

TRIỂN KHAI ĐIỀU TRA THỐNG KÊ QUA MẠNG INTERNET Ở THỐNG KÊ HÀ LAN

Jelke Bethlehem^(*)

Giới thiệu: Các cơ quan Thống kê quốc gia (NSI's) Hà Lan thu thập dữ liệu thống kê theo hai hình thức truyền thống, đó là thông tin ghi chép trên giấy và hình thức phỏng vấn điều tra có sự trợ giúp của máy tính (CAI). Thu thập dữ liệu có sự trợ giúp của máy tính có ba điểm thuận lợi quan trọng: (1) Đơn giản hoá công việc của người phỏng vấn, do không phải quan tâm nhiều đến việc chọn lựa lộ trình của bảng hỏi, (2) Nâng cao chất lượng dữ liệu thu thập được, do có thể kiểm tra được người trả lời và chỉnh sửa được trong quá trình phỏng vấn, và (3) Giảm đáng kể thời gian cần thiết để xử lý dữ liệu điều tra và như vậy rút ngắn thời gian công bố kết quả điều tra.

Cho đến nay Internet đã phát triển nhanh hơn một thập kỷ và có thể sử dụng nó để thu thập dữ liệu thống kê, đó là điều tra qua mạng. Theo đó bảng hỏi được thiết kế như một trang web mà đối tượng điều tra có thể truy cập vào được. Một dạng khác là điều tra bằng thư điện tử, với việc yêu cầu đối tượng điều tra điền thông tin cần thiết và gửi lại như một dạng thư điện tử hoặc gửi file đính kèm với thư điện tử.

Thoạt đầu, điều tra qua mạng có vẻ có một số lợi thế hấp dẫn. Thứ nhất, đây là phương thức đơn giản để truy cập đến một nhóm lớn đối tượng điều tra tiềm tàng. Thứ hai, không cần đến điều tra viên và không tốn chi phí bưu điện. Thứ ba, có thể tiến

hành điều tra rất nhanh. Khi đã có bảng hỏi không phải chờ đợi lâu mà có thể triển khai ngay điều tra ở địa bàn. Thứ tư, đặc biệt điều tra qua mạng cho ta những khả năng mới và hấp dẫn như sử dụng các hình thức đa phương tiện như âm thanh, hình ảnh, phim hoạt hình và phim ảnh. Tuy nhiên, điều tra qua mạng cũng có những vấn đề về phương pháp luận. Việc sử dụng internet để chọn đối tượng điều tra và sử dụng web làm công cụ đo lường cũng có những vấn đề cần phải bàn bạc và cân nhắc. Nếu như không chú tâm một cách nghiêm túc các bài toán này, kết quả chất lượng dữ liệu điều tra qua mạng sẽ thấp, không đưa ra kết luận phù hợp với tổng thể mục tiêu của điều tra đề ra.

2. Các vấn đề về phương pháp luận

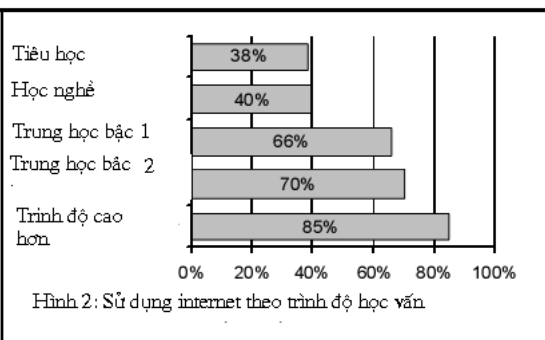
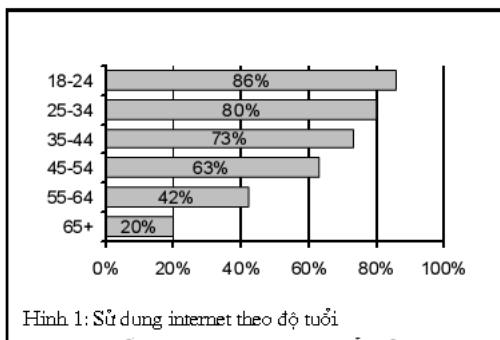
Trong bài viết của hai ông Horvitz và Thompson (năm 1952) chỉ ra là ta có thể tính được các ước lượng không chệch của các đặc trưng tổng thể nếu như tất cả các phần tử trong tổng thể có xác suất chọn lựa khác không (non-zero), các nhà nghiên cứu cũng biết rõ các xác suất này. Hơn nữa, chỉ trong những điều kiện như vậy thì mới có thể tính được độ chính xác của các ước lượng. Nhiều cuộc điều tra qua mạng là những cuộc điều tra tự chọn đối tượng điều tra. Ta không biết xác suất về khả năng mời được đối tượng điều tra tham gia như vậy, cũng không biết xác suất chấp nhận một đối

