OSTIS-2015
(Open Semantic Technologies for Intelligent Systems)

УДК 681

ИНФОРМАЦИОННАЯ МОДЕЛЬ ЛИЧНОСТИ

Харламов А.А.

Институт высшей нервной деятельности и нейрофизиологии РАН, г. Москва

kharlamov@analyst.ru

В работе рассмотрены вопросы автоматической обработки текстов, описывающих личность, а также сопутствующих текстов, эффективного представления этой информации, а также – разработки удобных интерфейсов упомянутой информации с пользователем.
Ключевые слова: информационная модель личности, семантические представления, лингвистическая обработка текстов, диалог с пользователем

Введение

Если не принимать во внимание психологических особенностей личности, за основу информационной модели личности можно взять языковую модель личности [Караулов, 1987], в которой структура личности рассматривается как совокупность трех уровней, а именно: 1) вербально-семантического, где лексикон личности, понимаемый в широком смысле, включает также фонд грамматических знаний личности; 2) лингвокогнитивного, представляющего тезаурус личности, в котором запечатлен "образ мира", или система знаний о мире; и 3) мотивационного уровня - уровня деятельностно-коммуникативных потребностей, отражающего прагматику личности: систему ее целей, мотивов, установок и интенциональностей.

Расширим это представление в содержательную сторону – добавим в модель совокупность информационных источников, характеризующих моделируемую личность. Имея в виду существующие на настоящий момент технологии хранения и обработки информации, совокупность текстов, принадлежащих моделируемой личности, или описывающих ее, в наибольшей степени содержит информацию для моделирования.

Существующие медийные технологии лишь в небольшой степени содержат информацию для моделирования, причем, в основном, внешнего к содержанию характера. Поэтому, в работе речь пойдет исключительно об использовании текстовой информации для моделирования личности.

В работе рассмотрены вопросы создания информационного автомата, который реализует автоматическую обработку текстовой информации, формируя те самые три уровня языковой личности, и их расширение в сторону использования базы знаний [Харламов, 2006].

1. Модель мира человека

С информационной точки зрения модель личности в наибольшей степени определяется моделью мира упомянутой личности. Поэтому сначала поговорим о модели мира человека. Модель мира человека состоит из трех компонентов, в том числе, индивидуального многомодального образного правополушарного (у правшей), социализированного многомодального схематического левополушарного, и языкового, который также локализован в левом (у правшей) полушарии. Многомодальные компоненты не просты по структуре, по обработке информации, и потому, плохо моделируются. А, главное, они в малой степени архивируются в процессе деятельности человека. Языковой компонент, напротив, как правило, хорошо заархивирован, хорошо обрабатывается и визуализируется. А, главное, языковой компонент, будучи, запараллелен в процессе формирования с двумя многомодальными, поэтому, изоморфен по представляемой информации упомянутым многомодальным компонентам, и потому, может быть полноправным их представителем.

Языковой компонент модели мира человека представляет собой, с одной стороны, иерархию лингвистический представлений, а с другой – надлингвистические описания семантики и

прагматики модели мира человека. Языковой компонент моделирует, в своей нижней части, формальные основы языка – графематику, морфологию, синтаксис и семантику отдельного предложения. Надлингвистическая часть языкового описания картины мира включает статику семантических представлений и динамику прагматического описания модели мира – ее описательную и алгоритмическую части.

1.1. Модель языка

Модель языка в составе модели мира человека является иерархией представлений, включающих в себя словари, учитывающие формальную сочетаемость языковых единиц разных уровней. В том числе, словарь первого уровня – $\{B_i\}_1$ – морфологический словарь [Харламов с соавт., 2013a], второго уровня – $\{B_k\}_2$ – лексикон, словарь третьего уровня – $\{B_l\}_3$ – словарь синтаксических структур [Харламов с соавт., 2013b], и словарь четвертого уровня – $\{B_m\}_4$ – словарь попарной сочетаемости слов в тексте [Рахилина, 2000].

