Multiple Choice Intelligent Means of Automatic Alternate Choice Selection

Inventor: Andrew Ferlitsch

Abstract: Hand generation of alternate answers (wrong choices) for multiple choice questions can be problematic. Given an existing corpus of questions and correct answers (e.g., 1000 Q/A), one needs to hand generate three alternate choices per question (e.g., 3000 alternate choices). Beyond the tediousness, there is also the question of quality. If the alternate choices are too far away, they will be obvious for elimination and if too close, they will be confusing. Therefore, the human generator must be highly skilled and experienced within the domain of the corpus.

This invention discloses a means of an intelligent method of automated generation of alternate multiple choice answers that solve the aforementioned problem in an existing corpus of questions. The method is in the field of artificial intelligence and natural language processing.

Overview:

The system 100, depicted in Figure 1, consists of a corpus of questions and corresponding answers (Q&A) 110, a natural language processing (NLP) agent 120, which produces a set of similarity matches 130 from the Q&A corpus 110. A second agent 140 produces sets of multiple choice questions 150 from a combination of the Q&A corpus 110 and the similarity matches 130.

The Q&A Corpus 200 and 110, depicted in Figure 2, consists of a database schema 210 and rows of data 220. The schema consists of the following:

1. Unique ID for the Question/Answer.
2. The Category (Corpus) the Question/Answer belongs to.
3. The Question.
4. The Answer.
5. The Question converted into a Bag of Words.
6. The Answer converted into a Bag of Words.
7. Data on Similarity to other Questions/Answers in the same Corpus
8. A measurement of the Question’s difficulty to answer.

First Embodiment:

In the first embodiment of this invention, the text of the question and the text of the answer 310 and 210 are ran through a NLP process 300 and 120, depicted in Figure 3, to generate two sets of bags of words 360, one for the question and one for the answer 360.

The text of the question and the text of the answer 310 and 210 are first canonically processed 320 to remove punctuation and lowercasing. The post-processed text is then processed to remove stop words 330. These are common words that add little or no value in distinguishing the uniqueness of the text (e.g., the, and, but). There are many standard dictionaries of stop words available for use. Once stop word removal 330 is completed, the post processed text then is processed for lemmatization 340. In this process, words are converted to their root forms. For example, reads and reading are converted to read, splitting and splitted are converted to split. Once lemmatization 340 is completed, the post processed text is then scanned to remove duplicate words 350.

The result are two bag of words list 360, one for the question and one for the answer, that contain non-duplicated unique words from the corresponding question and answer text 310. Below is an example representation:

{ id: 1, ‘qBagofWords’: ‘product,owner,role’,’aBagofWords’: ‘person,product,stakeholder,responsibility,decide,story,backlog,sprint’ }

Each question and answer 310 is then processed against each of the remaining questions for similarity 400, depicted in Figure 4. A compactor 420 compares the number of unique matches of words from the question’s bag of words to all the question bag of words for all the remaining questions and answers in the same corpus 430. The process is repeated by the comparator 420 between the answer’s bag of words to all the answer bag of words for all the remaining questions and answers in the same corpus 430.

The total number of matches for both the question and answer bag of words are then combined. For example, if the question bag of words had two matches and the answer bag of words had three matches to another question/answer in the corpus, the value of the combined match would be five. The matches for a question/answer across the entire corpus is kept in a similarity list 440 and sorted by highest matching count. Below is an example representation:

{ id: 1, similarity: [{ id: 3, count: 5 }, { id: 17, count: 4 }] }

In the above example, the question with id 1 has two similarity matches. One with a question with id 3, with a count of 5, and one with question with id 17, with count 4.

In another aspect of this embodiment, the answer bag of words similarity matches are processed for the removal of duplicate (identical) answers 450. If the answer bag of words is identical to any other answer bag of words in the remaining questions/answer, it is eliminated from the similarity list 440. Likewise, the answer bag of words of each new similarity match is compared to the existing matches, and eliminated from the similarity list 440 if already identical to one in the existing similarity list 440.

Second Embodiment:

In a second embodiment, the automated multiple choice generation process 500 and 140, depicted in Figure 5, selects a combined similarity matches 510 and 440 for a question/answer 310 from the Q&A corpus 110. A multiple choice selector 520 selects the three entries 530 from the combined similarity matches 510 and 440.