^(*) Vụ Phương pháp và Thông tin - Cục Thống kê Trung ương Hà Lan

tượng là bao nhiêu. Do vậy, ta không thể tính được ước lượng chệch của các đặc trưng của tổng thể.

Với việc sử dụng giàn mẫu, hầu như chúng ta luôn chọn được một mẫu xác suất. Tuy nhiên, sẽ nảy sinh vấn đề nếu khi tổng thể đó được đại diện bằng một dàn mẫu khác so với tổng thể mục tiêu điều tra. Khi các phần tử của tổng thể mục tiêu không xuất hiện trong dàn mẫu thì diện điều tra không sẽ không đầy đủ. Đối với các cuộc điều tra qua mạng đây là vấn đề nghiêm trọng. Thông thường, tổng thể mục tiêu của các cuộc điều tra của các Cơ quan Thống kê quốc gia lớn hơn nhiều so với những người nối mạng internet. Do đó, một bộ phận đáng kể sẽ không bao giờ được chọn vào nếu như sử dụng internet làm dàn mẫu. Thực tế, nhiều người dân vẫn chưa thể truy cập vào internet. Như được giới thiệu trong hình dưới. Người càng có tuổi thì tỷ lệ sử

dụng internet càng ít và tỷ lệ đó tăng lên cùng với với trình độ học vấn (nguồn Thống kê Hà Lan 2003). Vấn đề đối tượng không trả lời câu hỏi cũng là vấn đề hệ trọng khi điều tra qua mạng. Đó là hiện tượng mà các đối tượng điều tra (con người, hộ gia đình, công ty) trong dàn được chọn không cung cấp thông tin cho cơ quan điều tra, hoặc thông tin cung cấp không sử dụng được. Bảng hỏi điều tra qua mạng là những câu hỏi mà đối tượng điều tra tự thực hiện, nên tỷ lệ trả lời có thể không cao, cũng có thể một phần do vấn đề kỹ thuật như tốc độ modem chậm, kết nối mạng không tốt, chi phí kết nối mạng cao, hướng dẫn không rõ ràng gây khó cho đối tượng điều tra. Để đảm bảo cho thông tin điều tra ở mức có thể chấp nhận được, phải tránh được những vấn đề như vậy. Điều này đòi hỏi khâu thiết kế công cụ bảng hỏi điều tra qua mạng phải cẩn thận.



3. Điều tra theo phương thức hỗn hợp

Các cuộc điều tra hỗn hợp là có sự kết hợp nhiều phương pháp thu thập dữ liệu trong cùng một cuộc điều tra. Về truyền thống có ba hình thức thu thập dữ liệu: phỏng vấn trực tiếp, phỏng vấn qua điện thoại và điều tra qua bưu điện. Từng loại một có những thuận lợi và bất lợi riêng về chất lượng dữ liệu thu thập và chi phí cho công việc đó. Kết hợp được ba hình thức điều tra này chúng ta có thể có được hình

