Здесь будет реферат статьи.

Sentiment Lexicon Expansion Based on Neural PU Learning, Double Dictionary Lookup, and Polarity AssociationSentiment Lexicon Expansion Based on Neural PU Learning, Double Dictionary Lookup, and Polarity Association

*Как же я люблю BTS! Вот они, слева направо: Yasheng Wang, Yang Zhang, Bing Liu*

*Здравствуйте, это Макдональдс? Мне пожалуйста Неурал Пию Лернинг, Двойной Дикшинэри-Лукап и Поларити Ассосиэйшн-Сентимент, соуса не надо*

Рассмотрев посты в социальной сети Weibo, Yasheng Wang, Yang Zhang и Bing Liu выяснили, что многие китайские эмоциональные слова не включены в словари, потому что эти слова слишком новые или употребляются в письменной речи более “низкого стиля”. Авторы поставили перед собой задачу ~~нажраться и не спалить дачу~~ расширить sentiment словарь китайского языка за счёт машинного обучения. Это задача PU Learning. Что такое PU (Positive/Unlabeled) Learning? PU Learning - это задача где у нас есть выборка, часть объектов из которой размечены как принадлежащие к одному классу (P), а все остальные - не размечены вообще (U). При этом подразумевается, что среди неразмеченных объектов как принадлежащие к этому классу, так и не принадлежащие к нему. Задача PU Learning состоит в том, чтобы построить такой классификатор, который разделял бы тестовую выборку на объекты, принадлежащие к заданному и не принадлежащие к нему. Отличие от бинарной классификации состоит в том, что мы сначала размечаем часть наиболее “надёжных” примеров в тестовой подвыборке и на них обучаем классификатор.

“A typical PU learning algorithm works by first identifying a small set of reliable N class examples (RN) from the unlabeled set U and then running a supervised learning method (e.g., SVM) it- eratively to add more and more data to the RN set to finally build a classifier (Liu, 2011)”