1.2. Семантическая сеть

Словарь попарной сочетаемости корневых основ $\{B_m\}_4$ позволяет построить сеть, характеризующую смысловую структуру текста. Получается так называемая ассоциативная (однородная семантическая) сеть N как совокупность несимметричных пар понятий (корневых основ) $< c_i c_j >$, где c_i и c_j – понятия (корневые основы), связанные между собой отношением ассоциативности (совместной встречаемости в некотором фрагменте текста, например, в предложении) [Харламов с соавт., 2008]: $< c_i c_j >= B_m \in \{B_m\}_4$:

$$N = \cup_{ij} < c_i c_j > . \qquad (1)$$

Эта сеть получается также, если предварительно пары слов объединить в звездочки: все пары слов с одинаковым первым словом это одна звездочка – $d =< c_i < c_j >>= \cup_j < c_i c_j >$, где $< c_j >$ - множество семантических признаков главного слова c_i.

$$N = \cup_i < c_i < c_j >> . \qquad (2)$$

Семантическая сеть, характеризующая сочетание ключевых понятий в рамках целого текста, является первым надлингвистическим уровнем языковой модели мира. В ней учитываются значимость отдельных ключевых понятий и их взаимосвязи в целом тексте. Если рассматривать модель мира как множество текстов, описывающих мир человека (можно говорить также о множестве моделей предметных областей, которые также описываются текстами), то взаимосвязи ключевых понятий и их ранги и составляют основное содержание семантического представления модели мира.

Для более тонкого анализа семантики можно заменить ассоциативные связи между ключевыми словами на весь спектр связей, используемых в языке. Для этого на морфологическом уровне выявляется вся морфологическая информация о словах $\{B_i\}_1 = \{m_i\}$, а на синтаксическом – информация о связях слов в группах и между группами $\{B_k\}_3 = \{r_k\}$, где r_1 – предикативная связь субъекта с главным объектом, а $r_k \mid k > 1$ - все остальные типы связей. При этом структуры синтаксического уровня укладываются в рамки словаря шаблонов минимальных структурных схем предложения, и словаря валентностей глаголов.

Тогда для каждого простого предложения можно построить расширенную предикатную структуру, которая после небольших преобразований сводится тоже к звездочке $d = \cup_j < c_i r_k c_j >$. Правда, в отличие от звездочки с простыми ассоциативными связями, в звездочке, построенной из расширенной предикатной структуры, вместо пар понятий используются тройки $< c_i r_k c_j >$, где между парой понятий имеется связь, размеченная одним из k типов отношений [Kharlamov et al., 2008].

При этом ассоциативная сеть N может быть построена и из таких звездочек тоже:

$$N = \cup_i d_i = \cup_i < c_i < r_{ij} c_j >> . \qquad (3)$$

1.3. Прагматическое описание

Необходимо заметить, что семантическая сеть текста одновременно включает в себя все понятия текста. Но если спроецировать предложения текста на эту сеть, то мы получим последовательность понятий сети, которые следуют друг за другом последовательно во времени. Эти понятия включены в фрагменты текста, которые либо являются описаниями чего-либо, либо описывают алгоритм реализации чего-либо.

Отдельные предложения этих фрагментов описывают отдельные фрагменты ситуации. Расширенной предикатной структуре отдельного предложения соответствует, после описанных выше преобразований, звездочка $d = \cup_j < c_i r_{ij} c_j >$. Тогда цепочка расширенных предикатных структур содержит смысл этих фрагментов – описаний, или алгоритмов:

$$D = (d_i \mid i = \overline{1, N}) . \qquad (4)$$

1.4. Модель предметной области

Подберем корпус текстов таким образом, чтобы он описывал некоторую предметную область. Смысл последовательностей предложений этого корпуса текстов может быть представлен

последовательностями расширенных предикатных структур этих предложений. То есть последовательность расширенных предикатных структур (и цепочек соответствующих звездочек), соответствующих предложениям корпуса текстов и является моделью предметной области:

$$M = \cup_i D_i . \quad (5)$$

1.5. Гипертекстовая структура

Семантическая сеть текста (или всего корпуса текстов) вместе с самим текстом представляют собой гипертекстовую структуру, которая позволяет ассоциативно навигировать по тексту. При этом каждое ключевое понятие семантической сети соотносится с множеством предложений текста, в которых оно содержится, ну и с позициями этих предложений в тексте.