thức điều tra tốt nhất, đó là chất lượng dữ liệu điều tra tốt và chi phí điều tra thấp. Tỷ lệ cung cấp thông tin điều tra của ba hình thức điều tra này có khác nhau, hình thức phỏng vấn trực tiếp có tỷ lệ cao nhất, tiếp đến là phỏng vấn qua điện thoại và cuối cùng là qua bưu điện. Năm 2001 ông Dillman et al đã chỉ ra là hình thức điều tra hỗn hợp có khả năng nâng tỷ lệ trả lời, đặc biệt là khi các hình thức được tiến hành kế tiếp nhau, ngay sau một thời gian ngắn.

Điều tra hỗn hợp cũng có những vấn đề cần phải bàn là liệu đối tượng điều tra cung cấp thông tin ở dạng điều tra này có khác với câu trả lời ở hình thức điều tra khác không. Trong thực tế chúng ta đã thấy những tác dụng của những hình thức điều tra này. Năm 1997 hai ông De Leeuw và Collins đã chỉ ra rằng đối với những câu hỏi nhạy cảm các cuộc điều tra gửi qua bưu điện thực hiện tốt hơn điều tra trực tiếp và điều tra qua điện thoại, còn điều tra trực tiếp và điều tra bằng điện thoại có số mục không trả lời ít hơn là điều tra gửi qua bưu điện.

Trong thiết kế và thực hiện một cuộc điều tra hỗn hợp, khâu tổ chức thường gặp nhiều khó khăn, phức tạp. Thứ nhất, để tránh được những tác động không tốt của hình thức điều tra đến mức có thể, các dạng bảng hỏi điều tra được dùng trong các hình thức điều tra càng phải tương đồng với nhau ở mức cần thiết. Thứ hai, thông thường ta sử dụng những hệ thống khác nhau đối với các hình thức điều tra khác nhau. Như vậy sẽ không dễ cho khâu thiết kế, thực hiện và duy trì các công cụ thu thập dữ liệu sao cho công việc đúng như vậy đối với các hệ thống khác nhau. Thứ ba, nếu như triển khai các phương pháp khác nhau theo những hệ thống khác nhau có thể sẽ phức tạp trong khâu quản lý. Thứ tư, nếu sử dụng theo các hệ thống khác nhau, dữ liệu có thể được lưu trữ trong những dạng file khác nhau và ở trong những vị trí khác nhau. Khi kết nối những file này lại, file phân tích dữ liệu có thể trở nên nặng nề.

Để tránh được những vấn đề trên, cần có một hệ thống thu thập dữ liệu hài hoà. Hệ thống như vậy không những khuyến khích các phương thức thu thập dữ liệu truyền thống mà cả đối với hình thức thu thập qua internet.

Đã vài năm trước đây, Thống kê Hà Lan đã triển khai ý tưởng của Trung tâm điều khiển xử lý điều tra tự động (với sự trợ giúp của máy tính), được thực hiện bằng một hệ thống phần mềm được gọi là *Blaise*. Để tiến hành một cuộc điều tra, *Blaise* yêu cầu phải có định nghĩa đối với kiểu dữ liệu. Đó là mô tả về các câu hỏi và những quy định yêu cầu trả lời cần phải tuân thủ. *Blaise* lưu trữ loại dữ liệu này ở dạng thuật ngữ máy tính (machine-readable). Một số loại hình thu thập dữ liệu phù hợp với *Blaise* như thu thập dữ liệu truyền thống – các bảng hỏi thiết kế trên giấy.

