Toward Understanding Causes of Anomaly in Dynamic Restaurant Rating

Lei Li[1], Zhao Lu[2], Danchen Zhang[3], Sanqiang Zhao[3], Ke Zhang[3],
[1]Nanjing University of Science and Technology
[2]East China Normal University
[3]University of Pittsburgh

Abstract
Rating score and text review are the most common features provided in online review systems to gather the opinions shared by users. Product rating distributions usually evolve dynamically over time and potentially accompany with some unusual changes, namely anomalies, which might be caused by product quality change or spamming attacks. In this preliminary study, we analyze the time-series of rating score distributions by using the data collected from Yelp restaurants, and we apply Principal Component Analysis (PCA) to detect anomalous time points. Through manually checking the corresponding review texts, we further investigate the underlying reasons leading to anomalous rating scores. The potential reasons we identified include food/service quality change, user preference, and review spam. Our study is envisioned to help business owners respond timely to unusual feedbacks and manage their business more efficiently.

Keywords: rating score; anomaly detection; PCA; online restaurant review

Citation: Li, L., Lu, Z., Zhang, D., Zhao, S. & Zhang, K. (2017). Toward Understanding Causes of Anomaly in Dynamic Restaurant Rating. In iConference 2017 Proceedings, Vol. 2 (pp. 134-137). https://doi.org/10.9776/17320

Contact:lileiwelldone@gmail.com; zlu@cs.ecnu.edu.cn; daz45@pitt.edu; saz31@pitt.edu; kez11@pitt.edu

1 Introduction

Online review systems, such as Yelp and Amazon, allow users to provide the rating scores associated with text reviews about products and services. On one hand, the ratings and reviews play an important role in shaping the decisions of the potential customers. On the other hand, they provide the business with timely feedbacks from their customers. User behaviors in online review systems usually rely on personal preferences (X. Li & Hitt, 2008; Feng & Qian, 2013), word-of-mouth effects (Cheung, Luo, Sia, & Chen, 2009), and even peer influence via their social connections (Zhang & Pelechrinis, 2014; L. Li, Zhang, Zhou, & Zhang, 2016).

Online ratings are usually represented as a score, typically ranging from 1 to 5. The rating score distribution in the different time periods varies dynamically. Figure 1 shows the distribution of rating scores over time for a restaurant named "Lombardi's Pizza". Different colors represent the percentages of ratings with scores from 1 (blue) to 5 (pink). We can see that the distribution sometimes suffers unusual changes. For example, at the very beginning, i.e., from time point t_1 to t_2, there is a sharp increase in the number of low rating scores. Such unusual changes are defined as anomalies (N. Günnemann, Günnemann, & Faloutsos, 2014).

In terms of anomaly detection, there are many related topics such as detecting spam reviews (Mukherjee, Liu, & Glance, 2012; Lim, Nguyen, Jindal, Liu, & Lauw, 2010) and fake reviews (Mukherjee et al., 2013; Lin et al., 2014). These studies mainly focus on detecting outliers, which are often attributed to random data corruptions (such as the spam or the measurement errors). However, not only outliers, but also anomalies exist in the ratings. Anomalies are the irregular data but follow a specific pattern, e.g., a restaurant receives consistently a lower rating scores due to a temporal decline of its service quality. Previous studies (N. Günnemann et al., 2014; S. Günnemann, Günnemann, & Faloutsos, 2014) have focused on anomaly detection by analyzing the temporal dynamics of rating scores. In particular, they found that rating scores cannot accurately reflect the real quality of products. They extracted the base rating that reflects the regular quality of products, and discovered the time points at which the product's rating shows anomalies (N. Günnemann et al., 2014). However, the hidden reasons underneath the occurrence of these rating anomalies are still not well examined.

In this work, we collect the rating and review data for restaurants in Yelp, and then we apply Principal Component Analysis (PCA) to detect time periods with anomalous rating distributions. We further analyze the content of corresponding reviews to examine various potential causes of anomalous ratings. Our study is envisioned to help business owners respond timely to unusual feedbacks and manage their business more efficiently.

