ARTICLE IN PRESS
DECSUP-12926; No of Pages 9
Decision Support Systems xxx (2018) xxx–xxx
Contents lists available at ScienceDirect
Decision Support Systems
journal homepage: www.elsevier.com/locate/dss

Unsupervised tip-mining from customer reviews

Di Zhu[a], Theodoros Lappas[a,*], Juheng Zhang[b]

[a] Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA
[b] University of Massachusetts Lowell, 220 Pawtucket St, Lowell, MA 01854, USA

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form 30 December 2017
Accepted 30 January 2018
Available online xxxx

Keywords:
Online reputation
Reviews
Unsupervised learning
Tips

ABSTRACT

In recent years, large review-hosting platforms have extended their functionality to allow their users to submit tips: short pieces of text that deliver valuable insight on a specific aspect of the reviewed business. These tips are meant to serve as a concise source of information that complements the often overwhelming number of customer reviews. Recent work has tackled the problem of automatically generating tips by mining review text. The motivation for this effort is to obtain tips for businesses or business aspects that have been overlooked by users. Another motivating factor is the quality of the user-submitted tips, which often provide trivial or redundant information. Existing tip-mining methods are limited by a reliance on training data, which is unlikely to be available and is also very costly to create for different domains. In this work, we present TipSelector, a completely unsupervised algorithm that delivers high quality-tips without the need for annotated training data. We verify the efficacy of TipSelector via an evaluation that includes real data from the hospitality industry and comparisons with the state-of-the-art. A secondary contribution of our work is a method for automatically evaluating tip-mining algorithms without humans in the loop. As we demonstrate in our experiments, this method can be used to enable large-scale evaluations and complement the user studies that are typically used for this purpose.

1. Introduction

Customer reviews have been established as an integral part of modern e-commerce. Recent studies reveal that 91% of consumers regularly or occasionally read online reviews, while 84% of people trust online reviews as much as personal recommendations [5]. Previous work has repeatedly verified the strong influence of reviews on the sales and success of products across domains [9, 16, 38, 57, 58]. Despite the established benefits, the volume and unstructured format of customer reviews have introduced challenges for consumers who are trying to collect information on competitive options prior to making a purchase decision. In recent years, the number of reviews on products and services has been rapidly increasing to the point where a platform can host hundreds or even thousands of reviews on a single item or business. For instance, the *Flamingo Las Vegas Hotel & Casino* had around 10,000 reviews on TripAdvisor.com in 2011, and the number was over 30,000 reviews as of 2017. Given that it is clearly impossible for a user to process such overwhelming volumes of information, both platforms and researchers have explored various methods for addressing this challenge. Examples include methods for ranking [2, 33, 39, 52], selection [35, 36, 44, 51], and summarization [13, 27, 48, 60] of customer reviews.

In an effort to further help their users make a purchase decision, leading online platforms, such as TripAdvisor and Yelp, have recently extended their functionality by hosting user-submitted tips: short text snippets that deliver useful information on a specific aspect of the reviewed entity. Consider the following examples of hotel tips:

- *The street noise is an issue, as the fire station is about a block away (with engines coming and going).*
- *Whole Foods is in the neighborhood, along with the naval observatory, where the vice president resides.*

Even if such information is hidden within the large volumes of available reviews or somewhere on the hotel's website, it is unlikely that the user will be able to retrieve them. In fact, studies show that 90% of consumers read less than 10 reviews before forming an opinion about a business, while 68% form an opinion by reading just 1–6 reviews [5]. User-submitted tips serve as a concise way to encode such valuable insight, thus facilitating the user's information-processing task.

Despite their benefits, user-submitted tips may overlook key aspects of the reviewed entity and can also suffer from issues that

* Corresponding author.
E-mail addresses: dzhu1@stevens.edu (D. Zhu), tlappas@stevens.edu (T. Lappas), Juheng_Zhang@uml.edu (J. Zhang).

https://doi.org/10.1016/j.dss.2018.01.011
0167-9236/

are common with user generated content (UGC), such as redundancy and low information quality [1, 15, 25]. In addition, as is the case with online reviews and UGC in general, it is likely that the platfrom will receive numerous tips for a small set of popular businesses, while new or less prominent options will be overlooked. In order to address such issues, recent work has introduced a method for automatically extracting tips from customer reviews [25]. The proposed method starts by extracting meaningful templates from a set of pre-existing tips that are assumed (or manually verified) to be informative. After human annotators filter out generic and irrelevant templates, the final template-set is used to locate tip candidates in customer reviews. Human annotators label each candidate as useful or not useful. The annotated tips serve as a training dataset for a classifier, which can then be used to automatically label new candidates. We describe the process in detail in Section 2.

The pioneering method proposed by Guy et al. [25] is the first of its kind and, as its creators demonstrate, can deliver informative tips. However, its applicability is limited by its supervised nature, its reliance on existing tips, and the need for manual annotations. In this work, we overcome such limitations and extend the literature by making the following contributions:

1. We present TipSelector: a completely unsupervised algorithm for automated tip-mining from customer reviews. As we demonstrate in our experiments, TipSelector delivers a low-redundancy set of highly useful tips, as evaluated by human annotators.
2. We describe a method for evaluating the informativeness of a given set of tips without humans in the loop. Our method can be used to complement studies with human annotators, which are currently the only option for evaluating tip-mining algorithms.