Phiên bản mới của nhất của *Blaise* cũng được thiết kế dành cho phỏng vấn điều tra qua mạng có sử dụng máy tính (CAWI). Bảng hỏi điều tra qua mạng cũng được xây dựng theo kiểu dữ liệu tương tự. Hai phương pháp phỏng vấn qua mạng được thực hiện theo hình thức này. Phương pháp định hướng theo câu hỏi có thể được sử dụng đối với bảng hỏi dài và phức tạp theo một quy trình nhất định và có kiểm tra. Đối tượng điều tra trong quá trình phỏng vấn luôn giữ nguyên trạng thái trực tuyến khi xem câu hỏi. Phương pháp điều tra thông qua bảng hỏi được sử dụng cho các cuộc điều tra ngắn hạn và đơn giản, nhập dữ liệu trực tiếp vào và không cần theo lộ trình câu hỏi. Trang web gồm tất cả những dạng câu hỏi đơn giản, có thể hoàn thành dạng này theo dạng độc lập.

4. Kinh nghiệm ban đầu và Kế hoạch đặt ra

Điều tra thử nghiệm qua internet của Thống kê Hà Lan đầu tiên được tiến hành vào năm 1998. Điều tra là bước khám phá và kiểm tra xem nội dung thông tin e-mail nào được dùng để thu thập dữ liệu về điều tra xu thế kinh doanh. Đây là cuộc điều tra theo nhóm, các công ty trả lời một số câu

hỏi ngắn về kỳ vọng sản xuất, đơn đặt hàng và vốn. 1 600 công ty được mời tham gia điều tra thông qua thư điện tử. Cuộc thử nghiệm thành công với tỷ lệ trả lời của các công ty tham gia là gần 90%, không gặp khó khăn gì về mặt chuyên môn. Toàn bộ các đối tượng được hỏi đều trả lời tốt. Tuy nhiên, bảng hỏi chỉ dựa theo dạng mã ASCII, lạc hậu, không thân thiện với người sử dụng.

Tiếp sau thành công về điều tra thử nghiệm này, năm 1999 quyết định tiến hành thử nghiệm với *Blaise* IS, phiên bản đầu của Hệ thống *Blaise* có khả năng tạo ra các dạng bảng hỏi của web. Tổng thể mục tiêu là nhóm gần 1 500 công ty xây dựng. Dạng điều tra được áp dụng là thống kê hàng tháng về xây dựng, nội dung chỉ gồm ba câu hỏi về doanh thu và số lượng lao động. Các công ty có thể chọn phương pháp thu thập dữ liệu theo ba hình thức: điều tra qua thư điện tử, điều tra trực tuyến qua mạng và điều tra không trực tuyến qua mạng. Trong số 188 công ty xây dựng sẵn sàng tham gia điều tra thử nghiệm. Rất cuộc có 74% trong đó điền vào bảng hỏi theo dạng điện tử. So với hình thức điều tra sử dụng giấy, tỷ lệ trả lời này có cao hơn chút ít. Trong số các công ty trả lời điều tra, 23% có thư điện tử nhưng không có khả năng thực hiện, 38 % công ty có khả năng thực hiện thư điện tử được chọn vào dạng điều tra không trực tuyến và 39 % nữa được chọn vào dạng điều tra trực tuyến. Một tỷ lệ nhỏ với 8% trả lời theo dạng khác. Cùng với thử nghiệm trước đó, kết luận cuối cùng là phương pháp thu thập dữ liệu này là khả thi. Về mặt thời gian thì thu thập dữ liệu điều tra dạng này có hiệu quả hơn và tiết kiệm được chi phí.

Thành công của hai thử nghiệm này đã dẫn đường cho dự án *E-Quest*. Mục tiêu

của dự án này là thực thi các dạng điều tra qua mạng trong lĩnh vực sản xuất đối với 20 cuộc điều tra ngắn hạn về lĩnh vực kinh doanh theo nhóm hàng. Một mục tiêu khác nữa của dự án là thuyết phục 1/3 trong số 27 000 công ty trong nhóm tham gia vào dạng truyền dữ liệu điện tử này. Sau nhiều nỗ lực, khoảng 7 000 công ty đồng ý tham gia và cung cấp địa chỉ e-mail. Thư điện tử gửi đến công ty để hỏi xem họ có lướt web không. Hàng tháng, hàng quý thư điện tử được gửi về những công ty (chỉ đến nhưng nơi có địa chỉ thư điện tử) nhắc họ điền vào bảng hỏi và gửi lại. Những người lướt web cũng được gửi một thư điện tử có đính kèm một dạng HTML (ngôn ngữ siêu văn bản: Hypertext Markup Language). Dữ liệu loại này được sử dụng ở dạng mật mã để dữ bí mật.