2. Модель личности

Модель личности условно мы представили множеством текстов, которые, с одной стороны, были сгенерированы этой личностью («сто томов партийных книжек» В.И. Ленина), с другой стороны – это тексты, в которых эта личность описывается. Необходимо добавить, что модель личности обычно включает фоновые знания, характерные для ее эпохи, то есть множество текстов, описывающих текущее состояние общества (или последовательность таких состояний для разных промежутков времени). Первые два множества текстов (первое подмножество – в большей степени, чем второе) описывают языковой эквивалент индивидуального многомодального компонента модели мира человека. Второе множество соответствует социализированному компоненту модели мира человека, так как оно в большей степени, чем первые два, характерно для представлений всего общества.

2.1. Индивидуальная модель

Индивидуальная часть модели личности, таким образом, строится на основе персональных текстов, сгенерированных личностью. Это письма, дневники, художественные, и не очень произведения (вплоть до технических отчетов). Помимо персональных текстов для формирования индивидуальной части модели личности используются тексты, в которых эта личность описана – художественные и прочие произведения.

2.2. Социализированная модель

Социализированная часть модели личности также строится на основе текстов из СМИ, Интернета. Но для ее построения, в отличие от текстов, использованных для построения индивидуальной части модели, берутся тексты на темы, характерные для индивидуальной части модели, но не относящиеся к моделируемой личности (дополняющие тексты предыдущего раздела).

Так для запроса пользователя: «Что нам делать с Украиной?» из Интернета скачиваются тексты по запросу «Украина». Для этого из индивидуальной модели мира личности – семантической сети – извлекается минимальный древовидный подграф – тематическое дерево, в котором отыскивается тема «Украина». Рассматриваются первые ее подтемы, затем – ниже. Из Интернета скачиваются тексты по запросам, соответствующим этим темам. Для каждой темы формируется своя модель предметной области.

2.3. Целеполагание

Для эффективного взаимодействия модели личности с пользователем необходимо уяснить, что нужно от нее пользователю. И вообще пользователю предпочтительны не ответы на вопросы, а активное поведение модели. Однако внесение потребностей в систему является отдельной большой проблемой. Поэтому, мы ее попытаемся избежать. Заменим потребности системы потребностями пользователя. Для этого нам необходимо выяснить исходную точку интересов пользователя на семантической сети и конечную точку. Далее, поведение модели определяется цепочкой вершин – ключевых понятий – на семантической сети от начальной точки до конечной точки. Выяснение намерений пользователя осуществляется в процессе диалога, о котором пойдет речь ниже.

3. Интерфейс

3.1. Диалог

Модель личности хороша сама по себе. Но наиболее эффективное (и эффектное ее применение), несомненно, возможно в диалоге с пользователем (оппонентом): «Владимир Ильич, ну и где Ваш коммунизм?» - «Э, батенька, условия изменились, да и люди теперь не те!».

Так для запроса пользователя: «Что нам делать с Украиной?» (см. Раздел 2.2) строится их общая семантическая сеть предметной области «Украина». Выявляются таким же образом темы и подтемы суммарной модели предметной области. Пользователю генерируется вопрос: «Что Вас интересует в первую очередь?», и ему перечисляются все ключевые слова предметной области, имеющие максимальный вес. Так делается до тех пор, пока не станут понятными исходная и конечная темы.

3.2. Синтез текста

Синтез корректного текста также является непростой задачей. В настоящий момент не существует систем, реализующих корректный синтез текста. Поэтому можно идти двумя путями. Более простой вариант – просто подбирать предложения из множества подходящих

предложений, собранных в гипертекстовой структуре соответствующей модели предметной области (см. Раздел 1.5), описывающие конкретную последовательность тем (от исходной темы к конечной). Можно синтезировать предложения по правилам синтаксиса языка и с использованием шаблонов синтаксических структур для той же самой последовательности тем.

