

variable and/or the like) for use in machine learning models (e.g., deep learning models, neural network models, and/or the like). For example, embedding vectors may include multi-dimensional, learned, continuous vector representations of discrete variables that may reduce the dimensionality of categorical variables and meaningfully represent categories in the transformed space. In some non-limiting embodiments or aspects, embedding vectors may simplify machine learning on large inputs, such as sparse vectors representing words. For example, an embedding vector may capture semantics of an input (e.g., a word from a corpus of documents) by placing semantically similar inputs close together in the embedding space. In some non-limiting embodiments or aspects, an embedding vector may be learned and reused across models. In some non-limiting embodiments or aspects, an embedding vector may include a vector representation of a data string (e.g., a word, a merchant identifier (e.g., name, identification number, etc.)) based on said string's distributions in a corpus of data (e.g., a set of documents (e.g., books and/or the like), consumer transaction data (e.g., a series of merchant identifiers that each consumer account of a set of accounts transacted with and/or the like), etc.).

[0078] Non-limiting embodiments or aspects of the present disclosure are directed to a system, method, and computer program product for debiasing embedding vectors of neural network machine learning models. It will be appreciated that described systems and methods improve the accuracy and efficiency of computer-driven, decision-making systems by reducing and/or removing bias from embedding vectors in machine learning models. Because embedding vectors may similarly represent input variables having similar semantic values, and because bias may be amplified in embedding vectors, the removal of bias allows a machine learning model to detect semantic similarity resulting from other feature dimensions besides a source of bias. For example, semantic similarities that otherwise would have been detected but are not identified because of obfuscation due to bias may cause a decision-making system to not trigger a computer process when it should (e.g., false negative). Additionally or alternatively, semantic similarities that are detected but would not be identified if bias were removed may cause a decision-making system to trigger a computer process when it should not (e.g., false positive). For decision-making systems making a number of decisions to execute or not execute computer processes every second (e.g., fraud monitoring system), the reduction of false negatives and false positives

implementation 400 may be executed by one or more processors of modeling system 104. Additionally or alternatively, one or more steps of implementation 400 may be executed (e.g., completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 104, such as user device 102, fraud monitoring system 106, decision-making system 108, and/or the like. Each step of implementation 400 may be performed by a same or different processor.

[0124] As shown in FIG. 4A, in some non-limiting embodiments or aspects, a plurality of embedding vectors 401, 403 may be received. (For clarity, each embedding vector 401, 403 is illustrated as a dot rather than a ray. Each dot may represent the endpoint of the respective vector.) For example, modeling system 104 may receive a plurality of embedding vectors 401, 403, as described herein. Embedding vectors 401, 403 may include first embedding vectors 401 in a first category and second embedding vectors 403 in a second category, as described herein. In some non-limiting embodiments or aspects, the first category and the second category each may be a respective subcategory of a category of bias. For example, the first category may include a first location and the second category may include a second location, and location may be a category of bias for merchant embedding vectors. For example, the first category may include a first gender (e.g., male) and the second category may include a second gender (e.g., female), and gender may be a category of bias for word embedding vectors.

[0125] In some non-limiting embodiments or aspects, at least two clusters may be generated. For example, modeling system 104 may generate two clusters 421, 423 based on first embedding vector 401 and second embedding vector 403, as described herein.

[0126] In some non-limiting embodiments or aspects, first mean vector 411 of first cluster 421 and/or second mean vector 413 of second cluster 423 may be determined. For example, modeling system 104 may determine first mean vector 411 of first cluster 421 of first embedding vectors 401 and second mean vector 413 of second cluster 423 of second embedding vectors 403, as described herein. In some non-limiting embodiments or aspects, an initial direction 431a may be determined based on first mean vector 411 and second mean vector 413, as described herein.

[0127] In some non-limiting embodiments or aspects, second mean vector 413 may be modified by moving second mean vector 413 toward each (second) embedding