Tháng 10 năm 2004 đã bắt đầu triển khai thí điểm để tìm hiểu xem liệu có thể được dùng dạng bảng hỏi điện tử này để thu thập số liệu thống kê sản xuất hàng năm không. Hai phương pháp sẽ được thử nghiệm đó là: thứ nhất, báo cáo dữ liệu điện tử (EDR). Hệ thống này giúp đối tượng điều tra quản lý được các chương trình phỏng vấn *Blaise* trong máy tính riêng của mình. Sau khi phần mềm EDR được nạp vào trong máy tính, đối tượng điều tra mới được gửi đến một thư điện tử đến. Đối tượng điều tra này được tự động gia nhập vào danh sách EDR (báo cáo dữ liệu điện tử), và chỉ đơn giản nhấp chuột là chương trình bắt đầu phỏng vấn. Sau khi hoàn thành phỏng vấn ngoại vi, dữ liệu nhập vào được tự động chuyển thành dạng mã và gửi về Cơ quan Thống kê Hà Lan. Phương pháp thứ hai là điều tra trực tuyến qua mạng, bảng hỏi điều tra cũng là cùng dạng dữ liệu *Blaise*. Đối tượng điều tra được phép ngắt quãng cuộc phỏng vấn và lưu dữ liệu đã nhập vào, rồi tiếp tục vào một thời điểm

nào đó. Cuộc điều tra thử nghiệm được triển khai với một mẫu gồm 100 công ty thuộc bốn ngành khác nhau như vận tải hàng hoá bằng đường bộ, công nghệ thông tin, kế toán và học nghề.

Cục Thống kê Hà Lan có nhiều kinh nghiệm nhất ở lĩnh vực điều tra doanh nghiệp qua mạng, bởi vì điều kiện máy tính và internet xâm nhập vào khu vực này là lớn hơn nhiều so với các hộ gia đình. Hiện nay có nhiều hộ gia đình hơn (hơn 60%) tiếp xúc đến các lĩnh vực này. Thật là hấp dẫn nếu chúng ta xem xét các đối tượng này sử dụng internet vào điều tra hộ gia đình. Bước đầu tiên của hướng này là dự án *CBSquest*). Mục tiêu của dự án này là để xem xét xem liệu điều tra có khả năng sử dụng cho các nhóm dân cư mà khó tiếp cận điều tra trực tiếp (họ hầu như vắng nhà) và điều tra qua điện thoại (không có trong danh mục điện thoại cố định và di động). Mục đích nữa là khám phá và tìm khả năng làm giảm chi phí bằng phương pháp thu thập này. Đối

với thử nghiệm đầu tiên, sử dụng câu hỏi của cuộc điều tra tổng hợp về điều kiện sống (Pols). Để kiểm tra xem liệu toàn bộ phương pháp có đúng chuyên môn không, cuộc kiểm tra này đã được tiến hành trên một mẫu 40 lao động làm việc tại Thống kê Hà Lan. Cuộc kiểm tra này đã thành công. Bước tiếp theo là để triển khai điều tra cảm tình của người tiêu dùng theo loại hình điện tử. Bảng hỏi này sẽ dành cho đối tượng điều tra “thực”. Kiểm định này cũng xem xét phương thức tác dụng khả dĩ của nó■

Trần Mạnh Hùng dịch

Internet Survey Developments At
Statistics Netherlands

Jelke Bethlehem

Methods and Informatics Department

Prinses Beatrixlaan 428

2273 XJ Voorburg, The Netherlands

jbtm@cbs.nl