(a) summarizes the topic

Using text mining techniques to find the most consolidated topics in the field and how often they have been the subject of reviews and citations throughout the years.

They extracted topics from 1,744 review papers and 62,952 citation papers using VOSviewer, and N-grams, and convert them into TDM (881 non-sparse terms). Then they use latent semantic analysis to create 76 clusters after merging similar high-loading terms. Finally, they generate a concise term list and provide a co-occurrence network and heatmap of the specific term per cluster.

(b) discusses the importance of the topic for ML and DS

They use VOSviewer to extract one-word terms from the title and abstract of the document, or automatically extract author keywords and build clusters based on the number of co-occurrences among the terms. They successfully cluster these tokens and provide a heat map to show what new topics are being explored and how OR/MS methodologies are being incorporated into new fields.

The paper shows that we can use ML/DS techniques to get data insight from these academic papers, to see the yearly research trend, such as optimization, and supply chain management. Moreover, we can see that the correlation/relationship between these terms, such as supply chain, is highly related to operation management, so we can suggest collaboration for these two topics.

(c) offers feedback about the presentation.

I like the overall presentation, especially since it mentions the limitation of the paper such as the quality or impact of the articles, topic modeling, and only four journals analyzed. And the term–term matrix is straightforward to analyze the relationships between topics. The co-occurrence network looks amazing and gives viewers a lot of insights. One recommendation is for extracting topics, I would like to see how the original abstract/title converted into TDM.

(a) summarizes the topic

Find the most important predictors of an explicit recommendation in the review versus non-recommendation advice. They use IBM SPSS Modeler Text Analytics Engine to extract text from a random sample of 1,112,708 reviews. Generate 17 categories for sentiment analysis using NLTK, naïve Bayes, and sklearn. Using a Lexicon-based approach to evaluate characters of the words to determine text sentiment with a bag of words and NRCLex.

(b) discusses the importance of the topic for ML and DS

Sentiment analysis is important in sorting data at scale, real-time analysis, market research, etc. Due to online reviews driving the consumer decision-making process, understanding which keywords contribute to positive/neutral/negative reviews, helps owners of the business can design user-specific responses or marketing promotions.

(c) offers feedback about the presentation.

The presentation is easy to follow and relevant to our lives about explicit review. The limitation of Lexicons with examples helps me to understand the same word such as killing can express anger or happiness based on different situations. I really like the example of showing how original reviews transformed into a list of terms using stemming, grouping words based on part of speech, etc. There could be some more explanation about the five models used in the paper: Probit, Binary Logistics, CHAID, C&RT, and Random Forest, and why they all have similar accuracy around 66%.

Case study 4: 18 mins

(a) summarizes the topic

Emotional text mining is a kind of sentiment analysis to identify the elements setting people’s interactions, attitudes, and expectations. ETM is suitable for marketing, advertising, and e-commerce. Using an Unsupervised model used to detect both semantic and semiotic meaning in the text. ETM can extract useful information for business decision-making.

(b) discusses the importance of the topic for ML and DS

Emotional text mining is useful for ML and DS because it provides business insights for social media marketing because is fast, cheap, and simple. The data is easier to gather instead of doing a survey offline. Using k-means with cosine similarity as a distance metric to identify positive, negative, and neutral sentiment analysis. Studies on tweets can show the potential of the characteristics of Twitter user communities such as product preferences or sentiments.

(c) offers feedback about the presentation.

The presentation works well and at a good pace. I like how your guys explain the emotional text mining procedure, very clearly and understandable. The two findings correspondence analysis and network analysis from 107,500 English tweets containing the brand name help me to understand the approach. Some of the visuals are hard for me to interpret but the oral explanation solved that issue.

(a) summarizes the topic

Bilingual text mining on social media focuses on O2O (online to offline). Finding similarities and differences between English and Chinese tweets helps the decision maker to improve brand loyalty programs, increase brand awareness, and better data collection. Lemmatization to stem words, spell checking, calculation of weight for every word, and text parsing procedure helped in removing noise. The findings can assist businesses in understanding the various trends in O2O markets and providing different content or strategies for different language users

(b) discusses the importance of the topic for ML and DS

Twitter is the primary source of social media and studying tweets helps e-commerce understand customers via feature-based document representation. Frequency weight and term weight are used to recognize and calculate the weight of the terms. Concept linkage connects related data or documents by identifying commonly shared concepts based on their co-occurrence. Design plans from data insight to increase sales, and revenue, maximizing interactions. Trend analysis can provide businesses, with analyzing publicly available data from social media, identifying weaknesses, and discovering new opportunities.

(c) offers feedback about the presentation.

The presentation has a clear and detailed explanation of the data procedure, especially how they prepare the data. The original dataset has 19,273 English tweets, and 17,721 Chinese tweets then become 16,720 and 14,293 after deleting One2One Network for bias avoidance. And also for a customized stop list, term-by-document matrix (each entry represents the number of times a term appears in a document), and remove terms with low frequency and irrelevant terms. The ways to preprocess the Chinese language are interesting because the Chinese language does not have delimiters, low syntax, and segmentation so they only contain nouns and proper nouns, and their abbreviations while ignoring verbs, adjectives, adverbs, etc.