-интензивно на получаването на стимула). Независимо от бюджетните ограничения, разумно би било да се изключат от последващи списъци за таргетиране онези лица, които имат близки до нулата или отрицателни инкрементални оценки.

Таблица 1. Групи клиенти с еднаква инкрементална оценка

Брой НЕ	Брой ДА	Дял НЕ	Дял ДА	Нетен ефект	Описание на сегмента
1620	1562	0,0037	0,03009	0,02639	Пол(Жена)&Цвят(Кестенява, Черна, Рижа)&Р.(СИ, ЮЗ)&НМ(С)
3718	3760	0,00699	0,02739	0,0204	Пол(Жена)&Цвят(Кестенява, Черна, Рижа)&Р.(СИ, ЮЗ)&НМ(Г)
5405	5414	0,00925	0,02327	0,01402	Пол(Жена)&Цвят(Кестенява, Черна, Рижа)&Р.(СЗ, ЮИ)
1899	1805	0,01369	0,00277	-0,0109	Пол(Жена)&Цвят(Руса)&В.>=42&В.<54
1118	1081	0,00447	0,0074	0,00293	Пол(Жена)&Цвят(Руса)&В.>=42&В.>=54
3297	3275	0,00607	0,01893	0,01287	Пол(Жена)&Цвят(Руса)&В.<42
1913	2052	0,0068	0,01023	0,00344	Пол(Мъж)&Р.(ЮЗ)
5993	5887	0,00884	0,00595	-0,0029	Пол(Мъж)&Р.(СИ, ЮИ, СЗ)

Таблица 2. Относителна важност (принос) на предикторите за постигане на инкрементален ефект

Фактор	Брой деления	Принос	Принос	Отн. дял
Пол	1	30,9216098		0,3596
Възраст	2	29,6518723		0,3448
Цвят коса	1	13,6520089		0,1588
Регион	2	9,70729626		0,1129
Населено място	1	2,06361444		0,0240

Заклучение

Демонстрираната методология може да намира приложение не само при планиране на традиционни кампании за директен маркетинг, но и при директното офериране на сходни продукти от по-висок ценови клас (up-selling), при предлагането на допълващи продукти, подходящи за комбиниране и/или просто едновременно купуване (cross-selling), както и при планирането на кампании за задържане на клиенти. Подобни техники са типични преди всичко при управлението на взаимоотношенията с клиенти на телекомуникационните пазари, в търговията на дребно и/или при маркетинга на финансови услуги. Интересно, но все още неизследвано приложно поле, е прилагането на тази техника и при управлението на взаимоотношенията със студенти във висши училища, както и на собствения персонал в големи компании.