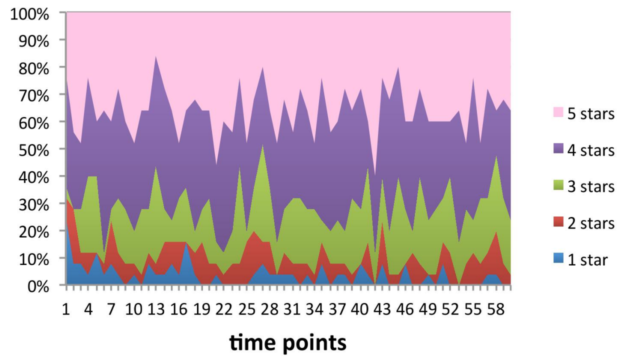

Figure 1: The Rating Distribution of a Restaurant "Lombardi's Pizza" Over Time

2 Research Design

2.1 Dataset

We collect review data from Yelp for restaurants in New York City. In particular, we query the webpage of a restaurant iteratively to get the full list of its historical reviews, and each review has a tuple format *<time, user id, rating score, review text>*. For example, the pizza restaurants shown in Figure 1 in our data has 3,492 reviews starting from 2009/1/10 to 2014/11/8.

2.2 Method

In this section, we introduce how to use PCA to identify abnormal rating time points and how we recognize the causes of these anomalies.

Following the method from previous work (N. Günnemann et al., 2014), we first discretize the whole collection time period into epochs, with epoch size as 2 weeks, and we calculate the percentages of rating scores using all reviews in each epoch. Eventually, we organize the time-series of rating distributions as a matrix X, where each column i represents the rating distribution in time epoch t_i.

We then apply PCA on matrix X. We first calculate the covariance matrix S, that is $S = \frac{1}{r-1}X^T X$. Actually, there is an interesting connection between Singular Value Decomposition (SVD) and PCA. In particular, let the SVD of matrix X be $X = U\Sigma V^T$. Then we calculate the eigenvalue decomposition for S:

$$S = \frac{1}{r-1} V\Sigma^T U^T U\Sigma V^T = V\Lambda V^T \tag{1}$$

where Λ is a diagonal matrix containing the eigenvalues λ_i of S in descending order.

In summary, PCA captures the dominant patterns in the original data by construct a k-dimensional normal subspace X^k in the original t-dimensional space. The remaining $t-k$ dimensions form the anomalous subspace X^a, where we have $X^a = X - X^k$. We select k using Equation 2, such that X^k can explain at least 95% of the variance in the original data. If we need all the components to explain the 95% variance, it means there is no potential anomalous ratings in the data.

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{t} \lambda_i} \geq 95\% \tag{2}$$

For the remaining $t-k$ components, i.e., with $\lambda_i > 0$, we calculate the root mean square of each column in X^a as the anomalous deviation d of the rating distribution in each epoch from the normal status.

3 Results

Figure 2 shows a ramen restaurant's abnormal deviations at different time epochs, and the mean anomaly value of this restaurant is set as the threshold. Time points 2, 5, 9, and 10, circled in the Figure 3, are the identified anomalies time points. Table 1 shows the four rating scores' anomalous values. The time points 2, 5, and 10 show an increase in low-star ratings and a decrease in high-star ratings. Then we read the related reviews and find the following facts: most of the customers complain about waiting time, for example: "The main negatives about Totto ramen is the wait time (about 30 minutes)." They also complain about the small size of the restaurant: "This place was pretty small" and the menu: "the menu selections were limited." We also find another interesting fact: the user's personal preference affects the rating score, as in: "the miso ramen had thin white noodles which I'm less a fan of." Moreover, we find there are users who try to advertise other competitive ramen restaurants: "Not as good as hide chan or ippudo." Meanwhile, time point 9 shows the increase of high rating values because of low prices: "the prices are great." The ratings also increase because of the staff's good performance: "staff is amazing."

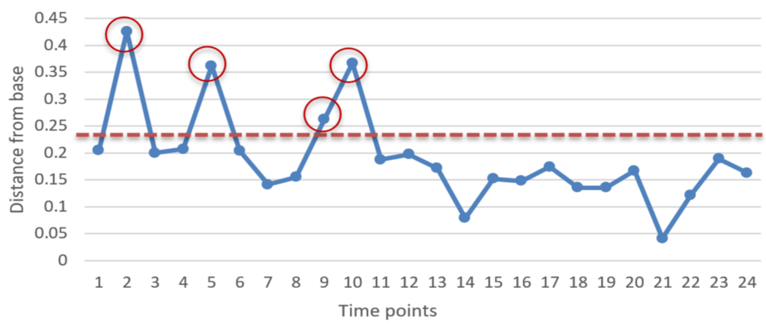

Figure 2: The Ramen Restaurant's Anomalies at Certain Time Points

Based on each anomaly time points' five rating scores anomalous value, the anomalous time points can be divided into two categories: high-to-low anomaly time points, and low-to-high anomaly time points.