The rest of paper is organized as follows. We review the relevant literature in Section 2. In Section 3, we propose an algorithm that extracts informative tips from customer reviews. In Section 4, we verify the efficacy of the proposed algorithm via an evaluation on real data, which includes multiple experiments and competitive baselines. Finally, we conclude in Section 5 with a discussion on managerial implications and a note on future work.

2. Background and related work

Online reviews have been established as the primary source of information for consumers [6, 14, 23, 29, 56]. Previous work has verified that reviews have a strong effect on customer decisions across industries [9, 16, 38, 57, 58]. Despite the established benefits of online reviews, their ever-increasing number has also given rise to challenges. Fake reviews have emerged as a troubling phenomenon that has motivated a significant body of work [28, 30, 37]. Relevant research has also focused on studying information quality in customer reviews [7, 33, 59].

Another relevant challenge is information overload, an inevitable consequence of the hundreds or even thousands of reviews that are nowadays available on a single entity [17, 31, 46]. A number of studies have verified that information overload leads to an increase in the time required to make a decision [50], a decrease in decision quality [10, 49], and lack of confidence about the decision [12]. Research has focused on both conversational overload and information entropy [31]. Conversational overload is the state when an individual fails to react adequately in view of the fact that too many messages are delivered [54], whereas information entropy represents the condition when incoming messages are not systematically arranged [26].

Previous work has explored three different directions toward addressing information overload. The first is review selection, which focuses on selecting a compact and representative subset of reviews [35, 36, 44, 51]. The second direction is review summarization, which aims to report aggregate statistics of negative and positive opinions about different features of the reviewed entity [13, 27, 48, 60]. The third direction is review ranking, which ranks the available reviews according to various measures, such as user-assigned helpfulness votes [2, 33, 39, 52].

In recent years, review-hosting platforms have explored user-submitted tips an alternative mechanism for addressing information overload. Tips are short pieces of text (typically no more than 1–2 sentences long) that deliver valuable insight on a specific aspect of the reviewed entity. Previous work on text mining has focused on extracting tips from sources such as question-answering websites. A relevant effort is that by Weber et al. [53], who define tips as "short, concrete and self-contained bits of non-obvious advice". The authors proposed a tip-extraction method to answer "how-to" questions. After collecting answers starting with a verb as tip candidates, they conducted a user study to evaluate the quality of each candidate. The annotated tips served as a training set for a classification algorithm, which could then be used to evaluate new candidates. Recently, Guy et al. [25] presented a similarly inspired tip-mining algorithm. We present the pseudocode in Algorithm 1.

Algorithm 1. Baseline tip extractor [25].

Input: sentence-similarity function $g()$, set of businesses $\mathcal{B}$, set of reviews $\mathcal{R}_b, \forall b \in \mathcal{B}$, set of noisy tips $\mathcal{T}$, frequency threshold FT, similarity threshold ST.
Output: Set of tips $\mathcal{T}_b$ for each business $b \in \mathcal{B}$
Manually select a subset of $\mathcal{T}^* \subseteq \mathcal{T}$ of useful tips.
Process $\mathcal{T}^*$ to extract a set of n-gram ($4 \leq n \leq 7$) sequences $\mathcal{W}$ that have at most 1 wildcard and occur at least FT times.
Create a final set of templates $\mathcal{W}^*$ by manually eliminating trivial templates from $\mathcal{W}$.
Process all available reviews to identify the set of sentences $\mathcal{M}$ that match at least one of the templates in $\mathcal{W}$.
Manually annotate a subset $\mathcal{M}^* \subseteq \mathcal{M}$ as useful/not useful.
Train a binary usefulness classifier CL_u on $\mathcal{M}^*$
for $b \in \mathcal{B}$ **do**
    Use CL_u to identify the set $\mathcal{U}_b$ of useful sentences in $\mathcal{R}_b$.
    Let $CL_u(s)$ be the score assigned by CL_u to a sentence s.
    $\mathcal{T}_b = \{s \in \mathcal{U}_b : \nexists s' \in \mathcal{U}_b \text{ with } g(s,s') \geq ST \text{ and } CL_u(s') > CL_u(s)\}$
Return $\{\mathcal{T}_b, \forall b \in \mathcal{B}\}$

The first phase of the algorithm focuses on the generation of tip templates. In order to do so, the authors collected a set of user-submitted tips $\mathcal{T}^*$ from 9847 travel guides on 54 different cities from TripAdvisor. Human annotators then identified a subset $\mathcal{T}^* \subseteq \mathcal{T}$ of useful tips. The second phase identifies n-gram sequences of 4–7 words that frequently appear in $\mathcal{T}^*$. These sequences are converted to a set of templates $\mathcal{W}$ by allowing one of the included words to be a wildcard that could be matched to *any* word. Irrelevant and trivial templates are eliminated by hand to produce a final set $\mathcal{W}^*$ of 150 templates.[1] The set of templates was then applied on a large dataset of over 3.5 million TripAdvisor reviews on various points of interest (POIs), such as tourist attractions across the USA. Out of the full set $\mathcal{M}$ of about 415,000 matching sentences, the authors sampled a subset $\mathcal{M}^*$ of 7500. The sentences in $\mathcal{M}^*$ were again annotated by

[1] http://www.ise.bgu.ac.il/downloadMe/templates.txt.

Please cite this article as: D. Zhu et al., Unsupervised tip-mining from customer reviews, Decision Support Systems (2018), https://doi.org/10.1016/j.dss.2018.01.011