Applications of text retrieval and analysis in research of food science and nutrition

Publicly available text data sources such as social networks, review websites, research literature etc. offer a variety of scopes for understanding different perspectives of dietetics and nutrition research. A number of studies have recently been analyzing text collections from these sources from different angles such as understanding food habits of individuals, identifying growing popularity of specific diet patterns, detecting foodborne illness outbreaks, assessing health quality of dietary supplements, recognizing different forms of bias in food security literature etc. These studies provide insight and directions on promising approaches and applications of utilizing text retrieval and analysis in advancing the broader domain of food science and nutrition research. Some example studies are briefed here.

Researchers at University of Illinois hypothesized that only a small proportion of people who fall sick due to contaminated food intake visit hospitals and so estimates of the severity and outspread of such diseases based on only medical records lack a lot of positive cases [1]. But analyzing discussions from social media can give a stronger signal to potential outbreaks. The researchers also demonstrated that mining text from scientific databases and literature abstracts provides scientists better understanding of relationships among various health related entities like diet, genes and disease.

Another limitation of studies based on food and nutrition experiments is that usually these are based on small cohorts participating in surveys or trials. But nowadays millions of people go online to post food reviews, share diet experiences, and discuss health conditions. Twitter and Instagram posts and Google searches have been used in studies to identify correlations between obesity related tweets and prevalence of obesity in U.S. adults.

Another study [2] focused on concepts of eating behavior from analyses of the content of food-related posts viewed on social networking sites, particularly Twitter. The study analyzed a collection of tweets about four meals - breakfast, lunch, dinner, and snacking. Text mining techniques and manual analysis were performed on more than 59000 unique tweets which reflected knowledge about the patterns of consumed foods, associated queries such as when, where, and why were the particular meals consumed. The study revealed that people liked

to describe eating experiences, and refer to foods and beverages consumed, as well as to express emotions towards food in their tweets. This study also provided suggested diet and health-related behavior change for tackling the obesity as well as improvement of health condition. The results of this study were suggested to be useful resource of recommendation for public health policies.

Another study [3] looked at utilizing text mining to qualify health benefits of dietary supplements. The researchers compared a fictional plant-based product was used with other products containing at least one of the tested ingredients registered in the years 2007–2019 in the register of dietary supplements kept by the Chief Sanitary Inspectorate (GIS), which were given either the "consistent" or "to be clarified" status. The obtained results concerned the frequency of occurrence of the individual ingredients (St John's wort/Hypericum, melissa, rose root/Rhodiola) in other products, considering their status in the register. The data obtained was subjected to classical statistical analysis in order to find correlations between the presence of a given ingredient and the product status. In view of the obtained results, the text mining analysis may be considered as a helpful tool in the process of internal risk assessment performed by manufacturers of dietary supplements.

Another study [4] looked at collection of PubMed publications to find out possible preventive nutrients and food items of metabolic syndrome. Collecting 992 abstracts from PubMed using appropriate search terms like metabolic syndrome, nutrition, and prevention, relevant nouns were extracted using a text mining tool. Then the most significant nouns were utilized to construct a network where connections among entities reflected strong association. The network was analyzed to figure out possible preventions of metabolic syndrome.

Another study [5] looked at food security literature through text mining approaches which revealed substantial spatial bias and thematic expansion over time. Mining research on food security conducted in 187 different countries, the study found little correlation between where food security related researches were conducted and where actual food security concern existed. It revealed a significant bias in where such researches were being conducted. Also, application of topic modelling - which is a specialized tool of text mining - identified that the themes of such researches gradually changed over time. Initially the researches focused primarily on economic policy and global issues, which later transitioned to concepts like livelihoods, health, and the environment.

Another study [6] analyzed nutrition and dietetics related publications through text mining techniques to identify popular and useful software and toolkit used in such research works. The study used a bag-of-word based text mining method for the task. The study identified different crucial statistical approaches such as numerical descriptive statistics, specific hypothesis tests, general hypothesis concepts, regression, ANOVA and tools such as IBM SPSS statistics to be essential for introducing in curricula of nutrition related academia.

Here we have looked at several studies which dealt with nutrition, food, diet related text data from a variety of sources and analyzed those through text mining techniques to gain important insights and perspectives. Text analysis in this domain is a relatively new area, and new studies are continuing to utilize power of text retrieval and analysis techniques in empowering knowledge of this domain. The presented studies offer an overview of scope and impact of such applications of text mining in nutrition and food science.

References

[1] *MS Windows NT Kernel Description*. http://web.archive.org/web/20080207010024/http://www.808multimedia.com/winnt/kernel.htm. Accessed: 2010-09-30.

[2] Ayuna Dondokova, Satyabrata Aich, Hee-Cheol Kim, and Gyung Hye Huh. "A Text Mining Approach to Study Individuals' Food Choices and Eating Behavior Using Twitter Feeds". In: *Lecture Notes in Electrical Engineering*. Springer Singapore, 2019, pp. 520–527. DOI: 10.1007/978-981-13-3648-5_60. URL: https://doi.org/10.1007/978-981-13-3648-5_60.

[3] Kacper Wróbel and Anna Justyna Milewska. "Using "Text Mining" Analysis for the Assessment of the Health Quality of Dietary Supplements". In: *Studies in Logic, Grammar and Rhetoric* 64.1 (Dec. 2020), pp. 15–26. DOI: 10.2478/slgr-2020-0037. URL: https://doi.org/10.2478/slgr-2020-0037.

[4] Hweesoo Jeong, Dongwook Lee, and Donggyu Jeong. "Metabolic Syndrome and Nutrition: Text Mining and Network Analysis Using Abstracts of PubMed". In: *Current Developments in Nutrition* 4.Supplement_2 (May 2020), pp. 211–211. DOI: 10.1093/cdn/nzaa043_062. URL: https://doi.org/10.1093/cdn/nzaa043_062.

[5] Matthew W. Cooper, Molly E. Brown, Meredith T. Niles, and Moataz M. ElQadi. "Text mining the food security literature reveals substantial spatial bias and thematic broadening over time". In: *Global Food Security* 26 (Sept. 2020), p. 100392. DOI: 10.1016/j.gfs.2020.100392. URL: https://doi.org/10.1016/j.gfs.2020.100392.

[6] Alison Coenen, Marijka J. Batterham, and Eleanor J. Beck. "Statistical methods and software used in nutrition and dietetics research: A review of the published literature using text mining". In: *Nutrition &amp Dietetics* 78.3 (June 2021), pp. 333–342. DOI: 10.1111/1747-0080.12678. URL: https://doi.org/10.1111/1747-0080.12678.