3.3. Говорящая голова

Интерфейс в виде диалога оказывается удобным реализовать с помощью аватара – говорящей головы, клонирующей голос моделируемой личности. Для этого можно использовать говорящую голову и клонирующую голос программу, разработанную Б.М. Лобановым.

Заключение

Целью настоящей работы было показать возможность реализации информационной модели личности, общение с которой не только дает представление о логике мышления моделируемого, но и позволяет взаимодействовать с моделью в реальных ситуациях.

Информационная модель личности основана на семантических представлениях моделируемой личности в виде семантической сети множества текстов, описывающих моделируемую личность, или сгенерированных ею. Семантические представления расширяются лингвистическим анализом до описания прагматики, которая моделирует динамику на семантической сети.

Модель строится как совокупность индивидуальной и социализированной частей, первая из которых представляет семантику самой моделируемой личности, вторая – семантику моделей предметных областей, характеризующих конкретные ситуации в предметных областях за конкретный промежуток времени на основе текстов из любых доступных источников.

Взаимодействие пользователя с моделью личности осуществляется в форме диалога, синтез ответа в котором осуществляется либо подбором соответствующего предложения из сформированного для модели предметной области гипертекстового представления, либо на основе синтаксических шаблонов и правил.

Модель может быть оформлена в виде аватара – говорящей головы – имитирующей голос моделируемой личности.

Работа была выполнена в рамках НИР «Исследование механизма ассоциативных связей в речемыслительной деятельности человека методом нейросетевого моделирования при анализе текстовой информации» (при финансовой поддержке Российского фонда фундаментальных исследований, Грант 14-06-00363).

Библиографический список

[Караулов, 1987] Караулов Ю.Н. Русский язык и языковая личность. /Отв. ред. член-кор. Д.Н. Шмелев. – М.: Наука, 1987.

[Харламов, 2006] Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). – М.: Радиотехника, 2006.

[Харламов с соавт., 2013a] Харламов А.А., Ермоленко Т.В., Дорохина Г.В., Журавлев А.О. Предсинтаксический анализ русско-английских текстов //Программная инженерия, № 10, 2013. – С. 37 – 47.

[Харламов с соавт., 2013b] Харламов А. А., Ермоленко Т. В. Разработка компонента синтаксического анализа предложений русского языка для интеллектуальной системы обработки естественно-языкового текста //Программная инженерия № 7, 2013. – С. 37-47.

[Рахилина, 2000] Рахилина Е.В. Когнитивный анализ предметных имен: семантика и сочетаемость. – М.: Русские словари, 2000.

[Харламов с соавт., 2008] Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации с использованием искусственных нейронных сетей, в условиях динамики внешней среды //Речевые технологии, N 3, 2008. – С. 27-35.

Alexander A. Kharlamov, Tatyana V. Yermolenko, Andrey A. Zhonin Text Understanding as Interpretation of Predicative Structure Strings of Main Text's Sentences as Result of Pragmatic Analysis (Combination of Linguistic and Statistic Approaches) //M. Zelezny, I. Habernal, A.Ronzhin Eds., LNAI 8113 Speech and Computer, Proceedings of 15th Int. Conf., SPECOM 2013, Pilsen, Czech Republic, September 2013г. – Pp. 333-339.

PERSON PHYSIOGNOMY INFORMATIONAL MODEL

Kharlamov A.A.

Institute of Higher Nervous Activity of RAS, Moscow

kharlamov@analyst.ru

Questions of automatical analysis of texts concerning person physiognomy description and also accompanied texts are represented in the paper. Questions of its effective representation and suitable user interface development are represented also.

Introduction

At the moment language person physiognomy model is well known now. For more exact representation of the person one need to extent the model in the direction of enlargement of the model specific content.

Main Part

Represented in the paper world model of person consists from three components: from individual one, socialized and language components. The language component includes the four-level language model. And the individual and the socialized components are constructed from semantic and pragmatic representations. The pragmatic representation is realized by full linguistic analysis.

Conclusion

In such a way we can consider the person physiognomy model in terms of semantic and pragmatic components which added by text corpuses of persons world model with talking head as an interface.