

Điều chỉnh điểm đánh giá dựa trên nội dung bình luận

Vấn đề sai lệch giữa rating và nội dung bình luận

Trong hệ thống đánh giá sản phẩm, lý tưởng là điểm số (rating 1-5 sao) phản ánh đúng cảm xúc và nội dung của bình luận. Tuy nhiên, thực tế không phải luôn như vậy. Đôi khi người dùng viết một bình luận **tích cực** nhưng lại cho điểm **thấp**, hoặc ngược lại, phàn nàn tiêu cực nhưng điểm **cao**. Hiện tượng này được gọi là **sự không nhất quán giữa đánh giá sao và nội dung bình luận**, hay *text-rating review discrepancy (TRRD)* – định nghĩa bởi sự không thống nhất giữa bài đánh giá dạng văn bản và điểm số đánh giá sao tương ứng. Nghiên cứu chỉ ra rằng sự không nhất quán này có thể gây **hoang mang** và làm giảm niềm tin của người đọc vào đánh giá đó. Do đó, việc phát hiện và điều chỉnh những rating "lệch tông" so với nội dung bình luận là cần thiết để đảm bảo dữ liệu đánh giá phản ánh chính xác ý kiến khách hàng.

Phương pháp phát hiện rating sai lệch dựa trên cảm xúc bình luận

Phân tích cảm xúc (sentiment analysis) của nội dung bình luận là chìa khóa để nhận biết sự sai lệch. Trước tiên, chúng ta áp dụng mô hình NLP để phân loại mỗi bình luận thành các nhãn cảm xúc: **tích cực (positive)**, **tiêu cực (negative)** hoặc **trung tính (neutral)**. Sau đó, so sánh nhãn này với **điểm số sao** mà người dùng đã cho:

- 4-5 sao**: phản hồi tích cực.
- 3 sao**: phản hồi trung tính.
- 1-2 sao**: phản hồi tiêu cực.

Khi **nhãn cảm xúc và điểm số không khớp nhau**, đặc biệt rơi vào nhóm đối lập, ta xác định đây là một trường hợp **rating sai lệch**.

Để phát hiện tự động, ta có thể thiết lập một số **quy tắc hoặc ngưỡng** đơn giản:

- Positive** nhưng **rating** ≤ 2 sao \rightarrow đánh giá không nhất quán.
- Negative** nhưng **rating** ≥ 4 sao \rightarrow đánh giá không nhất quán.
- Neutral** nhưng **rating** $\in \{1, 5\}$ sao \rightarrow đánh giá không nhất quán.

Những điều kiện trên giúp nhanh chóng **đánh dấu** các review nghi ngờ sai lệch.

Quy trình đề xuất điều chỉnh rating sai lệch

- Phân loại cảm xúc bình luận**: Sử dụng mô hình hoặc từ điển phân tích cảm xúc để gán nhãn.
- So sánh với điểm số sao**: Xác định nhóm điểm sao và kiểm tra sự nhất quán.
- Điều chỉnh điểm rating**:
 - Bình luận tích cực nhưng rating thấp: tăng lên 5 sao (hoặc 4 sao).
 - Bình luận tiêu cực nhưng rating cao: hạ xuống 1 sao (hoặc 2 sao).
 - Bình luận trung tính nhưng rating cực đoan: đưa về 3 sao.
 - Công thức tính toán**: $(\text{rating}_{\text{đề xuất}}) = 1 + 4 \times \frac{(s + 1)}{2}$ với $s \in [-1, +1]$ tương ứng tiêu cực đến tích cực.
- Cập nhật hoặc gắn cờ**: Tùy chính sách, có thể cập nhật hoặc flag để quản trị viên kiểm tra thủ công.

Ví dụ minh họa

- Ví dụ 1 (Tích cực vs. rating thấp)**: "Sản phẩm rất tuyệt vời" \rightarrow **1 sao** \rightarrow điều chỉnh thành **5 sao**.
- Ví dụ 2 (Tiêu cực vs. rating cao)**: "Chất lượng kém" \rightarrow **5 sao** \rightarrow điều chỉnh thành **1 sao**.
- Ví dụ 3 (Trung tính vs. rating cực đoan)**: "Bình thường" \rightarrow **5 sao** \rightarrow điều chỉnh thành **3 sao**.

Lưu ý khi áp dụng thực tế

- Cẩn trọng với trường hợp người dùng cố tình đánh giá lệch (ví dụ: mía mai).
- Ưu tiên gắn cờ cảnh báo hoặc hỏi lại người dùng thay vì tự động chỉnh sửa ngay lập tức.
- Kết hợp dữ liệu **định lượng** (rating) và **định tính** (nội dung bình luận) để có cái nhìn toàn diện hơn.

Tài liệu tham khảo

- Gao, Shan, et al. "Modeling rating and review text with deep learning." (2019).
- Wu, Chen, et al. "Detecting Review Manipulation via Generative Language Models and Sentiment Inconsistency." (2022).
- Chen, Y. et al. "Identifying Inconsistent Ratings with Textual Reviews for Data Cleaning in Recommender Systems." (2021).
- StackOverflow discussions about text-rating discrepancy detection (2023).