In one method of this embodiment, the top three closest matches with the highest match count are selected. In another method, a nearest neighbor approach is used. A "k" number, where k >= 3, of the closest matches is considered; whereby, a random selection is made within the set of k selections. In another method, a determination is made if any of the closest matches is too similar, thus may lead to confusion, and eliminated from the list for consideration.

This later method of selecting from a k >= 3 best-fit similarity selection system 600, is depicted in Figure 6. The combined similarity matches 510 and 440 for a question in the Q&A Corpus 110 is plotted on a similarity index 610, from zero (no similarity) 620 to N (maximum similarity) 630. A threshold of similarity 650 is determined; whereby a similarity index value greater than the threshold 650 is considered too similar 640 to the context of the question, and cause confusion.

A threshold of non-similarity 680 is determined; whereby a similarity index value less than the threshold 680 is considered too non-similar 690 to the context of the confusion, and would be considered obviously not the answer. Between the threshold of too similar 650 and too non-similar 680 is a range k 660. If range of k 660 exceeds three, then the best-fit similarity method will randomly pick three alternate choices between within the k 660 range.

Continuing with Figure 5, once the multiple choice selector 520 selects the three alternate choices, combined with the correct answer, 530, the order of the multiple choice answers 530 is randomized by the randomize position process 540. For example, the correct answer 720 may be put into choice position 2, and the alternate choices 730 in position 1, 3 and 4, when displayed as a multiple choice menu 700 along with the question 710, as depicted in figure 7.

Third Embodiment:

In the third embodiment, a timing metric is maintained on users answering the multiple choice questions, as depicted in figure 8. Over an extended period of time, a mean average time for answering multiple choice questions in the same corpus is computed. The mean average is used as the center (peak) 810 for a normal equation distribution 800. A minimum time threshold 820 and a maximum time threshold 830 are computed. Timing between the minimum 820 and maximum threshold 830 is considered normal range 840.

Additional, the mean average time per question is maintained. If the mean average time for a multiple choice question is below the minimum threshold 720, it is presumed that one or more of the alternate choices is too obvious, and if the mean average time exceeds the maximum threshold 730, it is presumed that one or more of the alternate choices is too confusing.

Fourth Embodiment:

In a fourth embodiment, the mean average time per multiple choice question is maintained and compared to the computed normal equation 700 for the corresponding corpus. If the mean average is outside of the thresholds 820 or 830, then the system will attempt to find a different selection of alternate choices 730. In one example method, the A/B Testing method is used. One of the existing alternate choices 730, within range k 660 is selected for replacement by an alternate choice not previously selected, such as choice k + 1. The swap is maintained for a period of time with a new mean average time calculated over that time period. If the new mean average is within the normal range 840, then the swap is kept; otherwise, another alternate choice is selected at random for the same swap.
If no combinations of swapping the existing alternate choices 730 results in a new mean average within the normal range 840, then the process is repeated again but with swapping with the next non-selected alternate choice k + 2, etc.

Fifth Embodiment:

In a fifth embodiment, the generated multiple choice question 700 in figure 7 is replaced with the generated multiple choice 900 as depicted in figure 9. In this version, the select correct option (radio box) is replaced with an exclude (X) selector. So instead of selecting the correct answer, the user excludes the non-correct answers 930 until only the correct answer remains 920.

The order and corresponding frequency in which the users exclude alternate choices is maintained, with the presumption that users exclude not in chronological order but in order of most to least obvious not correct.

The same normal equation depicted in figure 8 is computed as well, and the process of observing if the mean average of an individual multiple choice question is outside of the k range 840. If so, the frequency order of exclusion is used in selecting an alternate choice 930 for swapping with an unselected alternate choice k + 1. In the case that the mean average for the question is below the minimum threshold (too obvious) 820, then the alternate choice 930 computed to have the highest frequency of being excluded first is chosen first for swapping with k+1. In the case that the mean average for the question is above the maximum threshold (too confusing) 830, then the alternate choice 930 computed to have the lowest frequency of being excluded first is chosen first for swapping with k+1.

If no combinations of swapping the existing alternate choices 930 results in a new mean average within the normal range 840, then the process is repeated again but with swapping with the next non-selected alternate choice k + 2, etc.