For the high-to-low category, we examine the reviews with the low rating scores to identify the negative reasons mentioned by the customers. High price, long waiting time, low food quality, bad taste, too crowd, bad surroundings, inconvenient payment method, and a limited menu are basically all reasons for anomalies related to **the restaurant quality**. From the review text, we can further check the detailed reasons for the anomalous rating: for example, customers might complain that the food has few sauces. **Spam and personal preference** are the other two reasons that can lead to a low-to-high rating anomaly.

For the low-to-high category, some positive reasons overlap with the negative reasons, because some drawbacks of the restaurant have improved over time. We also found the positive reasons that related to **the restaurant quality**, such as great delivery, nice service, and good location, good taste, short waiting time, low price, good surroundings, and high food quality. Also, the detailed positive reasons can be detected. For example, a customer report that they enjoyed the perfectly boiled dumplings. **Personal preferences** are another reason that results in a positive rating anomaly.

4 Discussion and Conclusion

We employ PCA model to identify the anomalies rating time points. Then through reading the reviews written in the recognized anomalies time points, we explore the reasons that lead to the anomalies. Three restaurants' rating and reviews on Yelp are used to discuss the various causes of positive and negative anomalous ratings, including the reasons related to the restaurants quality, personal preference, and spam. Other time points that are not anomalous may have the same causes of getting high rating or low rating scores. However, the

Rating Scores	Time Points			
	2	5	9	10
1	-0.028	-0.039	-0.044	-0.036
2	0.216	-0.046	-0.053	0.047
3	0.134	0.077	-0.181	0.032
4	0.187	0.223	0.155	0.235
5	-0.284	-0.268	-0.086	-0.274

Table 1: The absolute values of anomalous deviations over time

anomalous time points make these causes easier to be noticed. The restaurant's management can improve the quality of their service by simply doing a deep survey at the time of the anomaly's occurrence.

This is a preliminary study on anomaly detection by only using sample datasets from Yelp. As a result, the reasons we have detected may not include all the causes of anomalous rating. In future work, we will use large-scale rating datasets to detect the various causes of anomalies for different kinds of entities by combining the textural mining method on reviews. Another area that we will further study is using other anomaly detection methods to identify anomalous time points more effectively.

References

Cheung, M. Y., Luo, C., Sia, C. L., & Chen, H. (2009). Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *International Journal of Electronic Commerce*, *13*(4), 9–38.

Feng, H., & Qian, X. (2013). Recommendation via user's personality and social contextual. In *Proceedings of the 22nd acm international conference on conference on information & knowledge management* (pp. 1521–1524).

Günnemann, N., Günnemann, S., & Faloutsos, C. (2014). Robust multivariate autoregression for anomaly detection in dynamic product ratings. In *Proceedings of the 23rd international conference on world wide web* (pp. 361–372).

Günnemann, S., Günnemann, N., & Faloutsos, C. (2014). Detecting anomalies in dynamic rating data: A robust probabilistic model for rating evolution. In *Proceedings of the 20th acm sigkdd international conference on knowledge discovery and data mining* (pp. 841–850).

Li, L., Zhang, K., Zhou, Q., & Zhang, C. (2016). Toward understanding review usefulness: a case study on yelp restaurants. *IConference 2016 Proceedings*.

Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, *19*(4), 456–474.

Lim, E.-P., Nguyen, V.-A., Jindal, N., Liu, B., & Lauw, H. W. (2010). Detecting product review spammers using rating behaviors. In *Proceedings of the 19th acm international conference on information and knowledge management* (pp. 939–948).

Lin, Y., Zhu, T., Wu, H., Zhang, J., Wang, X., & Zhou, A. (2014). Towards online anti-opinion spam: Spotting fake reviews from the review sequence. In *Asonam* (p. 261-264).

Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 632–640).

Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on world wide web* (pp. 191–200).

Zhang, K., & Pelechrinis, K. (2014). Understanding spatial homophily: the case of peer influence and social selection. In *Proceedings of the 23rd international conference on world wide web* (pp. 271–282).